Known and unknown confounding in genetic studies

Eleazar Eskin
University of California, Los Angeles
Natural Variation within the Neuronal Nicotinic Acetylcholine Receptor Cluster on Human Chromosome 15q24: Influence on Heritable Autonomic Traits in Twin Pairs

Integrated Computational and Experimental Analysis of the Neuroendocrine Transcriptome in Genetic Hypertension Identifies Novel Control Points for the

Genome-wide case/control studies in hypertension: only the ‘tip of the iceberg’

Kuixing Zhang⁵, Daniel T. O’Connor


Circulation 2007:115:2271-2281; originally published online Apr 16, 2007; DOI: 10.1161/CIRCULATIONAHA.106.628859

Circulation is published by the American Heart Association. 7272 Greenville Avenue, Dallas, TX 75231-0001

American Heart Association
Learn and Live™
Figure 3. Inferred phylogeny of human RGS2 haplotypes based on 7 of the most common variants (−161, −43, −38, −3, 1891 to 1892, 2138 to 2139, and 2297). Numbers of chromosomes (Num) for each haplotype are given. Solid arrows indicate probable point mutations between haplotypes. Two dashed arrows converging on a haplotype indicate a probable recombination event. Ancestral haplotypes differ from one another by ≥2 variants. D indicates deletion; I, insertion; CauNT, NT white; CauHT, hypertensive white; AfrNT, NT black; AfrHT, hypertensive black.
Discovering Genes Involved in Disease and the Mystery of Missing Heritability
Finding SNPs associated with phenotype

...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...
...ACATGCCCAGTTTCATAAGGCC...

Blood Pressure

180
175
170
165
160
145
140
130
120
115
110
110
105
100
Null hypothesis

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \]

\[ y = \mu + e \]
**Alternative hypothesis**

\[ H_1: \text{[Phenotype]} \sim \text{[SNP]} \]

\[ y = \mu + X\beta + e \]

Blood Pressure

\[ \begin{align*}
180 \\
175 \\
170 \\
165 \\
160 \\
145 \\
140 \\
130 \\
120 \\
115 \\
110 \\
110 \\
105 \\
100
\end{align*} \]
A typical single-SNP test

\[ y = \mu + X\beta + e \]

- **y**: phenotypes (size n)
- **X**: A SNP to test
- **\( \beta \)**: contribution from the SNP
- **e**: \((n \times 1)\) random effect, \( \text{Var}(e) = \sigma_e^2 I \)
A ‘hypothetical’ true genetic model

\[ y = \mu + \sum_{i=1}^{m} X_i \beta_i + e \]

- \( y \): phenotypes (size \( n \))
- \( X_i \): i-th SNP to test
- \( \beta_i \): contribution from the i-th SNP
- \( e \): \((n \times 1)\) random effect, \( \text{Var}(e) = \sigma_e^2 I \)
True effect of a single SNP

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]
Actual test is simple

\[
y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e
\]

\[
y = \hat{\mu} + X_k \hat{\beta}_k + e
\]
There are unmodeled genetic factors

**TRUE MODEL**

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

**SIMPLE LINEAR MODEL**

\[ y = \hat{\mu} + X_k \hat{\beta}_k + e \]
Unmodeled factors are not known

TRUE MODEL

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

SIMPLE LINEAR MODEL

\[ y = \hat{\mu} + X_k \hat{\beta}_k + e \]
Entering mouse genetics...
Classical inbred strains

Complex genetic relatedness of lab strains

Phylogeny of 38 inbred mouse strains using 140,000 mouse HapMap SNPs
Complex genetic relatedness of lab strains

Phylogeny of 38 inbred mouse strains using 140,000 mouse HapMap SNPs
Complex genetic relatedness of lab strains

Body weight phenotypes of 38 inbred mouse strains from JAX MPD
What we would expect

Genome-wide association map

Cumulative p-value distribution

Q-Q plot

Confounding effects in association and eQTL studies
What we actually observed

Confounding effects in association and eQTL studies
Example of spurious associations

Body weight phenotypes of 38 inbred mouse strains from JAX MPD
Example of spurious associations

Body weight phenotypes of 38 inbred mouse strains from JAX MPD
Source of spurious association

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \quad \quad \quad H_1: \text{[Phenotype]} \sim \text{[SNP]} \]
Source of spurious association

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \]
\[ H_1: \text{[Phenotype]} \sim \text{[SNP]} \]
Many SNPs are strongly correlated to the population structure

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \quad H_1: \text{[Phenotype]} \sim \text{[SNP]} \]
Some phenotypes are strongly correlated to population structure

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \quad \text{H}_1: \text{[Phenotype]} \sim \text{[SNP]} \]
Confounding effects in association and eQTL studies

SNPs and phenotypes become indirectly correlated

\[ H_0: \text{[Phenotype]} \perp \text{[SNP]} \]
\[ H_0: \text{[Phenotype]} \sim \text{[SNP]} \]
Use of a Dense Single Nucleotide Polymorphism Map for In Silico Mapping in the Mouse

Mathew T. Fletcher¹,², Philip McClurg¹, Serge Batalov¹, Andrew L. Su¹, S. Whitney Barnes¹, Erica Lagler¹, Ron Korstanje³, Xiaosong Wang⁵, Deborah Nusskern⁴, Molly A. Bogue³, Richard J. Mural⁵, Beverly Paigen⁶, Tim Wiltshire¹*

1 Genomics Institute of the Novartis Research Foundation, 2 States of America, 3 The Jackson Laboratory

Rapid expansion of available mouse strains along with the development of new methods for generating pedigrees provides an expedient way to identify both new polymorphisms for the purpose of fine-mapping and new candidate loci (QTL) for association studies. However, SNP data have lacked the density required to map these loci. To remedy this, 470,407 allele calls from whole-genome resequencing of 19 mouse strains provided a high-quality SNP set with statistical power to infer a denser haplotype for mouse strains. We used this method to high-density lipoprotein cholesterol (HDL-C) to identify a previously unknown haplotype that could successfully explain 65% of the variation in HDL-C. More generally, this method offers a way to associate genome-wide association (GWA) results with a dense haplotype map, which could be used to infer the location of QTLs. The inferred haplotype could successfully explain 65% of the variation in HDL-C. More generally, this method offers a way to associate genome-wide association (GWA) results with a dense haplotype map, which could be used to infer the location of QTLs.

In Silico Mapping of Complex Disease-Related Traits in Mice

Andrew Grupe,¹* Soren Germer,²* Jonathan Usuka,³* Dee Aud,¹
John K. Belknap,⁴ Robert F. Klein,⁴ Mandep M. Ahluwalia,²
Russel Higuchi,² Gary Peltz¹†

Experimental murine genetic models have great potential for understanding the genetic contributions to complex traits and for the development of therapeutic strategies. In a computational method for inferring genotype from a heavy background of strains, the phenotypic and genetic traits of unphenotyped strains can be predicted. Here, we revisit the three main aspects that affect in silico analysis: First, we report on the use of marker maps to infer a computational method for inferring genotype from a heavy background of strains. Second, we introduce a novel statistical approach: a novel approach for inferring genotype from a heavy background of strains. Third, we introduce a novel statistical approach: a novel approach for inferring genotype from a heavy background of strains. An integrated in silico gene mapping strategy in inbred mice

Alessandra C. L. Cervino,*¹* Ariel Darvasi,* Mohammad Fallahi,* Christopher C. Mader* and Nicholas F. Tsinoremas*

¹Department of Informatics, Scripps Florida, Jupiter, Florida 33458 and ²The Institute of Life Sciences, The Hebrew University, Jerusalem 91904, Israel

Manuscript received August 28, 2006
Accepted for publication September 28, 2006

ABSTRACT

In recent years in silico analysis of common laboratory mice has been introduced and subsequently applied, in slightly different ways, as a methodology for gene mapping. Previously, we have demonstrated some limitation of the methodology due to sporadic genetic correlations across the genome. Here, we revisit the three main aspects that affect in silico analysis: First, we report on the use of marker maps to infer a computational method for inferring genotype from a heavy background of strains. Second, we introduced a novel statistical approach: a novel approach for inferring genotype from a heavy background of strains. Third, we introduced a novel statistical approach: a novel approach for inferring genotype from a heavy background of strains. We have found that in our examples of complex traits, in silico analysis by itself does fail to uniquely identify quantitative trait gene (QTG)-containing regions. However, when combined with additional information, it may significantly help to prioritize candidate genes. We therefore recommend using an integrated workflow that uses other genomic information such as linkage regions, regions of shared ancestry, and gene expression information to obtain a list of candidate genes from the genome.
Use of a Dense Single Nucleotide Polymorphism Map for In Silico Mapping in the Mouse

