Leveraging molecular QTL to understand genetic architecture of diseases & complex traits

Farhad Hormozdiari
Postdoctoral Fellow, Price Lab
Department of Epidemiology, Harvard University
ASHG 2017
Outline

① Background & Goal of Project

② Our Method

③ Results on Real Phenotypes
Outline

① Background & Goal of Project

② Our Method

③ Results on Real Phenotypes
Partitioning Heritability Help Us Understand Genetic Architecture

• SNP-Heritability ($h_g^2$)
  - Fraction of the phenotypic variance explained by SNPs
  - Height SNP-heritability estimated at ~45%

• Partition heritability

• Annotation:
  - A group of SNPs that have the same biological function
  - Binary: which SNPs belong to the annotation
  - Continuous: weight for each SNP indicates contribution to annotation

(Yang et al. 2010 Nature Genetics, Yang et al. 2011 AJHG, Kostem et al. 2013 AJHG)
(Loh et al. 2015 Nature Genetics, Finucane et al. 2015 Nature Genetics)
Significant Fraction of GWAS Risk Loci Are eQTLs (Expression Quantitative trait loci)

• Expression Analysis
  - Understand the genetic basis of regulatory elements.
  - Gene expression is a “proxy” measure for transcription/translation event.

• eQTLs
  - Genomic loci that regulate expression levels of genes.
  - Analyze associations between SNPs and gene expression levels.

(Nicolae et al. 2010 PG, Nica et al. 2010 PG)
(Giambartolomei et al. 2014 PG, He et al. 2013 AJHG, Hormozdiari et 2016 AJHG)
(Gamazon et al. 2015 NG, Gusev et al. 2016 NG)
ciseQTL Are Enriched for Trait Heritability

- Davis et al. have shown that ciseQTL in cortex and cerebellum contributes significantly to the heritability of obsessive-compulsive disorder (OCD) and Tourette Syndrome (TS) → High Enrichment
- Torres et al. have shown that ciseQTL in muscle and adipose explain higher phenotypic variance than expected given the proportion of SNPs in T2D → High Enrichment
- Can we have construct a better annotation from QTL data?
- How about other molecular QTL (hQTL, meQTL)?

(Davis et al. 2013 PG, Torres et al. 2014 AJHG)
Stratified LD Score Regression (S-LDSC) Method to Partition Heritability

GWAS summary statistics (not only top GWAS hits)

Reference panel (1000G)

Functional annotations

Annotation Enrichment

Enrichment = \frac{\text{Prop. } h^2_g}{\text{Prop. SNP}}

PGC-SCZ 2014 Nature

Roadmap Epigenomics Consortium 2015 Nature

Annotation Effect Size(\mathcal{T}^*_C )

Conditional on other annotations

(Finucane et al. 2015 NG, Gazal et al. 2017 NG)
Goal: “Best” Annotation from QTL Datasets

- Generate an annotation from QTL to capture a significant fraction of complex trait heritability?

- Caveat: Cannot detect “Causal” relationship.
Outline

① Background & Goal of Project

② Our Method

③ Results on Real Phenotypes
Existing Method to Partition Heritability Using QTL (AllciseQTL)

- Generate an annotation from **ALL** significant variants
- Same variants can be significant in multiple genes

(example table)

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>SNP1</th>
<th>SNP2</th>
<th>SNPM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$S_{11}$</td>
<td>$S_{21}$</td>
<td>$S_{M1}$</td>
</tr>
</tbody>
</table>

If $S_{ij}$ is significant, $i$ is set to 1

(Binary Annotation)

(Davis et al. 2013 PG, Torres et al. 2014 AJHG)
Existing Method to Partition Heritability Using QTL (AllciseQTL)

- Generate an annotation from **ALL** significant variants
- Same variants can be significant in multiple genes

If $S_{ij}$ is significant, $i$ is set to 1

(Davis et al. 2013 PG, Torres et al. 2014 AJHG)
Existing Method to Partition Heritability Using QTL (AllciseQTL)

- Generate an annotation from **ALL** significant variants
- Same variants can be significant in multiple genes

(Davis et al. 2013 PG, Torres et al. 2014 AJHG)
95% Credible Set to Partition Heritability

- Generate an annotation from variants in the 95% Credible set
- 95% Credible set is obtained from CAVIAR

Binary Annotation

<table>
<thead>
<tr>
<th>Gene 1</th>
<th>Gene 2</th>
<th>Gene 22000</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{11}$</td>
<td>$S_{12}$</td>
<td></td>
</tr>
<tr>
<td>SNP2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{21}$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>SNP M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$S_{M1}$</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Variant i is in 95% Credible set it is set to 1

(Hormozdiari et al. 2014 Genetics)
MaxCPP to Partition Heritability

- Causal Posterior probability (CPP) variants
- Generate a continuous annotation with value for each SNP equal to the maximum probability of the SNP being causal in ALL genes.
MaxCPP to Partition Heritability

- Causal Posterior probability (CPP) variants
- Generate a continuous annotation with value for each SNP equal to the maximum probability of the SNP being causal in **ALL** genes.

