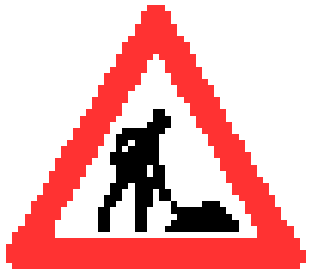
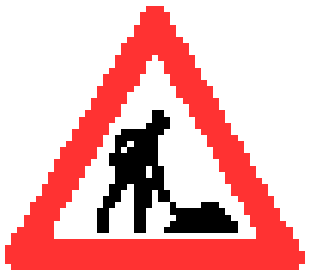
A simple method for detecting heterogeneity in meta-analyses of GWAS studies





David Evans

University of Bristol

wellcome trust

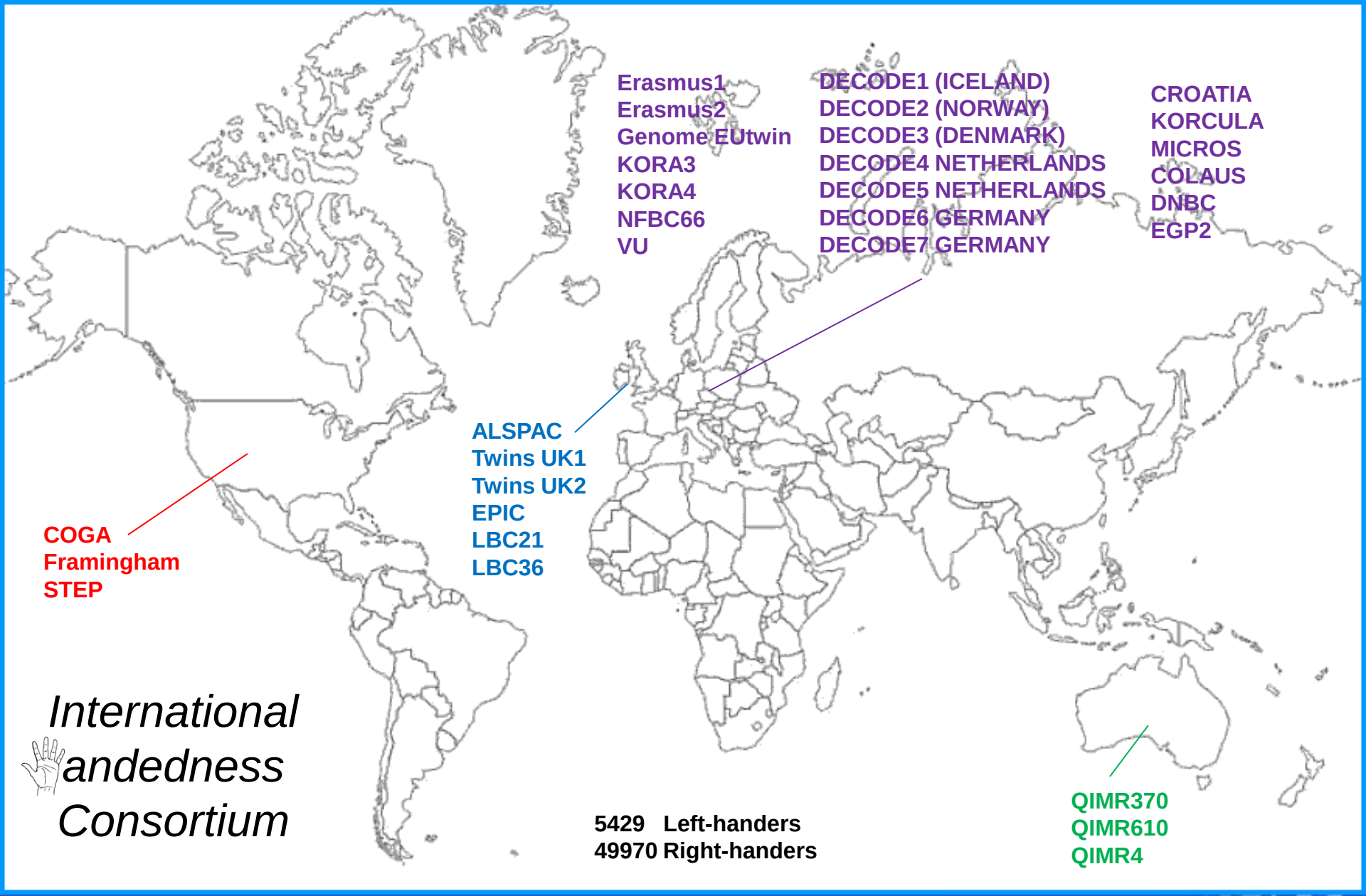
learning
discovery
enterprise

MRC Centre for Causal Analyses in Translational Epidemiology

CHILDREN OF THE 90s

Handedness

- First evidence at 9-10 weeks gestation
- Best predictor of cerebral lateralization
- “Which hand do you write with…”?
- Prevalence ~10%
  - Increasing since turn of century
  - Mirror-imaging?
- Heritability ~25% (Medland et al. 2009)
  - Single gene?
  - No evidence of C
  - Environmental effects include birth weight