Mathew T. Pletcher1,2, Philip McClurg1, Serge Batalov1, Andrew I. Su1, S. Whitney Barnes1, Erica Lagier1, Ron Korstanje3, Xiaosong Wang3, Deborah Nusskern4, Molly A. Bogue3, Richard J. Mural5, Beverly Paigen6, Tim Wiltshire1

1 Genomics Institute of the Novartis Research Foundation, United States of America, 2 The Jackson Laboratory, 3 Jackson Laboratory, Maine, USA

Rapid expansion of available mouse lines and our ability to perform high-throughput profiling of gene expression provides an expedient way to identify candidate genetic variation for complex traits. However, common variation for the purposes of association (SNP) data have lacked the density to sufficiently identify complex trait association signals. We sought to remedy this, 470,407 allele candidates were genotyped at a minimum of 300K SNP set with statistical power. We also developed a haplotype analysis method to high-density lipoprotein cholesterol (HDL-C) traits. The inferred haplotypes were used to more easily identify and characterize QTLs.
Unmodeled factors are not known

**TRUE MODEL**

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

**SIMPLE LINEAR MODEL**

\[ y = \hat{\mu} + X_k \hat{\beta}_k + e \]
Unmodeled factors & population structure

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

**TRUE MODEL**

**CAUSAL SNPS**

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C3H</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
</tbody>
</table>

**UNMODELED FACTORS**
Unmodeled factors & population structure

**TRUE MODEL**

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

**UNMODELED FACTORS**

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>CAST</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
</tr>
</tbody>
</table>

**CAUSAL SNPS**
Unmodeled factors & population structure

\[ y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e \]

**TRUE MODEL**

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C3H</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>DBA</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>T</td>
</tr>
<tr>
<td>129S1</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>CAST</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
</tr>
</tbody>
</table>
Unmodeled factors & population structure

$$y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e$$

### TRUE MODEL

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>C3H</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
</tr>
<tr>
<td>DBA</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>129S1</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>CAST</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
</tr>
</tbody>
</table>
Dependency among unmodeled factors are ignored

$$y = \mu + X_k \beta_k + \sum_{i \neq k} ? X_i \beta_i + e$$

**TRUE MODEL**

<table>
<thead>
<tr>
<th></th>
<th>B6</th>
<th>C3H</th>
<th>DBA</th>
<th>129S1</th>
<th>CAST</th>
</tr>
</thead>
<tbody>
<tr>
<td># of shared SNPs (K)</td>
<td>9</td>
<td>9</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
</tbody>
</table>

**UNMODELED FACTORS**

**SIMPLE LINEAR MODEL**

$$y = \hat{\mu} + X_k \hat{\beta}_k + e$$
Mixed model accounts for the dependency

\[
y = \mu + X_k \beta_k + \sum_{i \neq k} ?X_i \beta_i + e
\]

**TRUE MODEL**

<table>
<thead>
<tr>
<th># of shared SNPs (K)</th>
<th>B6</th>
<th>C3H</th>
<th>DBA</th>
<th>129S1</th>
<th>CAST</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>8</td>
<td>7</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>8</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td></td>
</tr>
</tbody>
</table>

**LINEAR MIXED MODEL**

\[
y = \hat{\mu} + X_k \hat{\beta}_k + u + e
\]

\[
u \sim N(0, \hat{\sigma}_g^2 K)
\]

\[
e \sim N(0, \hat{\sigma}_e^2 I)
\]
A ‘hypothetical’ true genetic model

\[ y = \mu + \sum_{i=1}^{m} X_i \beta_i + e \]

- \( y \) : phenotypes (size n)
- \( X_i \) : i-th SNP to test
- \( \beta_i \) : contribution from the i-th SNP
- \( e \) : \((n \times 1)\) random effect, \( \text{Var}(e) = \sigma_e^2 I \)

New assumption: \( \beta_i \sim N(0, \sigma_g^2) \)
Equivalence Mixed model and Bayesian Regression

TRUE MODEL

$$y = \mu + X_k \beta_k + \sum_{i \neq k} X_i \beta_i + e$$

with assumption: $$\beta_i \sim N(0, \sigma_g^2)$$

is equivalent to

LINEAR MIXED MODEL

$$y = \hat{\mu} + X_k \hat{\beta}_k + u + e$$

$$u \sim N(0, \hat{\sigma}_g^2 K)$$

$$e \sim N(0, \hat{\sigma}_e^2 I)$$

(after some normalization of the genotypes Hayes et al, 2009)
Mixed Model Heritability

**LINEAR MIXED MODEL**

\[ y = \hat{\mu} + X_k \hat{\beta}_k + u + e \]

\[ u \sim N(0, \hat{\sigma}_g^2 K) \]

\[ e \sim N(0, \hat{\sigma}_e^2 I) \]

**Heritability**

\[ h_M^2 = \frac{\hat{\sigma}_g^2}{\hat{\sigma}_g^2 + \hat{\sigma}_e^2} \]
Mixed Model Heritability

Common SNPs explain a large proportion of the heritability for human height

Jian Yang\textsuperscript{1}, Beben Benyamin\textsuperscript{1}, Brian P McEvoy\textsuperscript{1}, Scott Gordon\textsuperscript{1}, Anjali K Henders\textsuperscript{1}, Dale R Nyholt\textsuperscript{1}, Pamela A Madden\textsuperscript{2}, Andrew C Heath\textsuperscript{2}, Nicholas G Martin\textsuperscript{1}, Grant W Montgomery\textsuperscript{1}, Michael E Goddard\textsuperscript{3} & Peter M Visscher\textsuperscript{1}
Linear mixed model

\[ y = \mu + \beta X_k + u + e \]

\[ \text{Var}(u) = \sigma_g^2 K \quad \text{Var}(e) = \sigma_e^2 I \]

- \( \sigma_e \) and \( \sigma_g \) need to be estimated numerically
- Traditionally requires matrix inversion and stochastic optimization
- Newton-Raphson algorithm may result in local optimum or failure in convergence
Linear mixed model

\[ y = \mu + \beta X_k + u + e \]

\[ \text{Var}(u) = \sigma_g^2 K \quad \text{Var}(e) = \sigma_e^2 I \]

- \( \sigma_e \) and \( \sigma_g \) need to be estimated numerically
- Assume we know them.

\[ y \sim N(\mu + \beta X, V) \quad V = \sigma_g^2 K + \sigma_e^2 I \]

- But we know only how to estimate \( \beta \) when

\[ y \sim N(\mu + \beta X, \sigma_e^2 I) \]
Linear mixed model trick

\[ y = \mu + \beta X_k + u + e \]

- Assume we know them.

\[ y \sim N(\mu + \beta X, V) \quad V = \sigma_g^2 K + \sigma_e^2 I \]

- But we know only how to estimate \( \beta \) when

\[ y \sim N(\mu + \beta X, \sigma_e^2 I) \]

- We transform \( y \) and \( X \) to get the equation into the form where we can estimate \( \beta \) by multiplying by \( V^{-\frac{1}{2}} \)

\[ V^{-\frac{1}{2}} y \sim N(V^{-\frac{1}{2}} \mu + \beta V^{-\frac{1}{2}} X, V^{-\frac{1}{2}} V V^{-\frac{1}{2}}) = N(V^{-\frac{1}{2}} \mu + \beta V^{-\frac{1}{2}} X, I) \]

We can then apply standard estimation!
Likelihood Estimation

- We need to estimate $\sigma_e$ and $\sigma_g$.

$$y \sim N(X\beta, V) \quad V = \sigma_g^2 K + \sigma_e^2 I$$

- We can find the maximum likelihood values.

$$l(y) = -\frac{1}{2} \left[ n \log(2\pi) + \log|V| + (y - X\beta)^T V^{-1} (y - X\beta) \right]$$

- This is computationally difficult. There are several efficient algorithms.
EMMA (Efficient Mixed Model Association)

- Key Idea
  - Convert all the matrix operations to vector operations using one clever use of eigendecomposition

- Computational Efficiency
  - Thousands times faster than TASSEL/SAS implementation (C version)
  - $O(n)$ per iteration – more accurate and reliable optimization is possible
Likelihood computation takes $O(n^3)$

\[ y \sim N(X\beta, V) \]

\[ V = \sigma^2_g K + \sigma^2_e I \]

\[ l(y) = -\frac{1}{2} \left[ n \log(2\pi) + \log|V| + (y - X\beta)^T V^{-1} (y - X\beta) \right] \]
O(n) if we had eigendecomposition

\[ V = UDU^T \]

\[ z = U^T (y - X\beta) \]

\[ l(y) = -\frac{1}{2} \left[ n\log(2\pi) + \text{Tr}(D) + z^T D^{-1/2} z \right] \]
Eigenvectors are invariant

\[ K = UDU^T \quad \text{\color{blue}{O(n^3)}: needs to compute only once} \]

\[ V = U(\sigma_g^2 D + \sigma_e^2 I)U^T \]

\[ z = U^T (y - X\beta) \quad \text{\color{blue}{O(n^2)}: needs to compute only once} \]

\[ l(y) = -\frac{1}{2} \left[ n \log(2\pi) + \sigma_g^2 \text{Tr}(D) + n\sigma_e^2 + z^T \left( \sigma_g^2 D + \sigma_e^2 I \right)^{-1/2} z \right] \]

\[ \text{\color{blue}{O(n)}: needs to compute at each iteration} \]
EMMA with body weight association mapping
EMMA with liver weight association mapping

