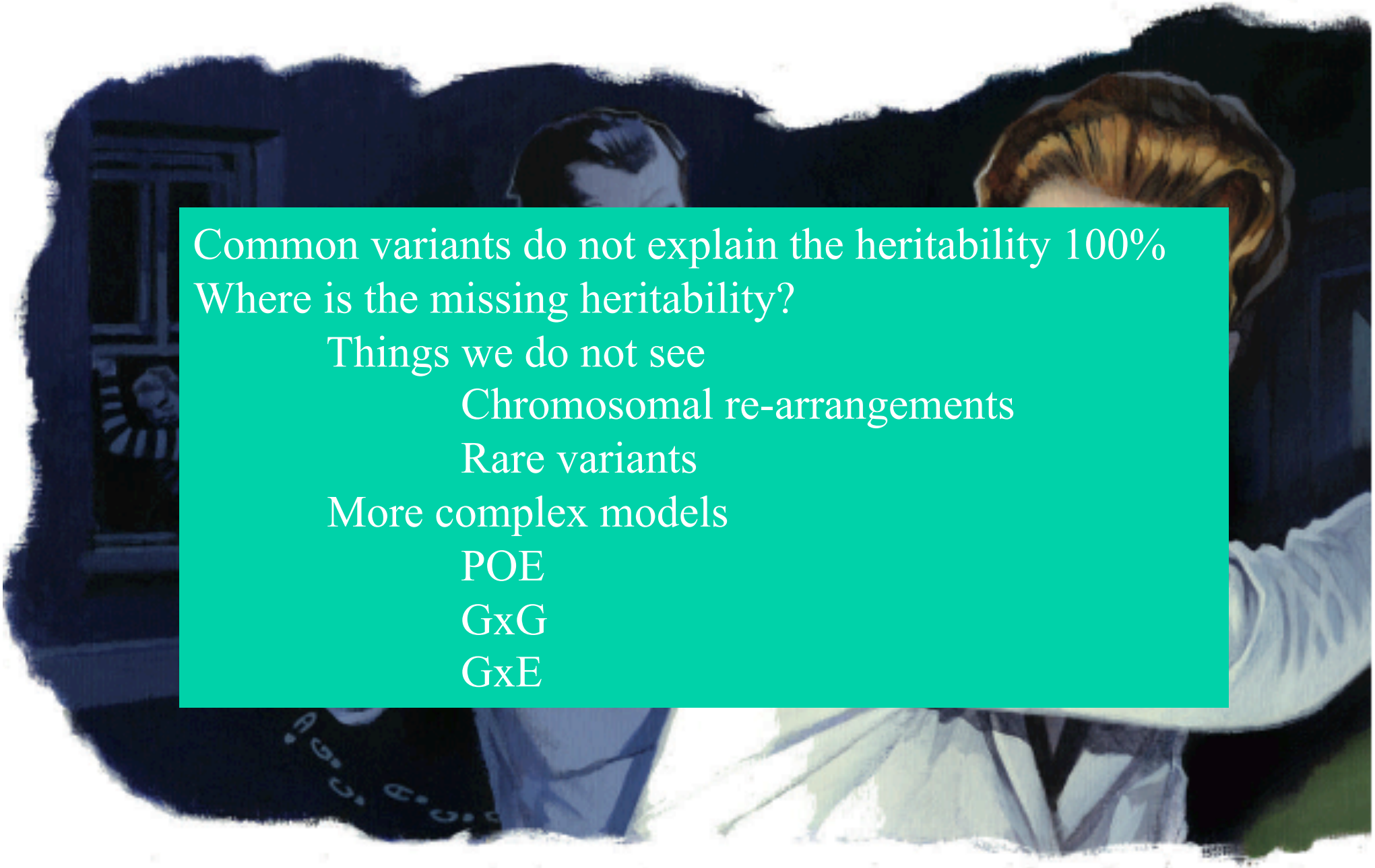
Searching for GxE in GWAS: statistical challenges and solutions

Yurii Aulchenko
Erasmus MC Rotterdam
10.02.2010



Common variants do not explain the heritability 100%

Where is the missing heritability?

