Using the transcriptome to determine the genetic mechanisms of disease susceptibility

Joseph Powell

Centre for Neurogenetics and Systems Genomics

Winter School in Mathematical and Computational Biology

Brisbane - 7th July 2014
Causes of Common Diseases - Genetics + Environment

Genetics

Environmental factors
How Do Mutations Influence Disease Susceptibility?

Most genetic variation underlying disease susceptibility is located in regulatory regions.
Move To Molecular Phenotypes - Systems Genetics

Mechanisms of disease susceptibility
By what means do most DNA variants alter cellular behaviour and ultimately contribute to differences in disease susceptibility?
Genome-wide epistasis

Variance heterogeneity
What is epistasis?

Definition

The effect on the phenotype caused by locus A depends on the genotype at locus B....

Epistasis has been reported in model organisms through artificial gene knockouts, artificial line crosses and hybridization.

Our aim was to systematically search for instances of epistasis amongst common variants for genetic variation that has arisen in natural populations.
Gene expression

- Transcription measured for thousands of genes
- Typically heritable
- Loci (eQTLs) commonly have very large effect sizes
- Good candidates to search for epistasis
Common Diseases

Epistasis

Variance Heterogeneity

Joseph Powell
Mechanisms of disease susceptibility
Discovery population

Brisbane Systems Genetics Study (BSGS)

- 846 healthy individuals
- 528,509 autosomal SNPs
- RNA measured for peripheral blood (Illumina HT-12v4.0)
- Expression levels for 7,339 probes
Epistasis

Initial exhaustive SNP-pair test for each expression trait

Apply a family-wise error rate of 5%

Filter results to remove artifacts

Second test for epistatic effects

Apply Bonferroni threshold

Replication in two independent populations

Functional characterization

Brisbane Systems Genetics Study (BSGS)

Fehrman EGCUT

434 Epistatic SNP pairs tested for replication

p-values < 2.5% confidence interval

High concordance of direction of epistatic effects (p=5.68e-10)
Computational analysis

- Exhaustive SNP x SNP testing for each probe
- epiGPU software and GPU clusters
- 8 d.f. F-test
- Over quadrillion tests

Initial exhaustive SNP-pair test for each expression trait
Design Of The Analysis

SNP x SNP 2 dimension analysis

Epistatic genotype by phenotype
Epistasis

- Initial exhaustive SNP-pair test for each expression trait
- Apply a family-wise error rate of 5%
- Filter results to remove artifacts
- Second test for epistatic effects
- Apply Bonferroni threshold
- Replication in two independent populations
- Functional characterization
- Brisbane Systems Genetics Study (BSGS)

Variance Heterogeneity

- Fehrman EGCUT
- 434 Epistatic SNP pairs tested for replication
- 345 replicate (p-values < 2.5% confidence interval)
- High concordance of direction of epistatic effects (p=5.16e-10)

Mechanisms of disease susceptibility
Correction for multiple testing

- Permutation and Bonferroni used to give 5% FWER
- Permutation: Single probe FWER; $T^* = 2.13 \times 10^{-12}$
- Correct for 7,339 probes
- Experiment wide FWER; $T^*/7,339 = 2.91 \times 10^{-16}$

Apply a family-wise error rate of 5%
Epistasis

Initial exhaustive SNP-pair test for each expression trait

Apply a family-wise error rate of 5%

Filter results to remove artifacts

Second test for epistatic effects

Apply Bonferroni threshold

Replication in two independent populations

Functional characterization

Brisbane Systems Genetics Study (BSGS)

Fehrmann EGCUT

434 Epistatic SNP pairs tested for replication
345 replicate (p-values < 2.5% confidence interval)
High concordance of direction of epistatic effects (p=5.56x10^-7)

Joseph Powell
Mechanisms of disease susceptibility
Removing artifacts

Filter results to remove artifacts

- For SNP pairs that passed the $2.91 \times 10^{-16}$ threshold
- All 9 genotypes classes present
- Minimum class size of 5
- No LD between SNP pairs ($r^2 < 0.1$ and $D' < 0.1$)
- No single loci additive or dominance effects
- 11,155 pairs carried forward
Epistasis

Initial exhaustive SNP-pair test for each expression trait

Apply a family-wise error rate of 5%

Filter results to remove artifacts

Second test for epistatic effects

Apply Bonferroni threshold

Replication in two independent populations

Functional characterization

Brisbane Systems Genetics Study (BSGS)

Fehrmann EGCUT

434 Epistatic SNP pairs tested for replication

435 replicate (p-values < 2.5% confidence interval)

High concordance of direction of epistatic effects (p=5.56x10^-10)

Joseph Powell
Mechanisms of disease susceptibility
Testing for non-additive effects

- Nested ANOVA for 11,155 pairs
- Contrasts full genetic (8 d.f.) vs marginal effects (4 d.f.)
- Thus, testing for the contribution of epistatic variance
Epistasis

- Initial exhaustive SNP-pair test for each expression trait
- Apply a family-wise error rate of 5%
- Filter results to remove artifacts
- Second test for epistatic effects
- Apply Bonferroni threshold