**Conventional test**

**EMMA**

*QTLs for Lvrq1 (liver weight), Orgwq2 (organ weight), Splq1 (spleen weight), Hrtq1 (heart weight), Lbm1 (lean body mass)*

EMMA in Model Organisms

- mouse (Bennet et al., 2010),
- Arabidopsis (Atwell et al., 2010)
- C. elegans (Rockman and Kruglyak, 2009; Palopoli et al., 2008)
- dogs (Boyko et al., 2010)
- Barley (Cockram et al., 2010)
- grape (Myles et al., 2011)
- rice (Famaso et al., 2011),
- potato (Stich and Gebhardt, 2011)
- malaria (Van Tyne et al., 2011) and
- mosquitos (Dorn et al., 2010)
EMMA in Mouse GWAS


<table>
<thead>
<tr>
<th>Strategy</th>
<th>Requires genotyping?</th>
<th>Requires breeding?</th>
<th>Genetic diversity?</th>
<th>Genome-wide power</th>
<th>Resolution</th>
<th>Refs*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classic cross</td>
<td>Yes</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>Low</td>
<td>50</td>
</tr>
<tr>
<td>Inbred strain association</td>
<td>No</td>
<td>No</td>
<td>No*</td>
<td>No</td>
<td>High</td>
<td>22</td>
</tr>
<tr>
<td>Hybrid Mouse Diversity Panel</td>
<td>No</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>16</td>
</tr>
<tr>
<td>Collaborative Cross</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Yes</td>
<td>Medium</td>
<td>15</td>
</tr>
<tr>
<td>Heterogeneous stock</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>High</td>
<td>38</td>
</tr>
<tr>
<td>Commercial outbred stock</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
<td>Yes</td>
<td>Very high</td>
<td>40</td>
</tr>
</tbody>
</table>
Mouse GWAS

a. Obtain and phenotype mouse population

b. Obtain genotypes for strains

<table>
<thead>
<tr>
<th>Strain</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNP3</th>
<th>SNP4</th>
<th>SNP5</th>
<th>SNP6</th>
<th>SNP7</th>
<th>SNP8</th>
<th>SNP9</th>
<th>SNP10</th>
</tr>
</thead>
<tbody>
<tr>
<td>B6</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>C3H</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>DBA</td>
<td>A</td>
<td>C</td>
<td>C</td>
<td>G</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>T</td>
</tr>
<tr>
<td>129S1</td>
<td>A</td>
<td>G</td>
<td>C</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>T</td>
<td>G</td>
<td>C</td>
<td>T</td>
</tr>
<tr>
<td>CAST</td>
<td>T</td>
<td>G</td>
<td>T</td>
<td>C</td>
<td>A</td>
<td>C</td>
<td>A</td>
<td>A</td>
<td>T</td>
<td>G</td>
</tr>
</tbody>
</table>

c. Carry out association

HMDP panel HDL cholesterol association

---

d. Examine implicated regions to identify candidate genes

Chr1: 173150000 - 173200000

Genes:
- Nr1i3
- Tomm40l
- Apoa2
- Fcer1g
- Adams4
- B4alt3
- Ndufs2
- Ppox
- Ufc1
- Usp21

SNPs:

- Position on Chr 1 (Mb)
- $-\log_{10}(P$-value)
Hybrid mouse diversity panel (HMDP)

- ~30 classical inbred strains
  - ~1,000,000 years ago
  - ~10,000 years ago
  - ~1,000 years ago
  - ~100 years ago

- ~70 recombinant inbred strains
  - Combined RI panels with CI

- Fine mapping resolution
- Much higher power
Power of hybrid mouse diversity panel

 causal SNPs are assumed to be polymorphic between BXA, BXH, BXD

- 30 CI: h²_{SNP}=0.1, h²_{background}=0.3, 5 replicates per strain
- 12 BXH
- 26 AXB
- 30 BXD
- HMDP

Statistical Power

- 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9
- 30 CI 12 BXH 26 AXB 30 BXD HMDP

Confounding effects in association studies
Proof of concept study

* 100 inbred mouse strains phenotyped
* mapped phenotypes using 140,000 mouse HapMap SNPs

- Brian Bennett
- Luz Orozco
- Charles Farber
- Anatole Ghazalpour
- Jaijam Suwanwela
- Peng-Zi Wen
- Xuping Wang
- Greg Gale
Association results in dramatically increased mapping resolution as compared to linkage

40 Mb
(~400 genes)

173 mb
# 17 HDL studies for meta analysis
(~5000 animals)

| Study ID          | Strains                        | Conditions                               | Age  | Sex | # Strains | # Samples | # Sig Loci | Ref |
|-------------------|--------------------------------|------------------------------------------|------|     |           |           |           |     |
| HMDPxB-chow(M)    | (HMDP x BL/6) F1              | Leiden/CETP TG, chow diet               | 8    |     |           |           | 1         | U   |
| HMDPxB-chow(F)    | (HMDP x BL/6) F1              | Leiden/CETP TG, chow diet               | 8    |     |           |           | 0         | U   |
| HMDPxB-ath(M)     | (HMDP x BL/6) F1              | Leiden/CETP TG, highfat diet            | 24   |     |           |           | 0         | U   |
| HMDPxB-ath(F)     | (HMDP x BL/6) F1              | Leiden/CETP TG, highfat diet            | 24   |     |           |           | 3         | U   |
| HMDP-chow(M)      | HMDP                           | chow diet                               | 12   | M   | 111       | 749       | 6         | [2] |
| HMDP-fat(M)       | HMDP                           | highfat diet                            | 16   | M   | 106       | 586       | 0         | [53]|
| HMDP-fat(F)       | HMDP                           | highfat diet                            | 16   | F   | 92        | 475       | 0         | [53]|
| BxD-db-12(M)      | (DBA x BL/6) F2               | BXD db/db, chow diet                    | 12   | M   | 125       | 125       | 0         | [8] |
| BxD-db-12(F)      | (DBA x BL/6) F2               | BXD db/db, chow diet                    | 12   | F   | 122       | 122       | 0         | [8] |
| BxD-db-5(M)       | (DBA x BL/6) F2               | BXD db/db, chow diet                    | 5    | M   | 109       | 109       | 1         | [8] |
| BxD-db-5(F)       | (DBA x BL/6) F2               | BXD db/db, chow diet                    | 5    | F   | 139       | 139       | 0         | [8] |
| BxH-apoe(M)       | (C3H x BL/6) F2               | BXH Apoe -/-                            | 24   | M   | 161       | 161       | 0         | [68]|
| BxH-apoe(F)       | (C3H x BL/6) F2               | BXH Apoe -/-                            | 24   | F   | 174       | 174       | 0         | [68]|
| BxH-wt(M)         | (C3H x BL/6) F2               | BXH wildtype, highfat diet              | 20   | M   | 164       | 164       | 0         | [66]|
| BxH-wt(F)         | (C3H x BL/6) F2               | BXH wildtype, highfat diet              | 20   | F   | 144       | 144       | 0         | [66]|
| CxB-ldlr(M)       | (BALB/cJ x BL/6) F2           | CXB LDLR -/-, highfat diet              | 12   | M   | 124       | 124       | 0         | U   |
| CxB-ldlr(F)       | (BALB/cJ x BL/6) F2           | CXB LDLR -/-, highfat diet              | 12   | F   | 64        | 64        | 0         | U   |
26 significant loci identified