International Handedness Consortium

University of BRISTOL

MRC Centre for Causal Analyses in Translational Epidemiology

Heterogeneity in Meta-analyses of GWAS

- GWAS studies can differ not only because of phenotypic and genetic heterogeneity, but also error

- Traditionally SNPs examined on a locus by locus basis
  - Tests have low power
  - Discards information present genome-wide

- Particularly relevant to handedness phenotype
  - Possibility of reverse coding
  - Particularly refractive to finding genes
  - No locus of strong effect to act as a positive control
  - Relevant to e.g. eczema etc

Polygenic Theory of Complex Traits

- Complex traits due to the action of hundreds of loci of small effect across the genome

- The results from GWA studies of the same (or similar) phenotypes should be more similar than GWA studies of etiologically different traits / GWA studies where there is added heterogeneity or systematic error

- Is it possible to harness this information to help detect heterogeneity between genome-wide association scans?

Human Molecular Genetics, 2009, Vol. 18, No. 18 **3525–3531**
doi:10.1093/hmg/ddp295
Advance Access published on June 24, 2009

Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk

David M. Evans[1,*], Peter M. Visscher[2] and Naomi R. Wray[2]

[1]Department of Social Medicine, MRC Centre for Causal Analyses in Translational Epidemiology, University of Bristol, Bristol, UK and [2]Genetic Epidemiology and Queensland Statistical Genetics, Queensland Institute of Medical Research, Australia

Table 1. Median AUC values for the seven diseases using the count and log odds methods

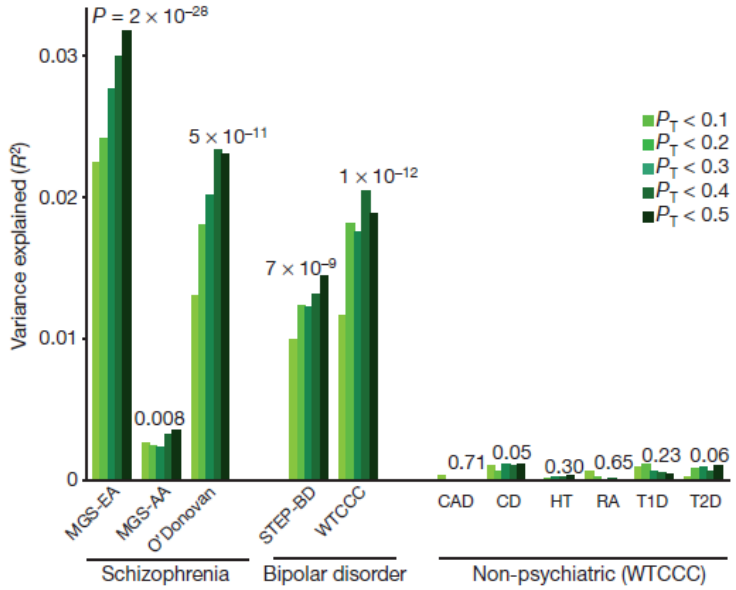
Threshold	BD	CHD	HT	CD	RA	T1D	T2D
Count method							
0.8	0.653 (0.527)	0.599 (0.527)	0.602 (0.538)	0.617 (0.554)	0.591 (0.522)	0.620 (0.513)	0.589 (0.520)
0.5	0.664 (0.527)	0.598 (0.533)	0.600 (0.534)	0.622 (0.553)	0.594 (0.528)	0.624 (0.515)	0.593 (0.513)
0.1	0.646 (0.537)	0.570 (0.532)	0.587 (0.534)	0.596 (0.524)	0.592 (0.515)	0.637 (0.515)	0.578 (0.516)
0.05	0.625 (0.537)	0.552 (0.525)	0.589 (0.532)	0.591 (0.521)	0.599 (0.524)	0.673 (0.537)	0.576 (0.529)
0.01	0.570 (0.555)	0.588 (0.508)	0.566 (0.518)	0.561 (0.514)	0.625 (0.522)	0.697 (0.531)	0.556 (0.516)
0.001	0.539 (0.534)	0.590 (0.534)	0.570 (0.521)	0.581 (0.532)	0.645 (0.546)	0.712 (0.544)	0.567 (0.549)
0.0001	0.533 (0.518)	0.551 (0.542)	0.568 (0.526)	0.624 (0.542)	0.647 (0.540)	0.716 (0.540)	0.565 (0.543)
0.00001	0.521 (0.525)	0.553 (0.509)	0.526 (0.536)	0.607 (0.558)	0.642 (0.528)	0.717 (0.540)	0.545 (0.515)
Log odds method							
0.8	0.668 (0.529)	0.595 (0.534)	0.610 (0.530)	0.614 (0.541)	0.646 (0.530)	0.721 (0.531)	0.601 (0.518)
0.5	0.668 (0.531)	0.592 (0.531)	0.610 (0.525)	0.618 (0.536)	0.642 (0.534)	0.724 (0.518)	0.601 (0.513)
0.1	0.636 (0.547)	0.580 (0.534)	0.599 (0.523)	0.598 (0.519)	0.652 (0.522)	0.743 (0.515)	0.574 (0.522)
0.05	0.620 (0.537)	0.560 (0.527)	0.596 (0.524)	0.592 (0.521)	0.656 (0.526)	0.747 (0.526)	0.568 (0.523)
0.01	0.567 (0.548)	0.600 (0.509)	0.585 (0.517)	0.574 (0.522)	0.666 (0.530)	0.749 (0.525)	0.569 (0.529)
0.001	0.533 (0.527)	0.590 (0.528)	0.580 (0.519)	0.597 (0.535)	0.661 (0.547)	0.749 (0.545)	0.575 (0.558)
0.0001	0.528 (0.520)	0.545 (0.544)	0.571 (0.534)	0.627 (0.544)	0.658 (0.557)	0.748 (0.534)	0.569 (0.549)
0.00001	0.529 (0.521)	0.556 (0.522)	0.520 (0.531)	0.612 (0.555)	0.655 (0.533)	0.749 (0.533)	0.544 (0.526)

