Detection and replication of epistasis influencing human transcription

Gibran Hemani Konstantin Shakhbazov
Grant W Montgomery Peter M Visscher Joseph E Powell

Queensland Brain Institute, University of Queensland

University of Queensland Diamantina Institute

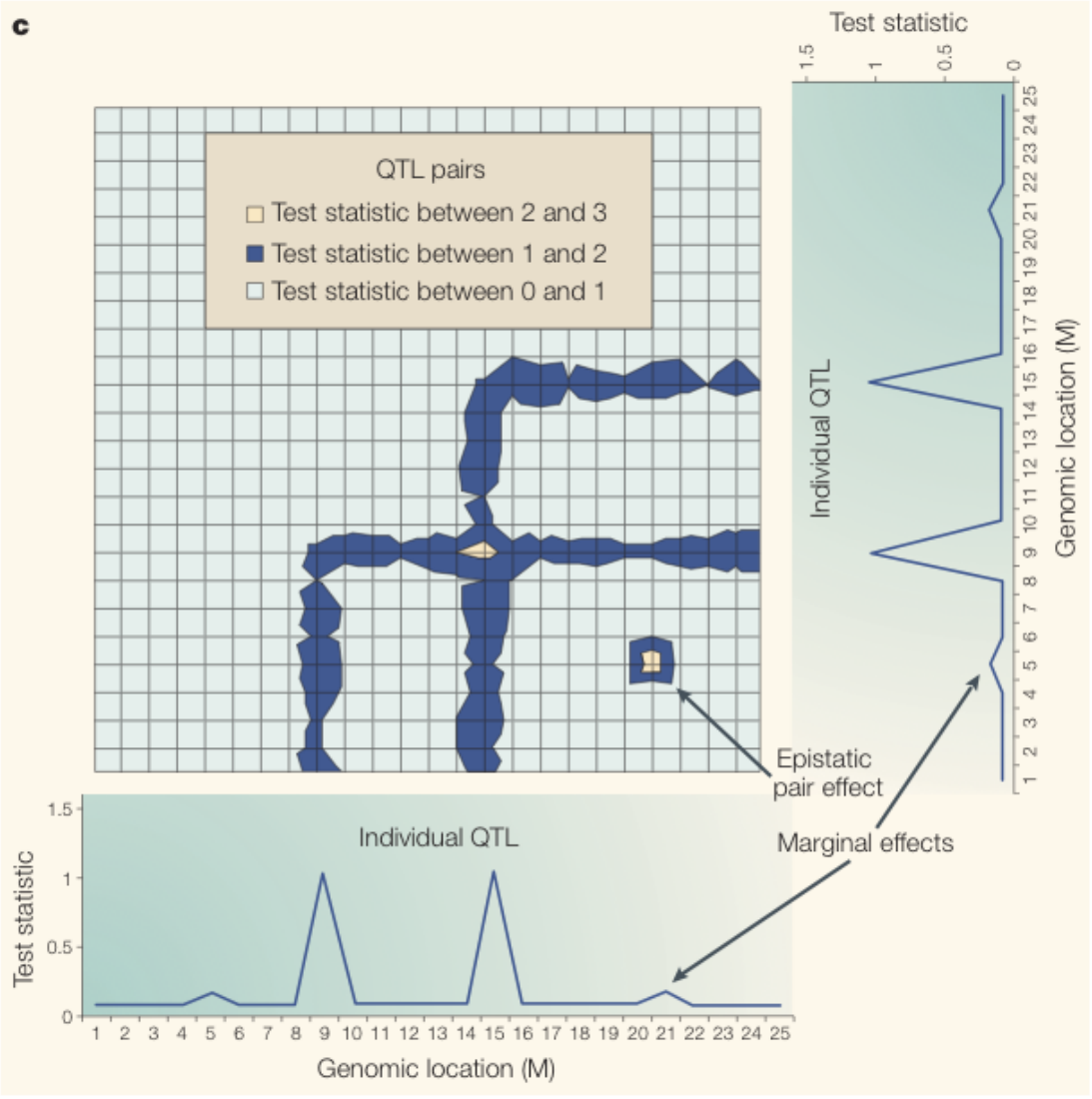
Outline

Epistasis

Definition

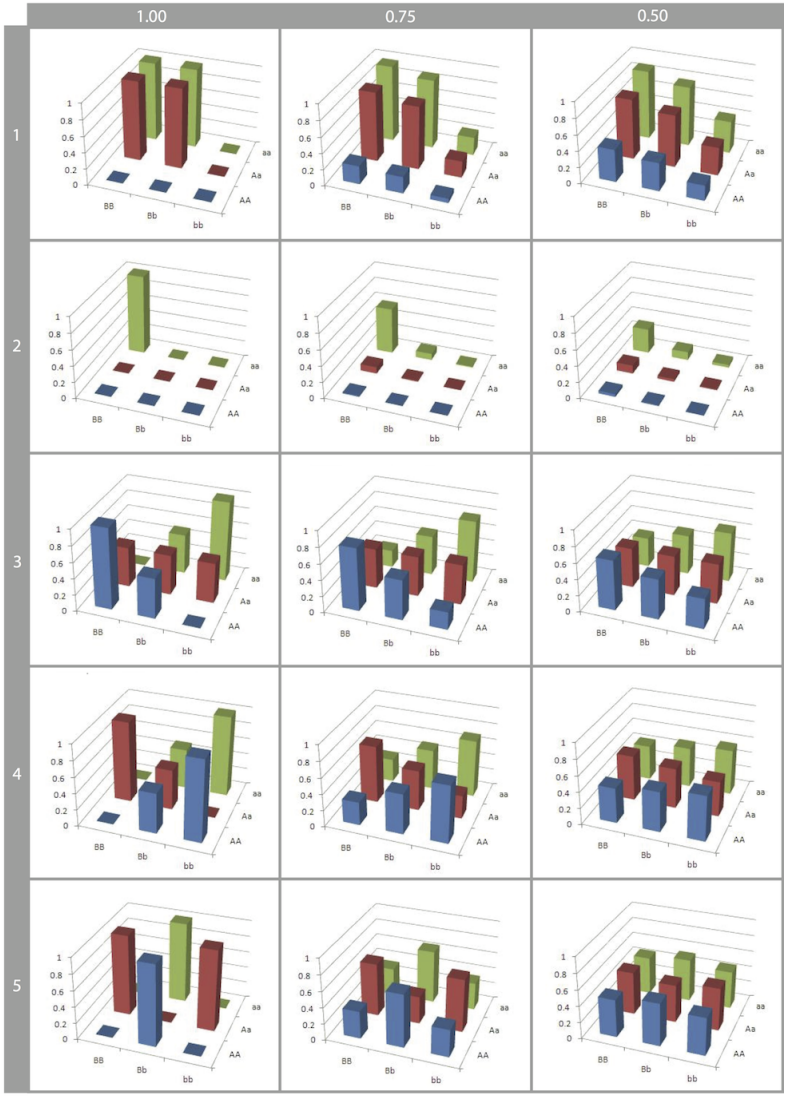
The effect on the phenotype caused by locus A depends on the genotype at locus B