Kang et al., PLoS Genetics, 2014
<table>
<thead>
<tr>
<th>SNP Location</th>
<th>Meta GxE P (Male)</th>
<th>Meta GxE P (Female)</th>
<th>Meta GxE P (Male+Female)</th>
<th># of Studies w/ Significant P</th>
<th>HE Meta P (Male+Female)</th>
<th># Studies E/A/N</th>
<th>Genes in Region</th>
<th>Gene Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chr1:64752822 (rs31078051)</td>
<td>$1.12 \times 10^{-6}$</td>
<td>$2.37 \times 10^{-3}$</td>
<td>$9.82 \times 10^{-8}$</td>
<td>0</td>
<td>$3.82 \times 10^{-3}$</td>
<td>2/14/1</td>
<td>Pikfyve</td>
<td>[62]</td>
</tr>
<tr>
<td>Chr1:107271282 (rs32203839)</td>
<td>$6.69 \times 10^{-4}$</td>
<td>$2.66 \times 10^{-4}$</td>
<td>$7.67 \times 10^{-7}$</td>
<td>0</td>
<td>$7.55 \times 10^{-2}$</td>
<td>6/0/11</td>
<td>Bcl2</td>
<td>[28]</td>
</tr>
<tr>
<td>Chr1:17199523 (rs32075748)</td>
<td>$3.45 \times 10^{-16}$</td>
<td>$1.39 \times 10^{-7}$</td>
<td>$4.41 \times 10^{-22}$</td>
<td>1</td>
<td>$5.80 \times 10^{-5}$</td>
<td>8/6/3</td>
<td>Apoa2</td>
<td>[55]</td>
</tr>
<tr>
<td>Chr2:77837584 (rs6273567)</td>
<td>$1.55 \times 10^{-4}$</td>
<td>$7.17 \times 10^{-4}$</td>
<td>$5.74 \times 10^{-7}$</td>
<td>0</td>
<td>$2.25 \times 10^{-2}$</td>
<td>3/14/0</td>
<td>Agps</td>
<td>[41]</td>
</tr>
<tr>
<td>Chr2:134421733 (rs27328693)</td>
<td>$7.69 \times 10^{-6}$</td>
<td>$1.66 \times 10^{-3}$</td>
<td>$1.09 \times 10^{-7}$</td>
<td>0</td>
<td>$1.63 \times 10^{-2}$</td>
<td>5/12/0</td>
<td>Jag1</td>
<td>[30]</td>
</tr>
<tr>
<td>Chr3:32944259 (rs29869794)</td>
<td>$2.97 \times 10^{-6}$</td>
<td>$2.49 \times 10^{-2}$</td>
<td>$3.68 \times 10^{-7}$</td>
<td>0</td>
<td>$2.36 \times 10^{-3}$</td>
<td>3/8/6</td>
<td>Prkci</td>
<td>[14]</td>
</tr>
<tr>
<td>Chr3:76066632 (rs31487078)</td>
<td>$5.89 \times 10^{-3}$</td>
<td>$2.29 \times 10^{-5}$</td>
<td>$5.03 \times 10^{-7}$</td>
<td>0</td>
<td>$6.56 \times 10^{-2}$</td>
<td>4/13/0</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr3:107430396 (rs30013147)</td>
<td>$9.59 \times 10^{-6}$</td>
<td>$3.84 \times 10^{-5}$</td>
<td>$1.56 \times 10^{-9}$</td>
<td>0</td>
<td>$7.74 \times 10^{-2}$</td>
<td>7/10/0</td>
<td>Csfl</td>
<td>[57]</td>
</tr>
<tr>
<td>Chr3:143466942 (rs30206761)</td>
<td>$1.82 \times 10^{-3}$</td>
<td>$3.97 \times 10^{-5}$</td>
<td>$3.34 \times 10^{-7}$</td>
<td>0</td>
<td>$8.60 \times 10^{-2}$</td>
<td>7/10/0</td>
<td>Hs2st1</td>
<td>[60]</td>
</tr>
<tr>
<td>Chr4:131925523 (rs32595861)</td>
<td>$1.72 \times 10^{-4}$</td>
<td>$2.84 \times 10^{-4}$</td>
<td>$1.41 \times 10^{-7}$</td>
<td>0</td>
<td>$8.65 \times 10^{-4}$</td>
<td>6/8/3</td>
<td>Fabp3</td>
<td>[47]</td>
</tr>
<tr>
<td>Chr5:119034507 (rs33131194)</td>
<td>$1.23 \times 10^{-4}$</td>
<td>$2.59 \times 10^{-3}$</td>
<td>$9.00 \times 10^{-7}$</td>
<td>0</td>
<td>$3.94 \times 10^{-1}$</td>
<td>9/8/0</td>
<td>Nos1</td>
<td>[39, 49]</td>
</tr>
<tr>
<td>Chr8:46903188 (rs33272858)</td>
<td>$1.47 \times 10^{-7}$</td>
<td>$6.52 \times 10^{-1}$</td>
<td>$1.66 \times 10^{-6}$</td>
<td>1</td>
<td>$1.60 \times 10^{-4}$</td>
<td>2/11/4</td>
<td>Acs1</td>
<td>[40]</td>
</tr>
<tr>
<td>Chr8:84150094 (rs31750594)</td>
<td>$1.96 \times 10^{-7}$</td>
<td>$1.89 \times 10^{-4}$</td>
<td>$1.33 \times 10^{-10}$</td>
<td>0</td>
<td>$8.34 \times 10^{-1}$</td>
<td>11/6/0</td>
<td>Cpe</td>
<td>[50]</td>
</tr>
<tr>
<td>Chr8:84073148 (rs33435859)</td>
<td>$1.95 \times 10^{-8}$</td>
<td>$4.53 \times 10^{-4}$</td>
<td>$4.94 \times 10^{-11}$</td>
<td>0</td>
<td>$8.33 \times 10^{-1}$</td>
<td>12/5/0</td>
<td>Prkaca</td>
<td>[51]</td>
</tr>
<tr>
<td>Chr9:101972687 (rs6333310)</td>
<td>$1.22 \times 10^{-4}$</td>
<td>$1.22 \times 10^{-5}$</td>
<td>$4.05 \times 10^{-9}$</td>
<td>0</td>
<td>$1.98 \times 10^{-8}$</td>
<td>2/1/4</td>
<td>Pik3cb</td>
<td>[5]</td>
</tr>
<tr>
<td>Chr10:21399819 (rs29363941)</td>
<td>$9.07 \times 10^{-4}$</td>
<td>$3.64 \times 10^{-7}$</td>
<td>$3.36 \times 10^{-9}$</td>
<td>0</td>
<td>$1.18 \times 10^{-2}$</td>
<td>3/12/2</td>
<td>Ifngr1</td>
<td>[23]</td>
</tr>
<tr>
<td>Chr10:90146088 (rs29370592)</td>
<td>$1.93 \times 10^{-7}$</td>
<td>0.756</td>
<td>$1.02 \times 10^{-5}$</td>
<td>1</td>
<td>$8.94 \times 10^{-4}$</td>
<td>2/14/1</td>
<td>Nr1h4</td>
<td>[37]</td>
</tr>
<tr>
<td>Chr11:69906552 (rs29477071)</td>
<td>$5.77 \times 10^{-8}$</td>
<td>$1.35 \times 10^{-5}$</td>
<td>$3.17 \times 10^{-12}$</td>
<td>0</td>
<td>$2.37 \times 10^{-9}$</td>
<td>6/9/2</td>
<td>Plscr3</td>
<td>[71]</td>
</tr>
<tr>
<td>Chr11:114083173 (rs29416888)</td>
<td>0.00011</td>
<td>$7.83 \times 10^{-5}$</td>
<td>$1.71 \times 10^{-7}$</td>
<td>0</td>
<td>$5.28 \times 10^{-5}$</td>
<td>3/13/1</td>
<td>Acox1</td>
<td>[13]</td>
</tr>
<tr>
<td>Chr14:33632464 (rs31061259)</td>
<td>$1.96 \times 10^{-8}$</td>
<td>$1.65 \times 10^{-3}$</td>
<td>$8.90 \times 10^{-7}$</td>
<td>0</td>
<td>$2.02 \times 10^{-5}$</td>
<td>3/10/4</td>
<td>Ppyr1</td>
<td>[58]</td>
</tr>
<tr>
<td>Chr15:21194226 (rs31670969)</td>
<td>$1.96 \times 10^{-8}$</td>
<td>$1.29 \times 10^{-2}$</td>
<td>$8.97 \times 10^{-10}$</td>
<td>1</td>
<td>$5.65 \times 10^{-7}$</td>
<td>3/2/12</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr15:59860191 (rs3718217)</td>
<td>$5.64 \times 10^{-6}$</td>
<td>$1.45 \times 10^{-5}$</td>
<td>$5.31 \times 10^{-10}$</td>
<td>0</td>
<td>$9.92 \times 10^{-5}$</td>
<td>5/10/2</td>
<td>Trib1, Sqle</td>
<td>[12, 16]</td>
</tr>
<tr>
<td>Chr17:46530712 (rs33259313)</td>
<td>$1.09 \times 10^{-5}$</td>
<td>0.0049</td>
<td>$3.26 \times 10^{-7}$</td>
<td>0</td>
<td>$3.53 \times 10^{-5}$</td>
<td>5/10/2</td>
<td>Gnmt</td>
<td>[43]</td>
</tr>
<tr>
<td>Chr18:82240606 (rs13483466)</td>