Replication in two independent populations

Functional characterization

Brisbane Systems Genetics Study (BSGS)

Fehrmann EGCUT

434 Epistatic SNP pairs tested for replication
345 replicate
(p-values < 2.5% confidence interval)
High concordance of direction of epistatic effects (p=5.56x10^-12)

Common Diseases

Variance Heterogeneity

Mechanisms of disease susceptibility

Joseph Powell
More correction for multiple testing

Epistatic effects significant at $p < 0.05/11, 155 = 4.48 \times 10^{-6}$

501 SNP pairs with significant interaction terms
Epistasis

- Initial exhaustive SNP-pair test for each expression trait
- Apply a family-wise error rate of 5%
- Filter results to remove artifacts
- Second test for epistatic effects
- Apply Bonferroni threshold
- Replication in two independent populations
  - Fehrman EGCUT
    - 434 Epistatic SNP pairs tested for replication
    - 345 replicate (p-values < 2.5% confidence interval)
    - High concordance of direction of epistatic effects (p=5.16x10^-7)
- Functional characterization

Variance Heterogeneity
Replication population

Fehrmann
- 1,240 individuals (Netherlands)

EGCUT
- 891 individuals (Estonian)
- RNA measured for peripheral blood (Illumina HT-12v3.0)
- Genotyped with Illumina arrays
**Epistasis**

1. Initial exhaustive SNP-pair test for each expression trait
2. Apply a family-wise error rate of 5%
3. Filter results to remove artifacts
4. Second test for epistatic effects
5. Apply Bonferroni threshold
6. Replication in two independent populations
7. Functional characterization

**Brisbane Systems Genetics Study (BSGS)**

**Fehrmann EGCUT**

454 Epistatic SNP pairs tested for replication
345 replicate (p-values < 2.5% confidence interval)
High concordance of direction of epistatic effects (p=5.46x10^{-10})

**Common Diseases**

**Variance Heterogeneity**

**Mechanisms of disease susceptibility**

Joseph Powell
Replication tests

- Replication of 501 epistatic hits using two independent cohorts
- Same filtering criteria
- 434 pairs which could be tested in both cohorts
Epistasis

- Initial exhaustive SNP-pair test for each expression trait
  - Apply a family-wise error rate of 5%
  - Filter results to remove artifacts
  - Second test for epistatic effects
  - Apply Bonferroni threshold
  - Replication in two independent populations

- Brisbane Systems Genetics Study (BSGS)

Variance Heterogeneity

- Fehrmann EGCUT
  - 434 Epistatic SNP pairs tested for replication
  - 345 replicate (p-values < 2.5% confidence interval)
  - High concordance of direction of epistatic effects (p=5.56x10^-7)

Joseph Powell
Mechanisms of disease susceptibility
Functional work

Analyses to elucidate possible functional mechanisms

- Non-synonymous mutations
- GWAS and known eQTL overlap
- Genome segmentation
- Chromosome interactions
- Tissue specific transcription regions
The top panel shows all 434 discovery SNPs that were tested for interactions.

The bottom panel shows the same data as the top panel but excluding the 30 most significant interactions.

Matched direction of epistatic effects $p = 5.56 \times 10^{-31}$
Interactive Figure: http://kn3in.github.io/detecting_epi/

Yellow dots: A Transcript

Grey dots: Epistatic SNPs
Example: Epistatic Control Of CAST Regulation

CAST: Associated with numerous rheumatic diseases

It’s regulation controlled by many epistatic SNPs throughout the genome

Novel genetic control of regulation potentially underlying disease susceptibility
Chromosome interactions identified in K562 cell lines using Hi-C and ChIP-seq
Red lines = N SNPs within 20kb, 500kb, 2Mb and 10Mb of known interacting regions
Histograms = null distributions
Conclusions

The first evidence for multiple instances of segregating common polymorphisms interacting to influence human traits.

Novel computational and statistical frameworks can identify and replicate epistasis.

New knowledge of the control of transcription of many genes implicated in common disease.
Searching for veQTL

There is evidence across several species for genetic control of phenotypic variation of complex traits.

This means that the variance of a phenotype is genotype dependent.

Understanding genetic control of variability is important in evolutionary biology, and medicine.
What causes them?

- Epistatic interactions
- Gene by Environment Interactions
- Other ..... ?

*Figure 1 Relation between a vQTL and an epistatic interaction.* Panel (a) plots phenotype values in arbitrary units for a population of 500 outbred individuals stratified by genotype at a hypothetical vQTL. Panel (b) shows how the pattern in (a) could have arisen through a simple (mean-controlling) epistatic interaction with a second locus, possibly on another chromosome, that segregates two genotypes (black and gray).
Brisbane Systems Genetics Study (BSGS)

- 846 healthy individuals
- 6,011,184 imputed (autosomal) SNPs
- RNA measured for peripheral blood (Illumina HT-12v4.0)
- Expression levels for 17,994 probes
Double Generalized Linear Model

\[ y = \mu + xb + e \]

\[ e \sim N(0, \sigma^2_e) \]

Model to detect variance and mean effects

\[ e \sim N(0, \sigma^2_{ei}); \log(\sigma^2_e) = c + vx \]

\[ y \sim N(\mu + xb, \exp(c + vx)) \]

Initial exhaustive test for each SNP x expression using DGLM
For statistical analysis, the informational content is the same if data is transformed by a monotonic function.