Two dimensional GWAS



Carlborg 2004

Impact of LD on detecting epistasis



Multiple testing problem

Curse of dimensionality

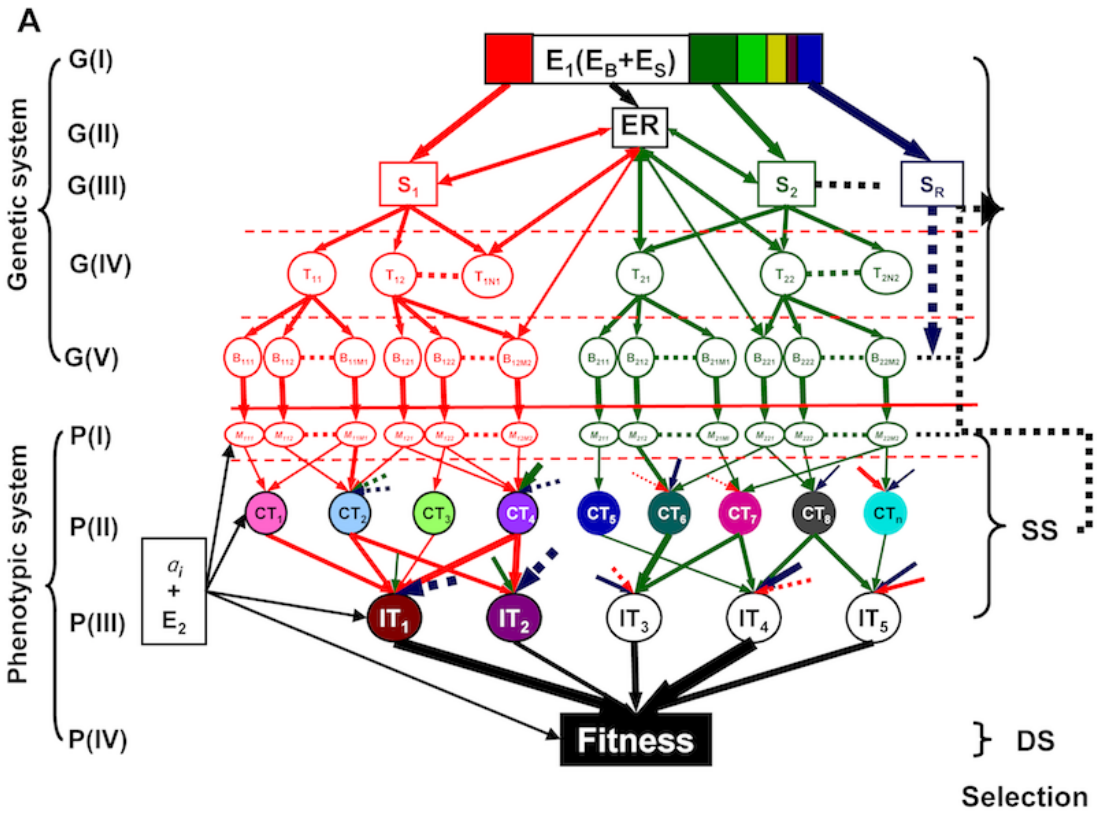
As the dimensionality of the search increases the background noise drowns out all real biological signals

$$N_{\text{tests}} = \frac{m \times (m-1)}{2}$$

e.g. $500000k$ SNPs $\rightarrow 1.25 \times 10^{11}$ tests

Outline

Expression traits likely have larger effect sizes



Discovery data

- BSGS data - 842 individuals
- Gene expression on whole blood
- 7339 traits with $n \geq 90\%$
- 528,509 SNPs

Replication data

- Fehrmann
  - $n = 1240$
  - Identical SNP chip and expression chip
- EGCUT
  - $n = 891$
  - Identical SNP chip and expression chip
- CHDWB
  - $n = 139$
  - Different SNP chip, same expression chip

Computation

Total number of tests

528,509 pairwise SNPs $\times$ 7,339 traits $=$ 1.02 quadrillion tests

epiGPU software

Performs $\sim$12 million association tests per second

GPU clusters

Supercomputers with 10s or 100s of GPUs can do this in a few weeks

Analysis outline

1. Discovery scan
2. Filtering of results based on threshold etc
3. Filtering based on interaction vs genetic effects
4. Replication in independent samples

Discovery and filtering

Perform 8 d.f. test for full genetic effect (additive + dominance + epistasis) at each SNP pair

1. Significance threshold $T = 2.91 \times 10^{-16}$
2. Remove SNP pairs with any class size < 5
3. Remove SNP pairs with LD $r^2 > 0.1$ or $D' > 0.1$
4. Keep the sentinel SNP pair for each chromosome $\times$ chromosome $\times$ trait
5. 11155 SNP pairs remain
6. Perform nested test of full genetic model (8 d.f.) vs marginal model (a + d, 4 d.f.)
7. Keep 4 d.f. interaction effects with $p < 0.05/11155$

501 significant interaction SNP pairs

Outline

501 significant interactions in 238 expression traits

Genomic positions

- 47 *cis-cis*
- 441 *cis-trans*
- 13 *trans-trans*

501 significant interactions in 238 expression traits

Genomic positions

- 47 *cis-cis*
- 441 *cis-trans*
- 13 *trans-trans*

Marginal effects ($p < 1.0 \times 10^{-10}$)

- 9 between two main effects
- 428 with only one main effect
- 64 with no main effects

501 significant interactions in 238 expression traits

Genomic positions

- 47 *cis-cis*
- 441 *cis-trans*
- 13 *trans-trans*

Marginal effects ($p < 1.0 \times 10^{-10}$)