<td>$2.05 \times 10^{-4}$</td>
<td>$2.23 \times 10^{-4}$</td>
<td>$1.30 \times 10^{-7}$</td>
<td>0</td>
<td>$9.05 \times 10^{-1}$</td>
<td>5/12/0</td>
<td>Mbp</td>
<td>[3]</td>
</tr>
<tr>
<td>Chr19:3319089 (rs31004232)</td>
<td>$5.58 \times 10^{-2}$</td>
<td>$8.53 \times 10^{-7}$</td>
<td>$4.56 \times 10^{-7}$</td>
<td>0</td>
<td>$1.08 \times 10^{-1}$</td>
<td>3/14/0</td>
<td>Lrp5</td>
<td>[18]</td>
</tr>
<tr>
<td>ChrX:151384614 (rs31202008)</td>
<td>$2.59 \times 10^{-4}$</td>
<td>$4.72 \times 10^{-6}$</td>
<td>$8.09 \times 10^{-9}$</td>
<td>0</td>
<td>$1.22 \times 10^{-1}$</td>
<td>5/5/0</td>
<td>Htr2c</td>
<td>[36]</td>
</tr>
<tr>
<td>SNP Location</td>
<td>Male GxE P (Male)</td>
<td>Female GxE P (Male)</td>
<td>Male GxE P (Male+Female)</td>
<td>Female GxE P (Male+Female)</td>
<td># of Studies w/ Significant P</td>
<td>HE Meta P (Male+Female)</td>
<td># Studies E/A/N</td>
<td>Genes in Region</td>
</tr>
<tr>
<td>--------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>----------------------------</td>
<td>-----------------------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Chr1:64752822 (rs31078051)</td>
<td>1.12x10^-6</td>
<td>2.37x10^-3</td>
<td>9.82x10^-8</td>
<td>0</td>
<td>3.82x10^-3</td>
<td>2/14/1</td>
<td>Pikfyve</td>
<td>[62]</td>
</tr>
<tr>
<td>Chr1:107271282 (rs32203839)</td>
<td>6.69x10^-4</td>
<td>2.66x10^-4</td>
<td>7.67x10^-7</td>
<td>0</td>
<td>7.55x10^-2</td>
<td>6/0/11</td>
<td>Bcl2</td>
<td>[28]</td>
</tr>
<tr>
<td>Chr1:171195523 (rs32075748)</td>
<td>3.45x10^-16</td>
<td>1.39x10^-7</td>
<td>4.41x10^-22</td>
<td>1</td>
<td>5.80x10^-5</td>
<td>8/6/3</td>
<td>Apoa2</td>
<td>[55]</td>
</tr>
<tr>
<td>Chr2:779837584 (rs6273567)</td>
<td>1.55x10^-4</td>
<td>7.17x10^-4</td>
<td>5.74x10^-7</td>
<td>0</td>
<td>2.25x10^-2</td>
<td>3/14/0</td>
<td>Agps</td>
<td>[41]</td>
</tr>
<tr>
<td>Chr2:134421733 (rs7238693)</td>
<td>7.69x10^-6</td>
<td>1.66x10^-3</td>
<td>1.09x10^-7</td>
<td>0</td>
<td>1.63x10^-2</td>
<td>5/12/0</td>
<td>Jag1</td>
<td>[30]</td>
</tr>
<tr>
<td>Chr3:32944259 (rs29869794)</td>
<td>2.97x10^-6</td>
<td>2.49x10^-2</td>
<td>3.68x10^-7</td>
<td>0</td>
<td>2.36x10^-5</td>
<td>3/8/6</td>
<td>Prkci</td>
<td>[14]</td>
</tr>
<tr>
<td>Chr3:76066632 (rs31487078)</td>
<td>5.89x10^-3</td>
<td>2.11x10^-5</td>
<td>5.03x10^-1</td>
<td>0</td>
<td>2.40x10^-2</td>
<td>4/13/0</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr3:107430396 (rs30013147)</td>
<td>9.59x10^-6</td>
<td>1.66x10^-6</td>
<td>1.66x10^-6</td>
<td>1</td>
<td>8.60x10^-2</td>
<td>7/10/0</td>
<td>Csfl</td>
<td>[57]</td>
</tr>
<tr>
<td>Chr3:143466942 (rs30206761)</td>
<td>1.82x10^-3</td>
<td>3.97x10^-5</td>
<td>3.34x10^-7</td>
<td>0</td>
<td>8.65x10^-4</td>
<td>6/8/3</td>
<td>FABP3</td>
<td>[47]</td>
</tr>
<tr>
<td>Chr4:131925523 (rs32595861)</td>
<td>1.72x10^-4</td>
<td>2.84x10^-4</td>
<td>8.41x10^-7</td>
<td>0</td>
<td>8.65x10^-4</td>
<td>9/8/0</td>
<td>Nos1</td>
<td>[39, 49]</td>
</tr>
<tr>
<td>Chr5:119034507 (rs33131194)</td>
<td>1.23x10^-4</td>
<td>2.56x10^-10</td>
<td>1.00x10^-7</td>
<td>0</td>
<td>3.94x10^-1</td>
<td>2/11/4</td>
<td>Acs1</td>
<td>[40]</td>
</tr>
<tr>
<td>Chr8:46903188 (rs33272858)</td>
<td>1.47x10^-7</td>
<td>6.52x10^-1</td>
<td>1.66x10^-6</td>
<td>1</td>
<td>1.60x10^-4</td>
<td>11/6/0</td>
<td>Cpe</td>
<td>[50]</td>
</tr>
<tr>
<td>Chr8:64150094 (rs31750594)</td>
<td>1.96x10^-7</td>
<td>1.89x10^-4</td>
<td>1.33x10^-10</td>
<td>0</td>
<td>8.34x10^-1</td>
<td>12/5/0</td>
<td>Prkca</td>
<td>[51]</td>
</tr>
<tr>
<td>Chr8:84073148 (rs33435859)</td>
<td>1.95x10^-8</td>
<td>4.53x10^-4</td>
<td>4.94x10^-11</td>
<td>0</td>
<td>8.33x10^-1</td>
<td>2/14/0</td>
<td>Pik3cb</td>
<td>[5]</td>
</tr>
<tr>
<td>Chr9:101972687 (rs6333310)</td>
<td>1.22x10^-4</td>
<td>1.22x10^-5</td>
<td>4.05x10^-9</td>
<td>0</td>
<td>1.98x10^-8</td>
<td>3/12/2</td>
<td>Ifngr1</td>
<td>[23]</td>
</tr>
<tr>
<td>Chr10:21399819 (rs29336941)</td>
<td>9.07x10^-4</td>
<td>3.64x10^-7</td>
<td>3.36x10^-9</td>
<td>0</td>
<td>1.18x10^-2</td>
<td>2/14/1</td>
<td>Nr1h4</td>
<td>[37]</td>
</tr>
<tr>
<td>Chr10:90146088 (rs29370592)</td>
<td>1.93x10^-7</td>
<td>0.756</td>
<td>1.02x10^-5</td>
<td>1</td>
<td>8.94x10^-4</td>
<td>2/14/1</td>
<td>Plscr3</td>
<td>[71]</td>
</tr>
<tr>
<td>Chr11:69906552 (rs29477071)</td>
<td>5.77x10^-8</td>
<td>1.35x10^-5</td>
<td>3.17x10^-12</td>
<td>0</td>
<td>2.37x10^-9</td>
<td>6/9/2</td>
<td>Acox1</td>
<td>[13]</td>
</tr>
<tr>
<td>Chr11:114083173 (rs29416888)</td>
<td>0.00011</td>
<td>7.83x10^-3</td>
<td>1.71x10^-7</td>
<td>0</td>
<td>5.28x10^-5</td>
<td>3/13/1</td>
<td>Acox1</td>
<td>[13]</td>
</tr>
<tr>
<td>Chr14:33632464 (rs31061259)</td>
<td>1.96x10^-4</td>
<td>1.65x10^-3</td>
<td>8.90x10^-7</td>
<td>0</td>
<td>2.02x10^-5</td>
<td>3/10/4</td>
<td>Ppyr1</td>
<td>[58]</td>
</tr>
<tr>
<td>Chr15:21194226 (rs31670969)</td>
<td>1.96x10^-8</td>
<td>1.29x10^-2</td>
<td>8.97x10^-10</td>
<td>1</td>
<td>5.65x10^-7</td>
<td>3/12/2</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr15:59860191 (rs3718217)</td>
<td>5.64x10^-6</td>
<td>1.45x10^-5</td>
<td>5.31x10^-10</td>
<td>0</td>
<td>9.92x10^-5</td>
<td>5/10/2</td>
<td>Trib1, Sdle</td>
<td>[12, 16]</td>
</tr>
<tr>
<td>Chr17:46530712 (rs33259313)</td>
<td>1.09x10^-5</td>
<td>0.0049</td>
<td>3.26x10^-7</td>
<td>0</td>
<td>3.53x10^-5</td>
<td>5/10/2</td>
<td>Gmnt</td>
<td>[43]</td>
</tr>
<tr>
<td>Chr18:82240606 (rs13483466)</td>
<td>2.05x10^-4</td>
<td>2.23x10^-4</td>
<td>1.30x10^-7</td>
<td>0</td>
<td>9.05x10^-1</td>
<td>5/12/0</td>
<td>Mbp</td>
<td>[3]</td>
</tr>
<tr>
<td>Chr19:3319089 (rs31004232)</td>
<td>5.58x10^-2</td>
<td>8.53x10^-7</td>
<td>4.56x10^-7</td>
<td>0</td>