nature Vol 460|6 August 2009|doi:10.1038/nature08185

LETTERS

Common polygenic variation contributes to risk of schizophrenia and bipolar disorder

The International Schizophrenia Consortium*



ISC (2009) Nature

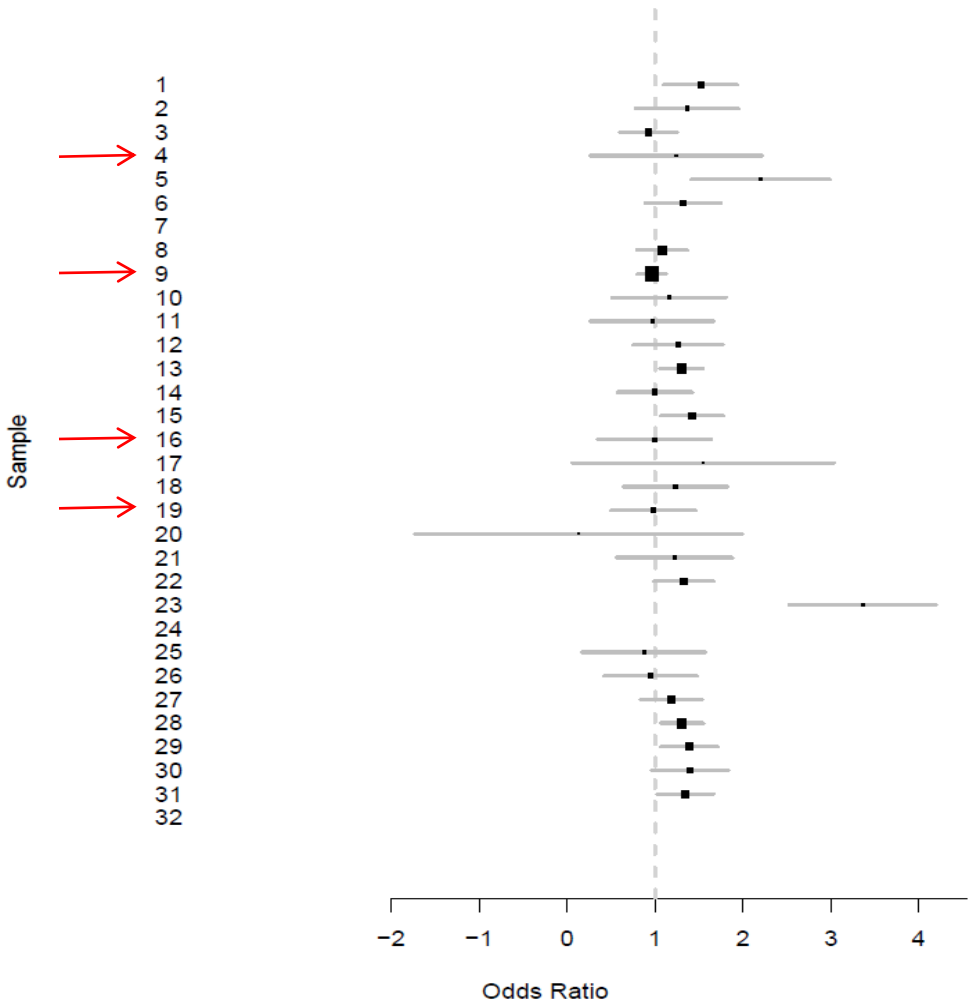
wellcome trust

University of BRISTOL

learning
discovery
enterprise

MRC Centre for Causal Analyses in Translational Epidemiology

Which cohorts show heterogeneity?