![Diagram showing computation of CPP for each SNP]
MaxCPP to Partition Heritability

- Causal Posterior probability (CPP) variants
- Generate a continuous annotation with value for each SNP equal to the maximum probability of the SNP being causal in **ALL** genes.

<table>
<thead>
<tr>
<th></th>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNP1</td>
<td>S_{11}</td>
<td>S_{12}</td>
</tr>
<tr>
<td>SNP2</td>
<td>S_{21}</td>
<td>S_{22}</td>
</tr>
<tr>
<td>SNP M</td>
<td>S_{M1}</td>
<td>S_{M2}</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1</td>
<td>0.9</td>
<td>0</td>
</tr>
<tr>
<td>0.8</td>
<td>0.1</td>
<td>0</td>
</tr>
</tbody>
</table>

**Continuous Annotation**

Max

<table>
<thead>
<tr>
<th></th>
<th>Gene 1</th>
<th>Gene 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.1</td>
<td>0.2</td>
<td></td>
</tr>
<tr>
<td>0.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

0.2
Outline

① Background & Goal of Project

② Our Method

③ Results on Real Phenotypes
Summary of Datasets Analyzed

• QTL datasets
  – GTEx v6p: eQTL for 44 tissues
  – BLUEPRINT: eQTL, 2 hQTL (H3K27ac and H3K4me1), meQTL, and sQTL
• We utilize summary statistics
  – 41 Independent traits and diseases (N=320K).
    • Loh et al. bioRxiv; BOLT-LMM UK Biobank summary statistics are publicly available.
    • Meta(τ*) is meta-analyzed over 41 traits.
  – Results are obtained by conditioning on the baseline LD

(Gazal et al. 2017 NG, Loh et al. bioRxiv)
MaxCPP Outperforms Existing QTL Annotations in GTEx Whole Blood

- AllciseQTL: All significant variants for every gene
- 95%CredibleSet: variants in the 95% Credible Set
- MaxCPP: maximum CPP score for each variant over ALL genes.

(GTEx Consortium 2017 Nature)
(Hormozdiari et al. 2017 bioRxiv)
Linear Relationship Between Sample Size & MaxCPP Effect Size in GTEx

\[ R^2 = 0.69 \]
\[ P = 1.36 \times 10^{-12} \]

(Hormozdiari et al. 2017 biorxiv)
GTEx Meta-Tissue Improves Results

\[ \text{Meta}(\tau^*) = 0.23 \]  
Whole Blood

\[ \text{Meta}(\tau^*) = 0.52 \]  
GTEx-Meta-Tissue

(Hormozdiari et al. 2017 biorxiv)
Cell Type Specific Signals for Blood Cell Traits

- We consider 5 blood cell traits: white blood cell count, red blood cell count, etc.
- GTEx Blood has higher enrichment than GTEx Meta-Tissue

BaselineLD + GTEx-Meta-Tissue + Whole Blood

Brain-related traits \((P=9.81e-05)\) and Autoimmune diseases \((P=9.15e-03)\) show similar tissue/cell type specific signal

(Hormozdiari et al. 2017 biorxiv)
MaxCPP Restricted to ExAC loss-of-function (LoF) Intolerant Genes is Extremely Enriched

- Investigate the effect of common variants on traits through regulatory elements
- ExAC detected 3230 LoF intolerant genes
- MaxCPP(ExAC) = MaxCPP restricted to ExAC genes
- MaxCPP(All Genes) = MaxCPP over all genes

eQTL, meQTL, and hQTL Are Jointly Enriched for Disease Heritability

- QTL enriched with disease heritability
  - GTEx-eQTL (4.88x)
  - BLUEPRINT-eQTL (3.79x)
  - BLUEPRINT-H3K4me1 (3.57x)
  - BLUEPRINT-H3K27ac (3.51x)
  - BLUEPRINT-meQTL (2.55x)
  - BLUEPRINT-sQTL (2.55x)
- All molecular QTL except sQTL remain significant in a joint model

(Hormozdiari et al. 2017 biorxiv)
Conclusions

- Fine-mapped eQTL are enriched with disease/trait heritability
- Cell type specific signal for blood cell type, brain-related and autoimmune traits.
- ExAC gene-set is enriched with disease/trait heritability
- eQTL, hQTL, meQTL and sQTL are individually enriched
- sQTL signal is captured by eQTL, hQTL, and meQTL in joint model
Acknowledgments

- Harvard Uni.
  - Alkes Price
  - Steven Gazal
  - Bryce van de Geijn
  - Xuanyao Liu
  - Hilary Finucane
  - Sasha Gusev
  - Po-Ro Loh
  - Armin Schoech
  - Yakir Reshef

- UCLA
  - Eleazar Eskin

Hormozdiari et al. “Leveraging molecular QTL to understand the genetic architecture of diseases and complex traits”, biorxiv.