<td>1.08x10^-1</td>
<td>3/14/0</td>
<td>Lrp5</td>
<td>[18]</td>
</tr>
<tr>
<td>ChrX:151384614 (rs31202008)</td>
<td>2.59x10^-4</td>
<td>4.72x10^-6</td>
<td>8.09x10^-9</td>
<td>0</td>
<td>1.22x10^-1</td>
<td>5/5/0</td>
<td>Htr2c</td>
<td>[36]</td>
</tr>
<tr>
<td>SNP Location</td>
<td>Meta GxE P (Male)</td>
<td>Meta GxE P (Female)</td>
<td>Meta GxE P (Male+Female)</td>
<td># of Studies w/ Significant P</td>
<td>HE Meta P (Male+Female)</td>
<td># Studies E/A/N</td>
<td>Genes in Region</td>
<td>Gene Refs</td>
</tr>
<tr>
<td>----------------------</td>
<td>------------------</td>
<td>--------------------</td>
<td>--------------------------</td>
<td>-------------------------------</td>
<td>-------------------------</td>
<td>----------------</td>
<td>----------------</td>
<td>-----------</td>
</tr>
<tr>
<td>Chr1:64752822 (rs31078051)</td>
<td>1.12×10^-6</td>
<td>2.37×10^-3</td>
<td>9.82×10^-8</td>
<td>0</td>
<td>3.82×10^-3</td>
<td>2/14/1</td>
<td>Pikfyve</td>
<td>[62]</td>
</tr>
<tr>
<td>Chr1:107271282 (rs32203839)</td>
<td>6.69×10^-4</td>
<td>2.66×10^-4</td>
<td>7.67×10^-7</td>
<td>0</td>
<td>7.55×10^-2</td>
<td>6/0/11</td>
<td>Bcl2</td>
<td>[28]</td>
</tr>
<tr>
<td>Chr1:171199523 (rs32075748)</td>
<td>3.45×10^-16</td>
<td>1.39×10^-7</td>
<td>4.41×10^-22</td>
<td>1</td>
<td>5.80×10^-5</td>
<td>8/6/3</td>
<td>Apoa2</td>
<td>[55]</td>
</tr>
<tr>
<td>Chr2:77837584 (rs6273567)</td>
<td>1.55×10^-4</td>
<td>7.17×10^-4</td>
<td>5.74×10^-7</td>
<td>0</td>
<td>2.25×10^-2</td>
<td>3/14/0</td>
<td>Agps</td>
<td>[41]</td>
</tr>
<tr>
<td>Chr2:134421733 (rs27238693)</td>
<td>7.69×10^-6</td>
<td>1.66×10^-3</td>
<td>1.09×10^-7</td>
<td>0</td>
<td>1.63×10^-2</td>
<td>5/12/0</td>
<td>Jag1</td>
<td>[30]</td>
</tr>
<tr>
<td>Chr3:32944259 (rs29869794)</td>
<td>2.97×10^-6</td>
<td>2.49×10^-2</td>
<td>3.68×10^-7</td>
<td>0</td>
<td>2.36×10^-5</td>
<td>3/8/6</td>
<td>Prkci</td>
<td>[14]</td>
</tr>
<tr>
<td>Chr3:76066632 (rs31487078)</td>
<td>5.89×10^-3</td>
<td>2.29×10^-5</td>
<td>5.03×10^-7</td>
<td>0</td>
<td>6.56×10^-4</td>
<td>4/13/0</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr3:107430396 (rs30013147)</td>
<td>9.59×10^-6</td>
<td>3.84×10^-3</td>
<td>3.15×10^-9</td>
<td>0</td>
<td>8.7×10^-7</td>
<td>7/10/0</td>
<td>Csfl</td>
<td>[57]</td>
</tr>
<tr>
<td>Chr3:143466942 (rs30206761)</td>
<td>1.82×10^-3</td>
<td>3.97×10^-5</td>
<td>3.34×10^-7</td>
<td>0</td>
<td>8.60×10^-2</td>
<td>7/10/0</td>
<td>Hs2st1</td>
<td>[60]</td>
</tr>
<tr>
<td>Chr4:131925523 (rs32595861)</td>
<td>1.72×10^-4</td>
<td>2.84×10^-4</td>
<td>8.41×10^-7</td>
<td>0</td>
<td>8.65×10^-4</td>
<td>6/8/3</td>
<td>Fabp3</td>
<td>[47]</td>
</tr>
<tr>
<td>Chr5:119034507 (rs33131194)</td>
<td>1.23×10^-4</td>
<td>2.56×10^-7</td>
<td>8.00×10^-7</td>
<td>0</td>
<td>3.94×10^-1</td>
<td>9/8/0</td>
<td>Nos1</td>
<td>[39, 49]</td>
</tr>
<tr>
<td>Chr8:46903188 (rs33272858)</td>
<td>1.47×10^-7</td>
<td>6.52×10^-1</td>
<td>1.66×10^-6</td>
<td>1</td>
<td>1.60×10^-4</td>
<td>2/11/4</td>
<td>Acsl1</td>
<td>[40]</td>
</tr>
<tr>
<td>Chr8:64150094 (rs31750594)</td>
<td>1.96×10^-7</td>
<td>1.89×10^-4</td>
<td>1.33×10^-10</td>
<td>0</td>
<td>8.34×10^-11</td>
<td>11/6/0</td>
<td>Cpe</td>
<td>[50]</td>
</tr>
<tr>
<td>Chr8:84073148 (rs33435859)</td>
<td>1.95×10^-8</td>
<td>4.53×10^-4</td>
<td>4.94×10^-11</td>
<td>0</td>
<td>8.33×10^-1</td>
<td>12/5/0</td>
<td>Prkaca</td>
<td>[51]</td>
</tr>
<tr>
<td>Chr9:101972687 (rs63333310)</td>
<td>1.22×10^-4</td>
<td>1.22×10^-5</td>
<td>4.05×10^-9</td>
<td>0</td>
<td>1.98×10^-8</td>
<td>2/1/14</td>
<td>Pik3cb</td>
<td>[5]</td>
</tr>
<tr>
<td>Chr10:21399819 (rs29363941)</td>
<td>9.07×10^-4</td>
<td>3.64×10^-7</td>
<td>3.36×10^-9</td>
<td>0</td>
<td>1.18×10^-2</td>
<td>3/12/2</td>
<td>Ifngr1</td>
<td>[23]</td>
</tr>
<tr>
<td>Chr10:90146088 (rs29370592)</td>
<td>1.93×10^-7</td>
<td>0.756</td>
<td>1.02×10^-5</td>
<td>1</td>
<td>8.94×10^-4</td>
<td>2/14/1</td>
<td>Nr1h4</td>
<td>[37]</td>
</tr>
<tr>
<td>Chr11:69906552 (rs29477071)</td>
<td>5.77×10^-8</td>
<td>1.35×10^-5</td>
<td>3.17×10^-12</td>
<td>0</td>
<td>2.37×10^-9</td>
<td>6/9/2</td>
<td>Plscr3</td>
<td>[71]</td>
</tr>
<tr>
<td>Chr11:114083173 (rs29416888)</td>
<td>0.00011</td>
<td>7.83×10^-5</td>
<td>1.71×10^-7</td>
<td>0</td>
<td>5.28×10^-5</td>
<td>3/13/1</td>
<td>Acox1</td>
<td>[13]</td>
</tr>
<tr>
<td>Chr14:33632464 (rs31061259)</td>
<td>1.96×10^-4</td>
<td>1.65×10^-3</td>
<td>8.90×10^-7</td>
<td>0</td>
<td>2.02×10^-5</td>
<td>3/10/4</td>
<td>Ppyr1</td>
<td>[58]</td>
</tr>
<tr>
<td>Chr15:21194226 (rs31670969)</td>
<td>1.96×10^-8</td>
<td>1.29×10^-2</td>
<td>8.97×10^-10</td>
<td>1</td>
<td>5.65×10^-7</td>
<td>3/2/12</td>
<td>Novel</td>
<td>-</td>
</tr>
<tr>
<td>Chr15:59860191 (rs3718217)</td>
<td>5.64×10^-6</td>
<td>1.45×10^-5</td>
<td>5.31×10^-10</td>
<td>0</td>
<td>9.92×10^-5</td>
<td>5/10/2</td>
<td>Trib1, Sacle</td>
<td>[12, 16]</td>
</tr>
<tr>
<td>Chr17:46530712 (rs33259313)</td>
<td>1.09×10^-5</td>
<td>0.0049</td>
<td>3.26×10^-7</td>
<td>0</td>
<td>3.53×10^-5</td>
<td>5/10/2</td>
<td>Gmnt</td>
<td>[43]</td>
</tr>
<tr>
<td>Chr18:82240606 (rs13483466)</td>
<td>2.05×10^-4</td>
<td>2.23×10^-4</td>
<td>1.30×10^-7</td>
<td>0</td>
<td>9.05×10^-1</td>
<td>5/12/0</td>
<td>Mbp</td>
<td>[3]</td>
</tr>
<tr>
<td>Chr19:3319089 (rs31004232)</td>
<td>5.58×10^-2</td>
<td>8.53×10^-7</td>
<td>4.56×10^-7</td>
<td>0</td>
<td>1.08×10^-1</td>
<td>3/14/0</td>
<td>Lrp5</td>
<td>[18]</td>
</tr>
<tr>
<td>ChrX:151384614 (rs31202008)</td>
<td>2.59×10^-4</td>
<td>4.72×10^-6</td>
<td>8.09×10^-9</td>
<td>0</td>
<td>1.22×10^-1</td>
<td>5/5/0</td>
<td>Htr2c</td>
<td>[36]</td>
</tr>
</tbody>
</table>