$p = 2.7 \times 10^{-6}$

METHOD

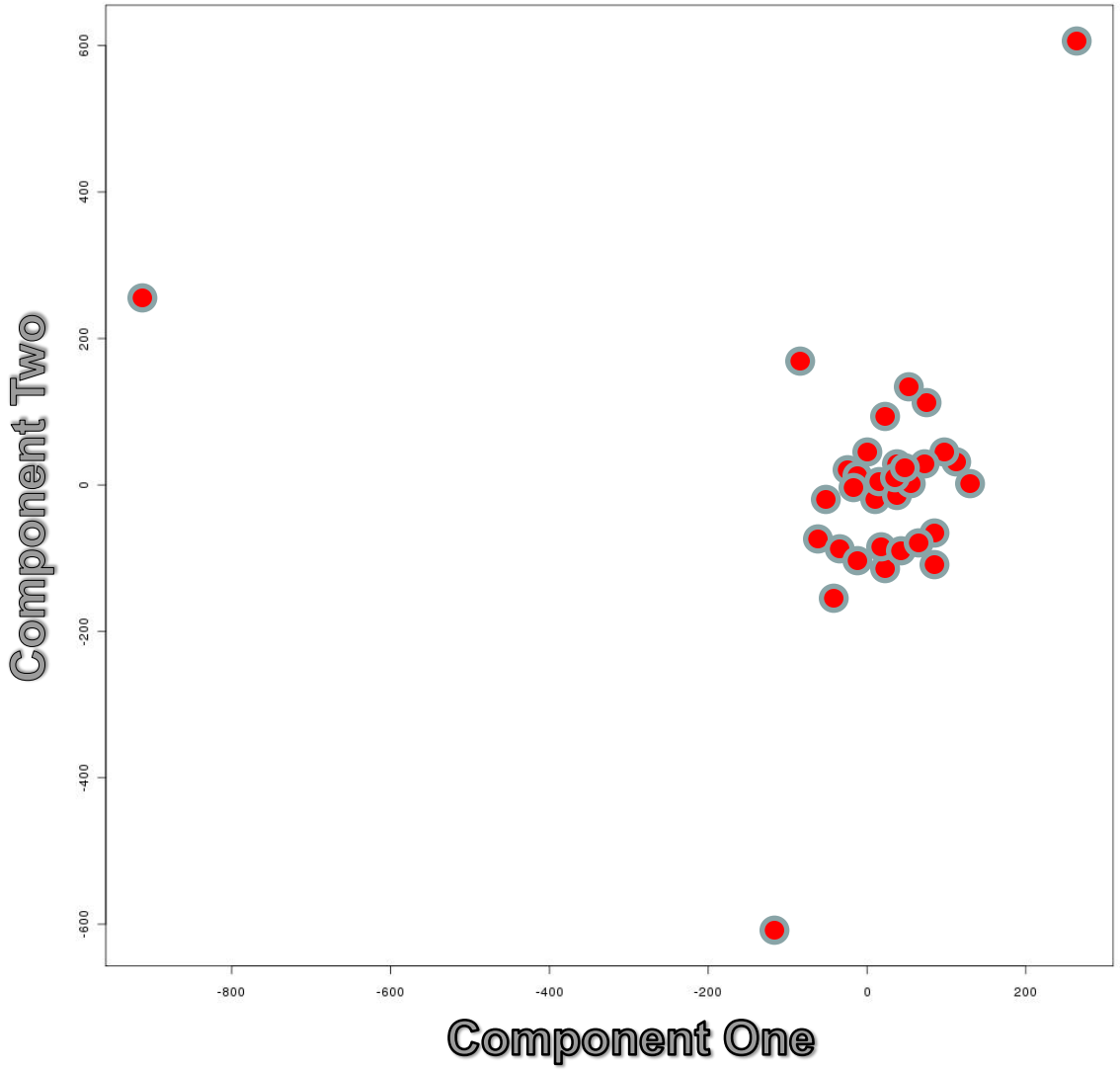
Start with N_{Cohort} x N_{SNP} matrix of p values