Provided the three means are different, there is always a monotonic transformation that will tend to equalize the three variances.

However, it’s relevance depends on the biological scale.

Test for effects of re-scaling
Original and normalised scales

![Graph showing original and normalised mRNA expression distributions.](image-url)
Initial scale = Z-score from inverse-normal transformation

- Adjust original scale using transformation functions provided in Sun et al. (2013)
  \[ \sigma^2(x) = f(m) = am^2 + bm + c \]  

- Express variance of \( x \) as a function of its mean \( m_x \), \( f(m_x) \)
  \[ y = g(x) \propto \int \frac{dx}{\sqrt{f(x)}} \]  

- Re-test using dgIm
- The aim is to remove main driven variance effects

Pictorial representation of four different non-linear changes to the scale. They represent four classes of monotonic transformation (Sun et al AJHG 2013).
Discovery results

Cis - $2.55 \times 10^{-7}$

- 747 veQTL (Potentially have some main effects)
- Only 88 have $P_{main} > 0.01$
- Transformations remove the effects of 655/659 (with main effects)
- Transformations remove the effects of 1/87 (without main effects)

Trans - $3.39 \times 10^{-12}$

- 1412 veQTL (Potentially have some main effects)
- 619 have $P_{main} > 0.01$
- Transformations remove the effects of 732/793 (with main effects)
- Transformations remove the effects of 21/619 (without main effects)
Fehrmann
- 1,240 individuals (Netherlands)

EGCUT
- 891 individuals (Estonian)
- RNA measured for peripheral blood (Illumina HT-12v3.0)
- Genotyped with Illumina arrays
Replication in two independent populations

- Replication of 7,768 veQTL hits using two independent cohorts
- DGLM
- Also Bartlett’s and Levene’s tests
Replication Meta-analysis

Just interested in the variance only veQTL

- At $FDR = 0.05$
- 89% of cis-veQTL
- 78% of trans-veQTL
- Same effect direction;
- 85% cis and 77% trans
Ripple effect across a network

Predicted TF binding sites tagged by expression probes. Blue lines have variance heterogeneity with rs3821224 ($p < 0.05$)
Expression levels for a probe in AHSA2 by genotype class of rs3821224
## GWAS - veSNP overlap

<table>
<thead>
<tr>
<th>Disease / Trait</th>
<th>N veSNPs in GWAS catalogue</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asthma</td>
<td>1</td>
</tr>
<tr>
<td>Autism spectrum disorder</td>
<td>2</td>
</tr>
<tr>
<td>Bipolar disorder</td>
<td>2</td>
</tr>
<tr>
<td>Cholesterol, total</td>
<td>3</td>
</tr>
<tr>
<td>Endometriosis</td>
<td>1</td>
</tr>
<tr>
<td>HDL cholesterol</td>
<td>2</td>
</tr>
<tr>
<td>Height</td>
<td>4</td>
</tr>
<tr>
<td>Inflammatory bowel disease</td>
<td>2</td>
</tr>
<tr>
<td>LDL cholesterol</td>
<td>6</td>
</tr>
<tr>
<td>N-glycan levels</td>
<td>4</td>
</tr>
<tr>
<td>Obesity-related traits</td>
<td>5</td>
</tr>
<tr>
<td>Parkinson’s disease</td>
<td>6</td>
</tr>
<tr>
<td>Rheumatoid arthritis</td>
<td>3</td>
</tr>
<tr>
<td>Schizophrenia</td>
<td>8</td>
</tr>
<tr>
<td>Sudden cardiac arrest</td>
<td>2</td>
</tr>
<tr>
<td>Systemic lupus erythematosus</td>
<td>3</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>7</td>
</tr>
<tr>
<td>Type 1 diabetes</td>
<td>3</td>
</tr>
<tr>
<td>Urate levels</td>
<td>2</td>
</tr>
</tbody>
</table>
Conclusions

Computational and statistical frameworks can identify and replicate variance heterogeneity for RNA transcripts.

Variance (only) eQTL represent another form of genetic control for phenotypes.

Functional characterization of variance loci suggests a number of possible mechanisms that underlie variance heterogeneity.
Acknowledgements

Centre for Neurogenetics and Systems Genomics
- Gib Hemani
- Kostya Shakhbazov
- Jian Yang
- Allan Mcrae
- Peter Visscher

University of Groningen
- Harm-Jan Westra
- Lude Franke

Dalama University
- Lars Ronnegard

QIMR Berghofer
- Nick Martin
- Anjali Henders
- Grant Montgomery

Georgia Institute of Technology
- Greg Gibson

University of Tartu
- Andres Metspalu
- Tonu Esko

Visit us at www.complextraitgenomics.com