- 9 between two main effects
- 428 with only one main effect
- 64 with no main effects

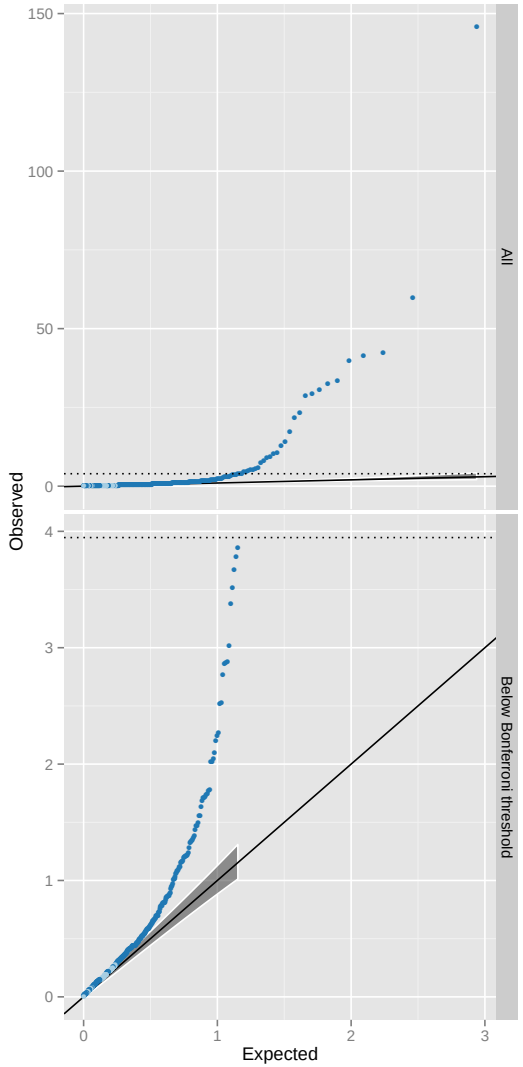
Largest epistatic variance component

- 120 A x A
- 255 A x D
- 126 D x D

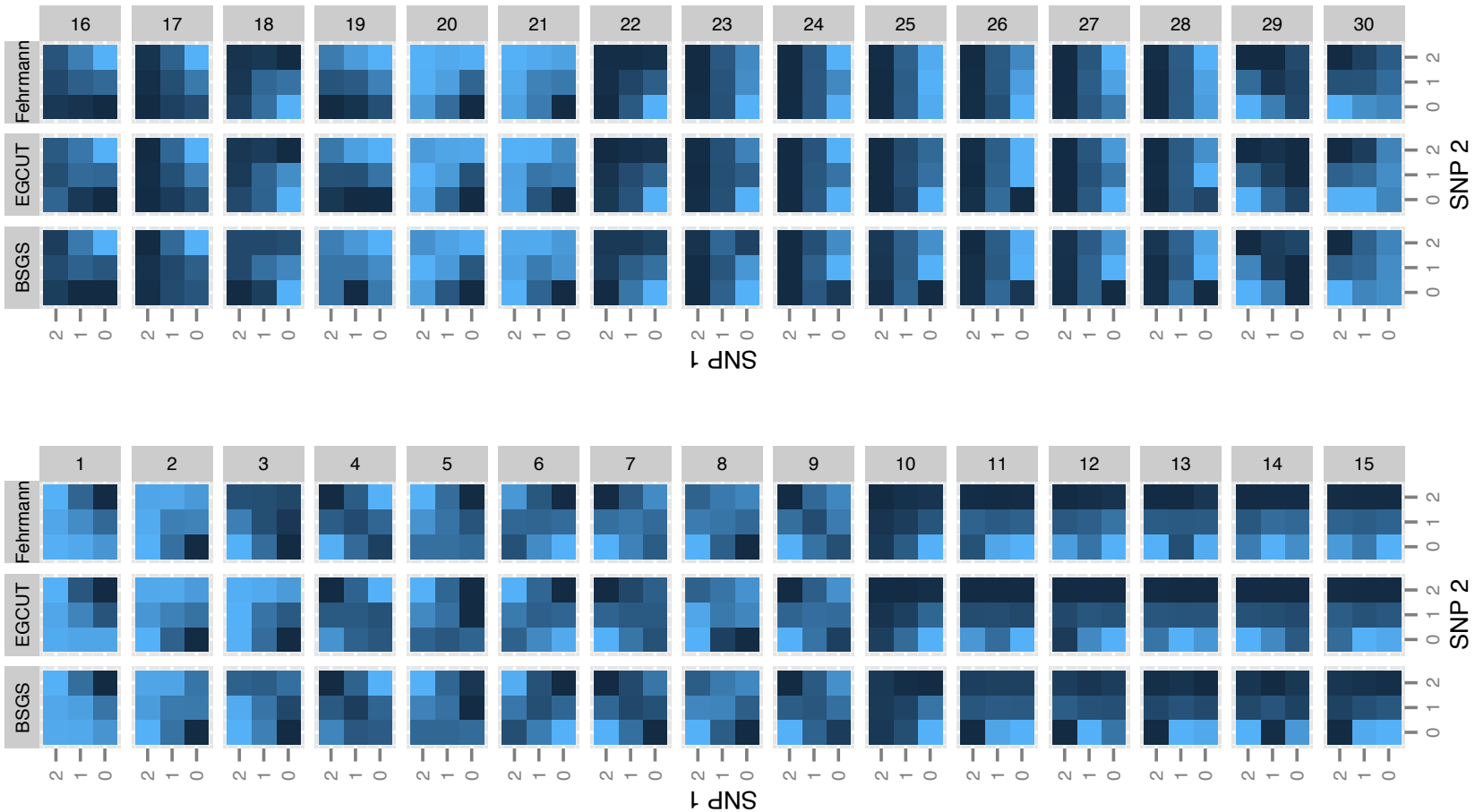
Replication