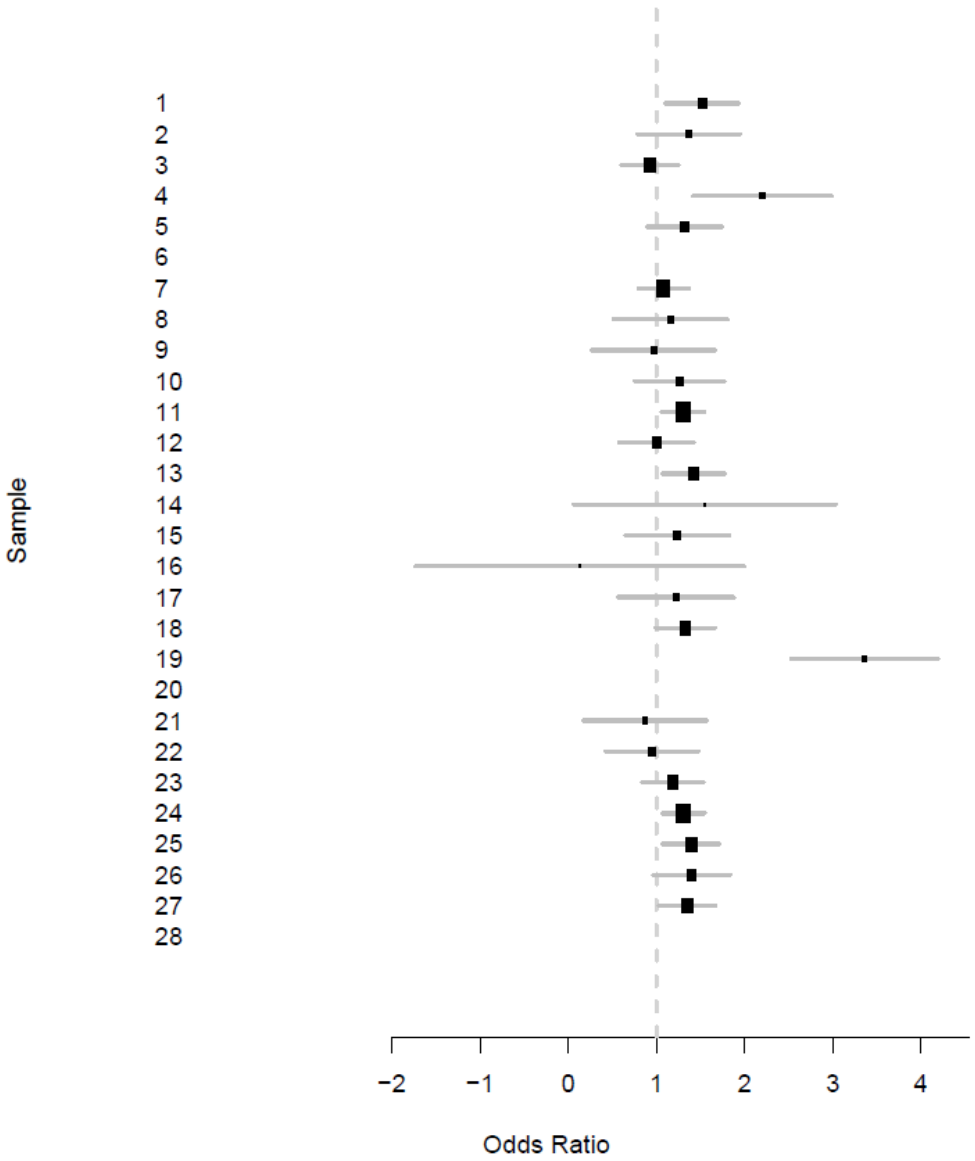
Transform each element to –log10pval x Direction of effect