**Female Specific Loci**
Chr1:171199523 [rs32075748] (Meta $P = 4.41 \times 10^{-22}$)

Gene: Apoa2

- Log odds ratio
- m-value

RE Summary

1. HMDPxB–chow(M)
2. HMDPxB–chow(F)
3. HMDPxB–ath(M)
4. HMDPxB–ath(F)
5. HMDP–chow(M)
6. HMDP–fat(M)
7. HMDP–fat(F)
8. BXD–db–12(M)
9. BXD–db–12(F)
10. BXD–db–5(M)
11. BXD–db–5(F)
12. BXH–apoe(M)
13. BXH–apoe(F)
14. BXH–wt(M)
15. BXH–wt(F)
16. CXB–ldlr(M)
17. CXB–ldlr(F)

(P-value)

- Study has an effect ($m > .9$)
- Study does not have an effect ($m < .1$)
- Study's effect is uncertain (.1 $< m < .9$)

Study Name

P−value

0.0886
0.1112
0.0054
0.0001
6.84 $\times 10^{-9}$
0.0029
0.0002
0.3805
0.6348
2.50 $\times 10^{-5}$
0.6062
0.0001
0.0016
0.5437
0.2249
0.1882
0.4412
Chr1:171199523 [rs32075748] (Meta $P = 4.41 \times 10^{-22}$)

Gene: Apoa2

RE Summary

<table>
<thead>
<tr>
<th>Study Name</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.HMDPxB−chow(M)</td>
<td>0.0886</td>
</tr>
<tr>
<td>2.HMDPxB−chow(F)</td>
<td>0.1112</td>
</tr>
<tr>
<td>3.HMDPxB−ath(M)</td>
<td>0.0054</td>
</tr>
<tr>
<td>4.HMDPxB−ath(F)</td>
<td>0.0001</td>
</tr>
<tr>
<td>5.HMDP−chow(M)</td>
<td>6.84 x 10^{-9}</td>
</tr>
<tr>
<td>6.HMDP−fat(M)</td>
<td>0.0029</td>
</tr>
<tr>
<td>7.HMDP−fat(F)</td>
<td>0.0002</td>
</tr>
<tr>
<td>8.BXD−db−12(M)</td>
<td>0.3805</td>
</tr>
<tr>
<td>9.BXD−db−12(F)</td>
<td>0.6348</td>
</tr>
<tr>
<td>10.BXD−db−5(M)</td>
<td>2.50 x 10^{-5}</td>
</tr>
<tr>
<td>11.BXD−db−5(F)</td>
<td>0.6062</td>
</tr>
<tr>
<td>12.BXH−apoe(M)</td>
<td>0.0001</td>
</tr>
<tr>
<td>13.BXH−apoe(F)</td>
<td>0.0016</td>
</tr>
<tr>
<td>14.BXH−wt(M)</td>
<td>0.5437</td>
</tr>
<tr>
<td>15.BXH−wt(F)</td>
<td>0.2249</td>
</tr>
<tr>
<td>16.CXB−ldlr(M)</td>
<td>0.1882</td>
</tr>
<tr>
<td>17.CXB−ldlr(F)</td>
<td>0.4412</td>
</tr>
</tbody>
</table>

**PM-Plot**

- Study has an effect ($m > .9$)
- Study does not have an effect ($m < .1$)
- Study's effect is uncertain (.1 < $m < .9$)
**Gene x Diet Interaction**
Chr1:171199523 [rs32075748] (Meta P = 4.41 x 10^{-22})

Gene: Apoa2

RE Summary

(A) Log odds ratio

(B) m-value

- Study has an effect (m > .9)
- Study does not have an effect (m < .1)
- Study's effect is uncertain (.1 < m < .9)
Chr1:171199523 [rs32075748] (Meta $P = 4.41 \times 10^{-22}$)

Gene: Apoa2

**Gene x Sex Interaction**
Chr1:171199523 [rs32075748] (Meta $P = 4.41 \times 10^{-22}$)

Gene: Apoa2

<table>
<thead>
<tr>
<th>Study Name</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.HMDPxB–chow(M)</td>
<td>0.0886</td>
</tr>
<tr>
<td>2.HMDPxB–chow(F)</td>
<td>0.1112</td>
</tr>
<tr>
<td>3.HMDPxB–ath(M)</td>
<td>0.0054</td>
</tr>
<tr>
<td>4.HMDPxB–ath(F)</td>
<td>0.0001</td>
</tr>
<tr>
<td>5.HMDP–chow(M)</td>
<td>6.84 x 10^{-9}</td>
</tr>
<tr>
<td>6.HMDP–fat(M)</td>
<td>0.0029</td>
</tr>
<tr>
<td>7.HMDP–fat(F)</td>
<td>0.0002</td>
</tr>
<tr>
<td>8.BXD–db–12(M)</td>
<td>0.3805</td>
</tr>
<tr>
<td>9.BXD–db–12(F)</td>
<td>0.6348</td>
</tr>
<tr>
<td>10.BXD–db–5(M)</td>
<td>2.50 x 10^{-5}</td>
</tr>
<tr>
<td>11.BXD–db–5(F)</td>
<td>0.6062</td>
</tr>
<tr>
<td>12.BXH–apoe(M)</td>
<td>0.0001</td>
</tr>
<tr>
<td>13.BXH–apoe(F)</td>
<td>0.0016</td>
</tr>
<tr>
<td>14.BXH–wt(M)</td>
<td>0.5437</td>
</tr>
<tr>
<td>15.BXH–wt(F)</td>
<td>0.2249</td>
</tr>
<tr>
<td>16.CXB–ldlr(M)</td>
<td>0.1882</td>
</tr>
<tr>
<td>17.CXB–ldlr(F)</td>
<td>0.4412</td>
</tr>
</tbody>
</table>

- Study has an effect ($m > .9$)
- Study does not have an effect ($m < .1$)
- Study's effect is uncertain ($0.1 < m < 0.9$)

(A) Log odds ratio

(B) $m$-value
Gene x Apoe Knockout Interaction
Chr8:86597047 (Meta $P = 4.94 \times 10^{-11}$)

Gene: *Prkaca*

<table>
<thead>
<tr>
<th>Study Name</th>
<th>$P$-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.HMDPxB–chow(M)</td>
<td>0.0060</td>
</tr>
<tr>
<td>2.HMDPxB–chow(F)</td>
<td>0.7069</td>
</tr>
<tr>
<td>3.HMDPxB–ath(M)</td>
<td>0.0049</td>
</tr>
<tr>
<td>4.HMDPxB–ath(F)</td>
<td>0.5384</td>
</tr>
<tr>
<td>5.HMDP–chow(M)</td>
<td>0.0272</td>
</tr>
<tr>
<td>6.HMDP–fat(M)</td>
<td>0.0364</td>
</tr>
<tr>
<td>7.HMDP–fat(F)</td>
<td>0.0540</td>
</tr>
<tr>
<td>8.BXD–db–12(M)</td>
<td>0.2936</td>
</tr>
<tr>
<td>9.BXD–db–12(F)</td>
<td>0.0424</td>
</tr>
<tr>
<td>10.BXD–db–5(M)</td>
<td>0.1029</td>
</tr>
<tr>
<td>11.BXD–db–5(F)</td>
<td>0.0102</td>
</tr>
<tr>
<td>12.BXH–apoe(M)</td>
<td>0.8838</td>
</tr>
<tr>
<td>13.BXH–apoe(F)</td>
<td>0.6201</td>
</tr>
<tr>
<td>14.BXH–wt(M)</td>
<td>0.0329</td>
</tr>
<tr>
<td>15.BXH–wt(F)</td>
<td>0.0707</td>
</tr>
<tr>
<td>16.CXB–ldlr(M)</td>
<td>0.0714</td>
</tr>
<tr>
<td>17.CXB–ldlr(F)</td>
<td>0.4422</td>
</tr>
</tbody>
</table>

**PM–Plot**
- **Study has an effect ($m > .9$)**
- **Study does not have an effect ($m < .1$)**
- **Study's effect is uncertain ($0.1 < m < .9$)**
Variance component model to account for sample structure in genome-wide association studies

Hyun Min Kang¹,²,⁸, Jae Hoon Sul³,⁸, Susan K Service⁴, Noah A Zaitlen⁵, Sit-yee Kong⁴, Nelson B Freimer⁴, Chiara Sabatti⁶ & Eleazar Eskin³,⁷
Complex trait analysis of gene expression uncovers polygenic and pleiotropic networks.