- Only 20 SNP pairs passed filtering in CHDWB
- 434 SNPs pairs passed QC in both EGCUT and Fehrmann
- 30 were significant for interaction p-values ($p < 0.05/434$) in EGCUT and Fehrmann

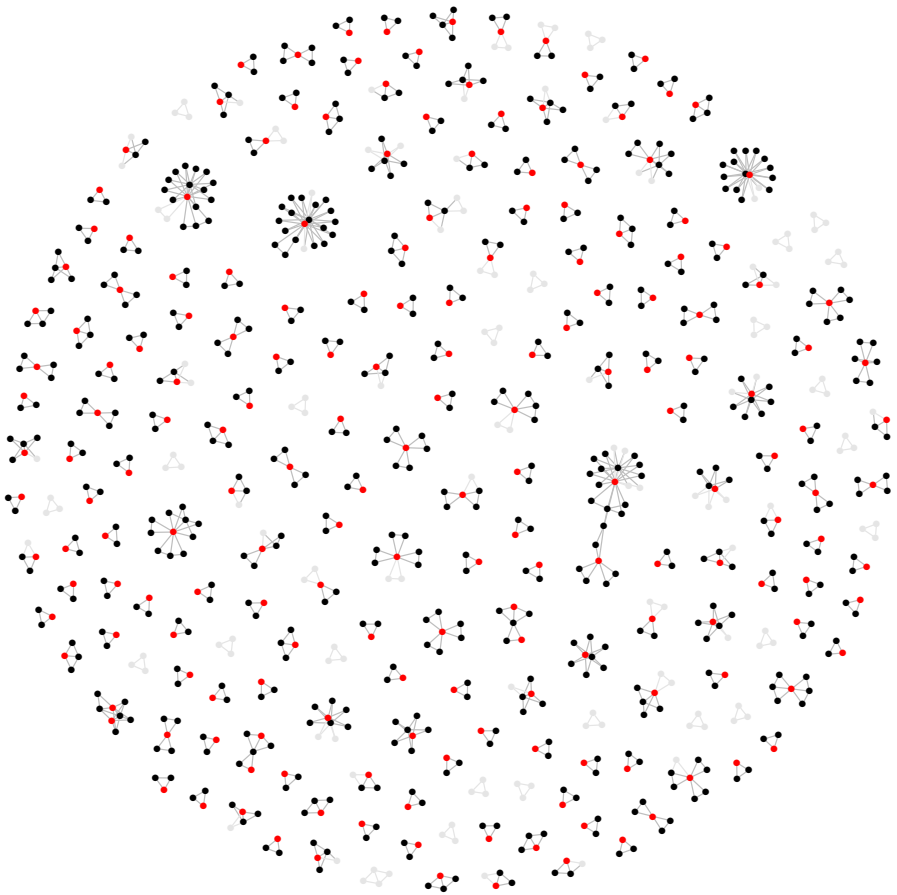
Q-Q plots of replication interaction *p*-values