Generate N_{Cohort} x N_{Cohort} matrix of Euclidean distances

Multi-dimensional scaling

Metric MDS

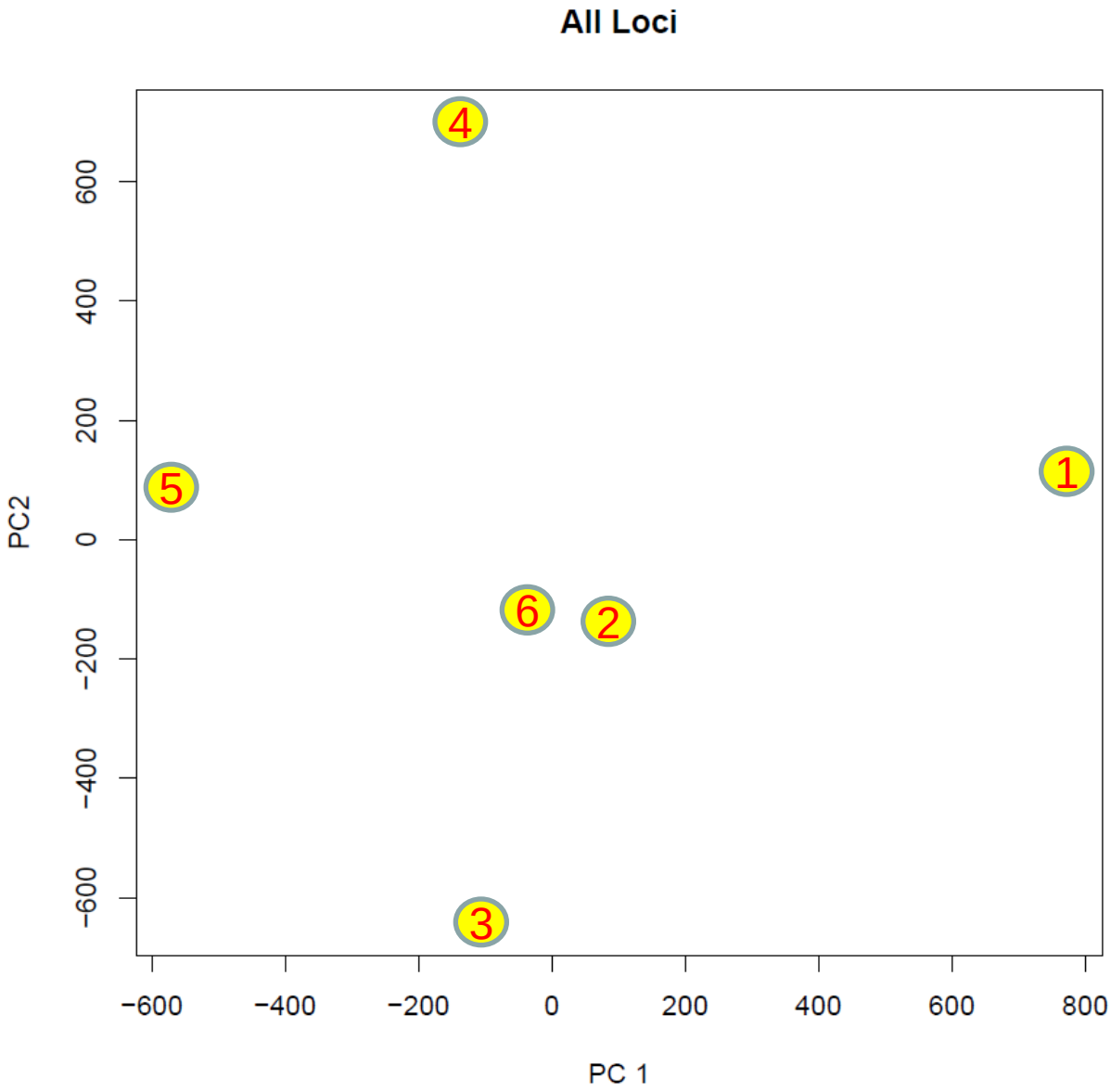

Component Two
600
400
200
0
-200
-400
-600
-800
-600
-400
-200
0
200
Component One



$p = 6.1 \times 10^{-8}$

Simulations

- 3000 individuals generated by HAPGEN
- 100 QTLs totalling 75% heritability randomly distributed across genome
- 6 x 500 individuals
- Dataset “6” has reverse coding


All Loci
PC2
PC 1
600
400
200
0
−200
−400
−600
−600
−400
−200
0
200
400
600
800
4
5
1
6
2
3

METHOD

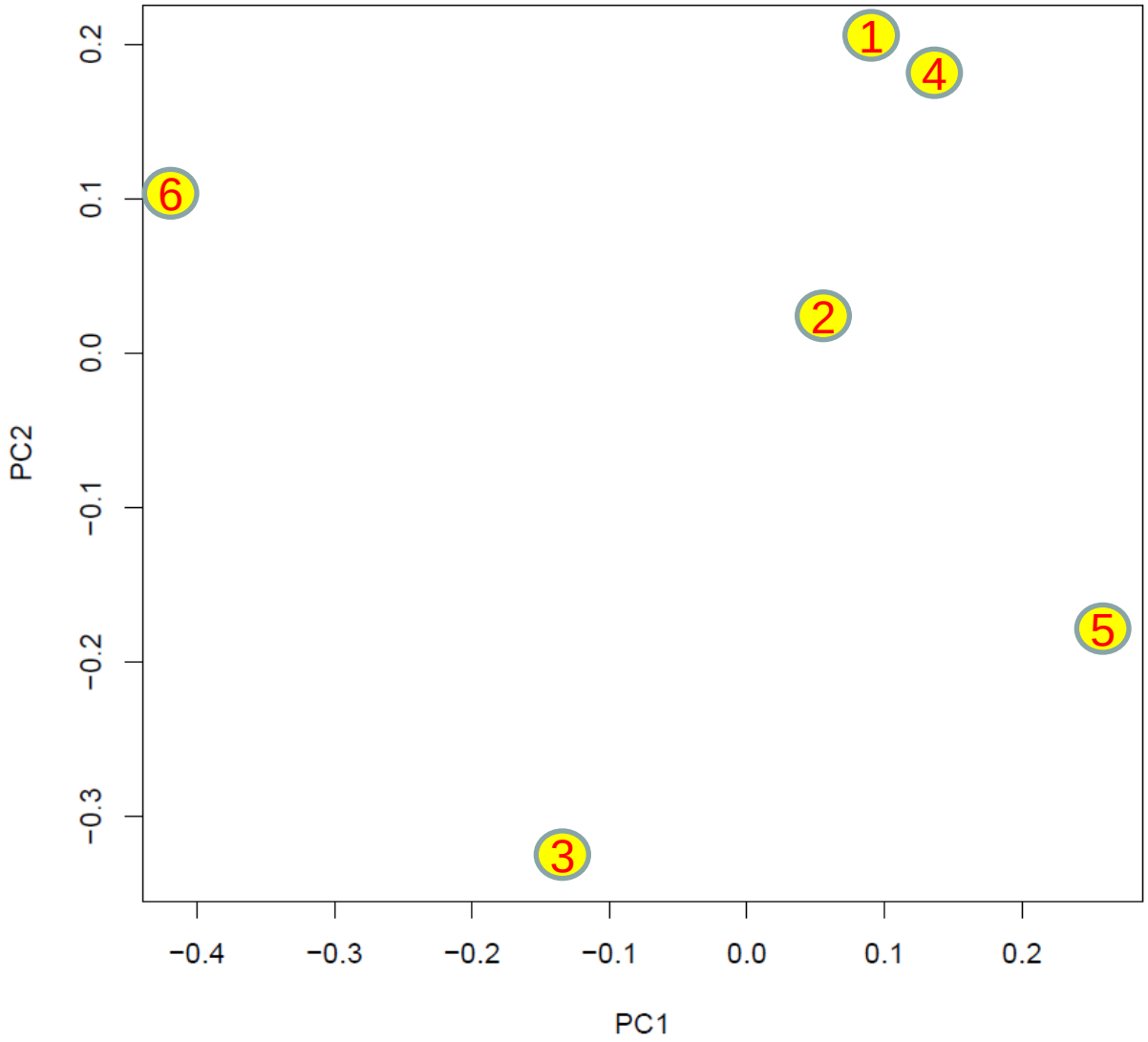
Start with N_{Cohort} x N_{SNP} matrix of p values