Relative Impact of Nucleotide and Copy Number Variation on Gene Expression Phenotypes
Barbara E. Stranger, Matthew S. Forrest, Mark Dunning, Catherine E. Ingle.

High-Resolution Mapping of Expression-QTLs Yields Insight into Human Gene Regulation

A genome-wide association study of global gene expression

Applying Gene Expression, Proteomics and Single-Nucleotide Polymorphism Analysis for Complex Trait Gene Identification
Ioannis M. Stylianou, Jason P. Affourtit, Keith R. Shockley, Robert Y. Wilpan, Fadi A. Abdi, Sanjeev Bhardwaj, Jarod Rollins, Gary A Churchill and Beverly Paigen.

Genetic Dissection of Transcriptional Regulation in Budding Yeast
Rachel B. Brem, Gaël Yvert, Rebecca Clinton, Leonid Kruglyak.

Genome-wide expression analysis in the wild and in the laboratory strain of Saccharomyces cerevisiae can be used to identify genes expressed between the wild and laboratory strain. The expression of a group of genes is correlated with growth rate and other phenotypic properties. This group of genes is enriched for genes involved in nutrient utilization, carbon utilization, and other metabolic functions. The expression of these genes is also correlated with the expression of other genes in the same pathway. This suggests that these genes may be important in the growth of the wild strain and that they may be targets for genetic selection in the laboratory strain.
eQTL mapping

- gene_1
- gene_2
- gene_m
eQTL map
eQTL map - Yeast (Brem, Kruglyak 2005)
eQTL map - Yeast (Brem, Kruglyak 2005)

- 112 Yeast segregants
- eQTL map between
  - 2,957 markers (SNP)
  - 5,700 expression levels
- Each pixel represents SNP × mRNA association
- **RED** pixels are strong associations

SNP location (DNA)

probe(gene) location (RNA)
**Cis-regulatory band**

- Strong associations between a SNP and a gene in close proximity
- More likely to be direct effect from a genetic variant
- A signature of true positives in the eQTL map
Trans-regulatory band

Confounding effects in association and eQTL studies
*Trans*-regulatory band

SNP location (DNA) → probe (gene) location (RNA)

- mRNA0
- mRNA1
- mRNA2
- mRNA3
- mRNA4
- mRNA5
- mRNA6
Trans-regulatory band
Trans-regulatory band

Confounding effects in association and eQTL studies
Leveraging biological replicates (Mouse)

**Strain 1, Rep 1**  
**Strain 1, Rep 2**  
**Strain 1, Avg**

**Strain 2, Rep 1**  
**Strain 2, Rep 2**  
**Strain 2, Avg**

**Strain 3, Rep 1**  
**Strain 3, Rep 2**  
**Strain 3, Avg**

**Strain 4, Rep 1**  
**Strain 4, Rep 2**  
**Strain 4, Avg**

**Strain n, Rep 1**  
**Strain n, Rep 2**  
**Strain n, Avg**

Genetic map

eQTL map avg
Leveraging biological replicates

- Strain 1, Rep 1
- Strain 1, Rep 2
- Strain 2, Rep 1
- Strain 2, Rep 2
- Strain 3, Rep 1
- Strain 3, Rep 2
- Strain 4, Rep 1
- Strain 4, Rep 2
- Strain n, Rep 1
- Strain n, Rep 2

Genetic map

eQTL map 1
eQTL map 2
Quantifying strength of hotspots
Hotspots are inconsistent between replicates
Random SNPs creates strong hotspots

p-value of the strongest hotspot with 1000 permutations = 0.89
Inter-sample correlation with replicates

**Brain BXD whole brain inter-sample correlation**

**HSC BXD HSC inter-sample correlation**

**Brain**

**HSC (hematopoetic stem cell)**
Inter-sample correlation with replicates

Diagonals represents correlation between replicated pairs
Lower-triangular matrix represent the pairwise correlation within replicated subset 1

brain

HSC (hematopoietic stem cell)
Inter-sample correlation with replicates

Upper-triangular matrix represent the pairwise correlation within replicated subset 2
The replicated pairs are not strongly correlated compared to different strain pairs.
A clear batch effect is observed

Higher within-batch correlations than between-batch correlations

**brain**

**HSC** (hematopoetic stem cell)
Spurious *trans*-regulatory band

Confounding effects in association and eQTL studies

SNP location (DNA)

probe(gene) location (RNA)

SNP

Systematic Confounding

mRNA1

mRNA2

mRNA3

mRNA4

mRNA5

mRNA6
Inter-sample correlation reproduces hotspots

Original Expression Data

Simulated Expression Data

Original Hotspots (Red)

Simulated Hotspots (Green)

Confounding effects in association and eQTL studies

10/19/15
Spurious *trans*-regulatory band

Confounding effects in association and eQTL studies
ICE (Inter-sample Correlation Emended) eQTL

\[ y = \mu + X\beta + u + e \]
ICE (Inter-sample Correlation Emended) eQTL

\[ y = \mu + X\beta + u + e \]
ICE eQTL map

t-test (linear model)

ICE eQTL (linear mixed model)
Concordance of eQTLs across replicates

Concordance of eQTLs between biological replicates in BXD whole brain

Proportion in common

Size of gene list

Concordance of eQTLs across replicates
Concordance of eQTLs across tissues
Do hotspots really exist?
ICE (Inter-sample Correlation Emended) eQTL

\[ y = \mu + X\beta + u + e \]
Do hotspots really exist?

Also supported by other data such as protein activity level hotspots.
But we eliminate the real hotspots...
New Idea: Use weak genes for correction

- NICE (Next-Generation ICE)
- Intuitions:
  - Confounding affects all genes
  - Genetic factors affect a subset of the genes
  - **True genetic effects will be stronger than confounding effects.**
- We correct for confounding using only weaker effects.
NICE preserves the real hotspots

true hotspots (overlap data)

NICE eQTL (linear mixed model)
eQTL Hotspot Results

- Overlap analysis identifies 10 hotspots.
- NICE identifies 9 hotspots with 2 spurious hotspots.
- ICE identifies only 4 hotspots with 2 spurious hotspots.
Other projects

- **Meta-Analysis**
  - Random Effects (Buhm Han, AJHG 2011)
  - Interpreting (Buhm Han, PLoS Genetics 2011)
  - Imputation Errors (Noah Zaitlen, GenEpi 2010)
  - Population Structure (Nick Furlotte, Genetics 2012)
  - Gene-by-Environment (Eun Yong Kang, PLoS Genetics, 2014)

- **eQTL Methods**
  - Multi-Tissue eQTLs (Jae Hoon Sul, PLoS Genetics 2013)
  - Speeding up computation (Emrah Kostem, JCB 2013, Lisa Gai, unpublished)
  - Efficient p-value computation (Jae Hoon Sul, AJHG 2015)
  - Incorporating Priors (Dat Duong, unpublished)

- **Mixed Models**
  - Longitudinal data (Nick Furlotte, Gen Epi 2012)
  - Population Structure and Selection (Jae Hoon Sul, NRG 2013)
  - Multivariate Models (Nick Furlotte, Genetics, 2015)
  - Gene-by-environment (Jae Hoon Sul, Michael Bilow, unpublished)

- **Fine Mapping** (Farhad Hormozdiari, Genetics 2014, Bioinformatics 2015)

- **RNAseq Assembly** (Serghei Mangul, Bioinformatics 2014, unpublished)
Acknowledgements

- Eskin Lab
  - Hyun Min Kang (U Michigan)
  - Chun Ye (UCSF)
  - Jong Wha (Joanne) Joo
  - Buhm Han (Asan, Korea)
  - Jae-hoon Sul (UCLA)
  - Noah Zaitlen (UCSF)

- UCLA
  - Jake A. Lusis
  - Chiara Sabatti (Stanford)

- http://zarlab.cs.ucla.edu/