Bonferroni level replicated GP maps

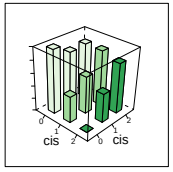


Map of interactions

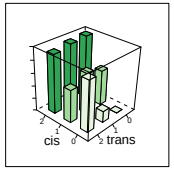


TMEM149

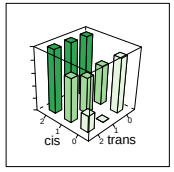
Chr 19 x 19
rs807491 x rs7254601



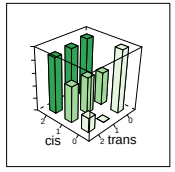
Chr 19 x 6
rs8106959 x rs6926382



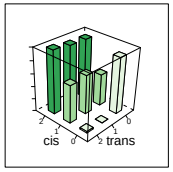
Chr 19 x 1
rs8106959 x rs914940



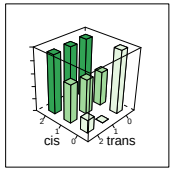
Chr 19 x 4
rs8106959 x rs2351458



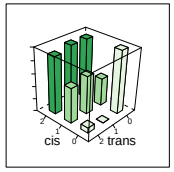
Chr 19 x 2
rs8106959 x rs6718480



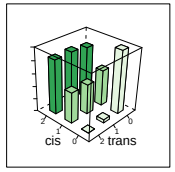
Chr 19 x 8
rs8106959 x rs1843357



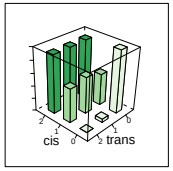
Chr 19 x 13
rs8106959 x rs9509428



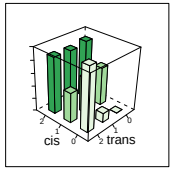
Chr 19 x 3
rs8106959 x rs10937361



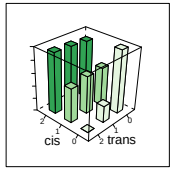
Chr 19 x 10
rs8106959 x rs10508289



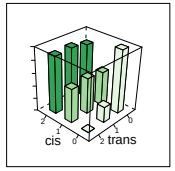
Chr 19 x 11
rs8106959 x rs471728



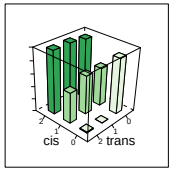
Chr 19 x 14
rs8106959 x rs17719594



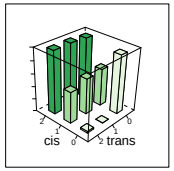
Chr 19 x 12
rs8106959 x rs1401098



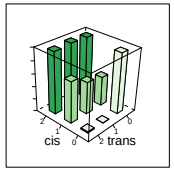
Chr 19 x 7
rs8106959 x rs2539000



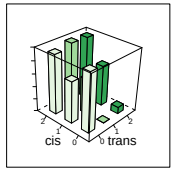
Chr 19 x 17
rs8106959 x rs7213338



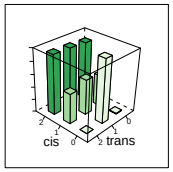
Chr 19 x 9