Similarity measure is n loci with p < threshold in both cohorts in same direction divided by n loci with p < threshold in both cohorts

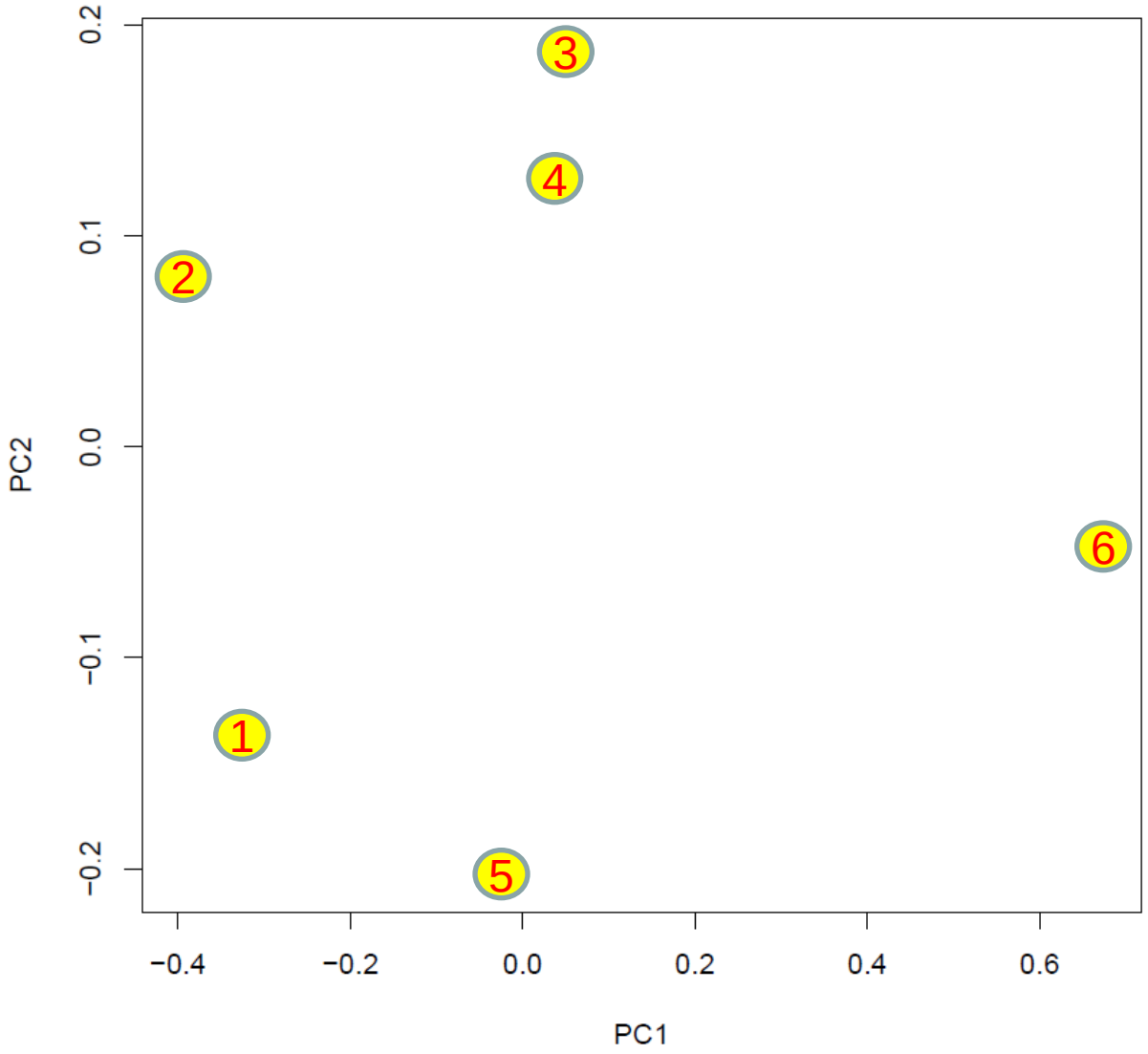
Generate N_{Cohort} x N_{Cohort} matrix of Euclidean distances

Multi-dimensional scaling

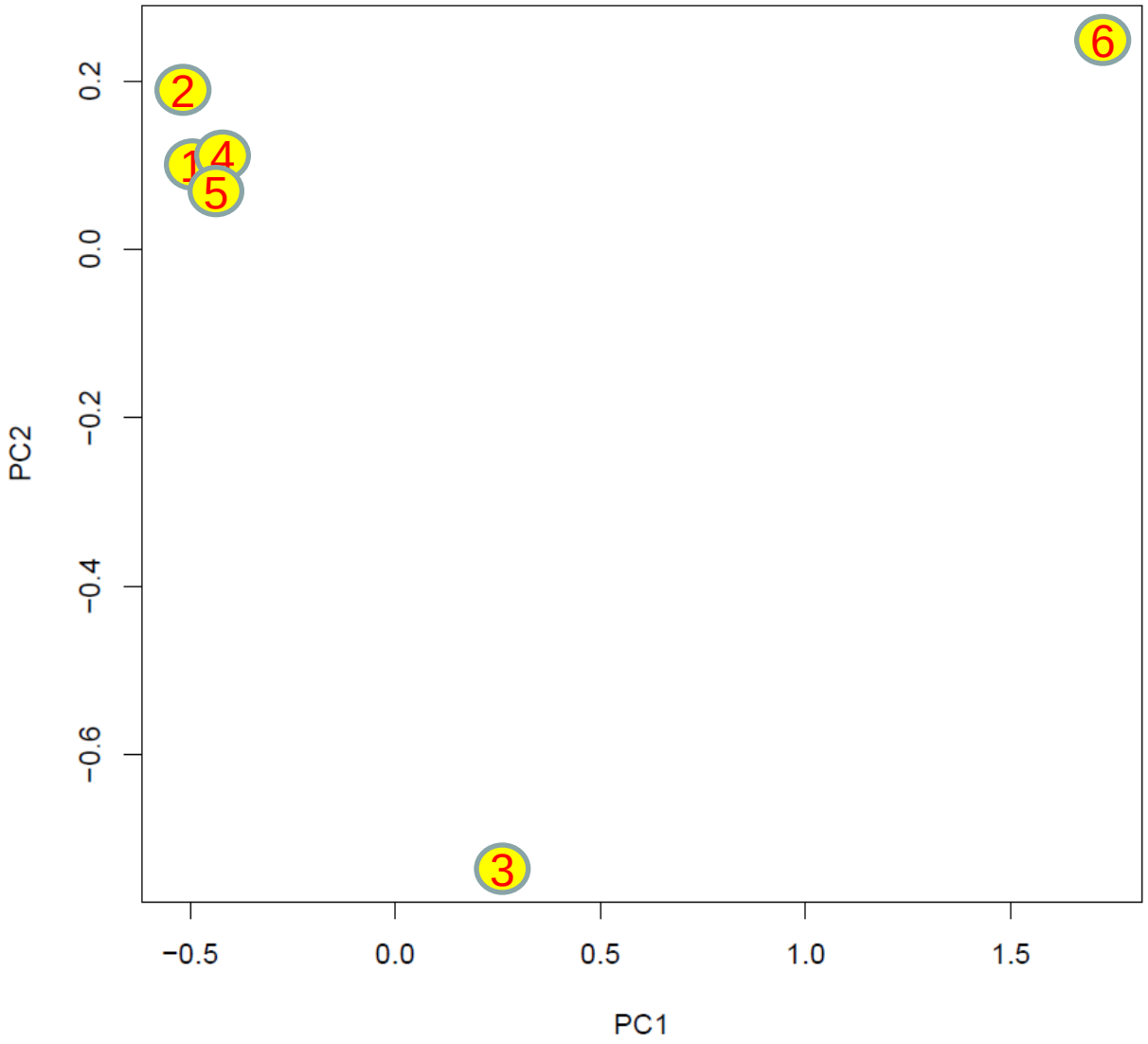
Shared Loci P < 0.1



Shared Loci P < 0.01



Shared Loci P < 0.001



Conclusions So Far?

- Clustering using number of shared “significant” loci may be better than using the whole genome (noisy)
- Stricter thresholds for selecting loci do better in terms of clustering properties
- Worryingly, cohorts without heterogeneity built into their model cluster strangely
- Strange clustering in Handedness Consortium may not be due to heterogeneity?

To Do

- Refine method
- Apply to simulated data with other sorts of heterogeneity (stratification etc), different genetic models, random and systematic genotyping error, sample sizes etc.
- Reapply to real data

Acknowledgments

- Sarah Medland Queensland Institute of Medical Research

- International Handedness Consortium