rs8106959 x rs10819626



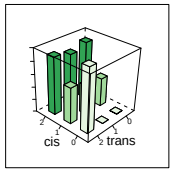
Chr 21 x 19
rs2839013 x rs8106959



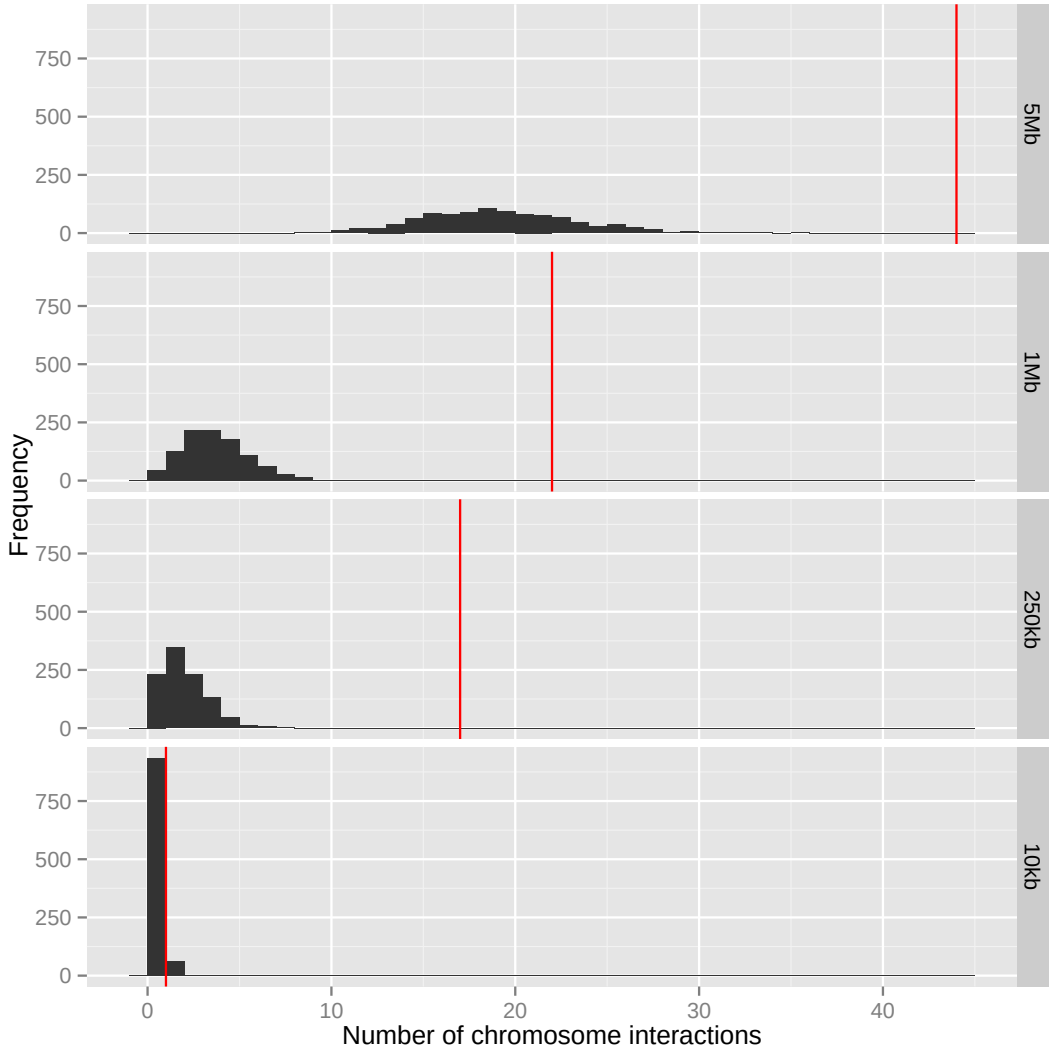
Chr 19 x 5
rs8106959 x rs2731711



Chr 19 x 18
rs8106959 x rs1557335



Chromosome interactions



Contribution relative to additive effects

At the same threshold (2.91×10^{-16})

- 453 expression traits have a significant additive effect
- 238 have a significant interaction effect

Contribution relative to additive effects

At the same threshold (2.91×10^{-16})

- 453 expression traits have a significant additive effect
- 238 have a significant interaction effect

At the same threshold (2.91×10^{-16})

- Significant additive effects explain 1.73% of phenotypic variance of 7339 traits
- Significant epistatic effects explain 0.25% (seven times less)

Outline

Acknowledgements

- University of Queensland
  - Joseph Powell
  - Konstantin Shakhbazov
  - Allan McRae
  - Jian Yang
  - Peter Visscher
- QIMR
  - Grant Montgomery
  - Nicholas Martin
  - Anjali Henders
- Estonian Genomics Centre
  - Tonu Esko
  - Andres Metspalu
- Georgia Tech University
  - Greg Gibson
- University of Groningen
  - Harm-Jan Westra
  - Lude Franke
- Computer resources
  - iVEC at University of Western Australia
  - MASSIVE project
  - QBI IT Team