- Things we do not see
  - Chromosomal re-arrangements
  - Rare variants
- More complex models
  - POE
  - GxG
  - GxE

The case of the missing heritability

GxE is a major player in complex traits

LETTERS

nature genetics

The sex-specific genetic architecture of quantitative traits in humans

Lauren A Weiss[1,2], Lin Pan[1], Mark Abney[1] & Carole Ober[1]

Mapping genetically complex traits remains one of the greatest challenges in human genetics today. In particular, gene-environment and gene-gene interactions, genetic heterogeneity and incomplete penetrance make thorough genetic dissection of complex traits difficult, if not impossible. Sex could be considered an environmental factor that can modify both penetrance and expressivity of a wide variety of traits. Sex is easily determined and has measurable effects on recognizable morphology; neurobiological circuits; susceptibility to autoimmune disease, diabetes, asthma, cardiovascular and psychiatric disease; and quantitative traits like blood pressure, obesity and lipid levels, among others. In this study, we evaluated sex-specific heritability and genome-wide linkages for 17 quantitative traits in the Hutterites. The results of this study could have important implications for mapping complex trait genes.

by sex. The importance of sex differences in disease course and prevalence and in response to drugs has recently been highlighted (for examples, see the 10 June 2005 issue of *Science*), but little is known about the underlying genetic architecture of these differences. Failing to model for sex-specific architecture may significantly hamper detection of susceptibility loci in genome-wide screens for complex traits[11]. In humans, although candidate genes are sometimes tested for sex-specific effects, few traits have been tested for sex-specific susceptibility loci in genome-wide screens. Psychiatric traits such as autism[12], neuroticism[13] and mood disorders[14], as well as immune-mediated disorders such as inflammatory bowel disease[15] and osteoarthritis[16] have shown sex-specific linkages. Serotonin and serum cortisol levels in the Hutterites, a founder population, have also shown marked sex-specific architecture[17,18].

In order to determine whether sex-specific genetics is limited to certain traits or is a more general phenomenon, in this study, we

Power of ENGAGE study

- Sample size up to 32,000 participants
- Under standard assumptions, we have
  - >90% power to detect loci explaining at least 0.2% of phenotypic variance
  - >50% power to detect loci explaining 0.1% of phenotypic variance
  - at genome-wide significance level !

- But nothing was detected at genome-wide significance level…
  - though there are many suggestive hits

A bunch of explanations

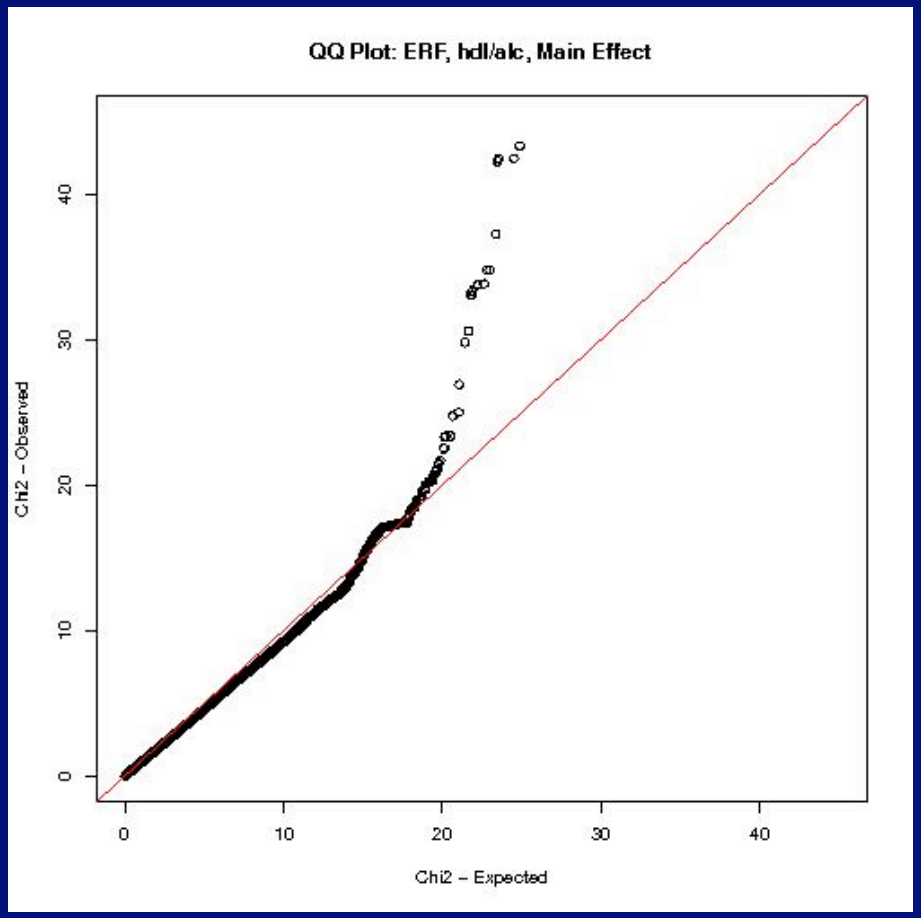
- Environmental factors
  - Are not well-measured
  - Have high missing rate
  - Are not so frequent

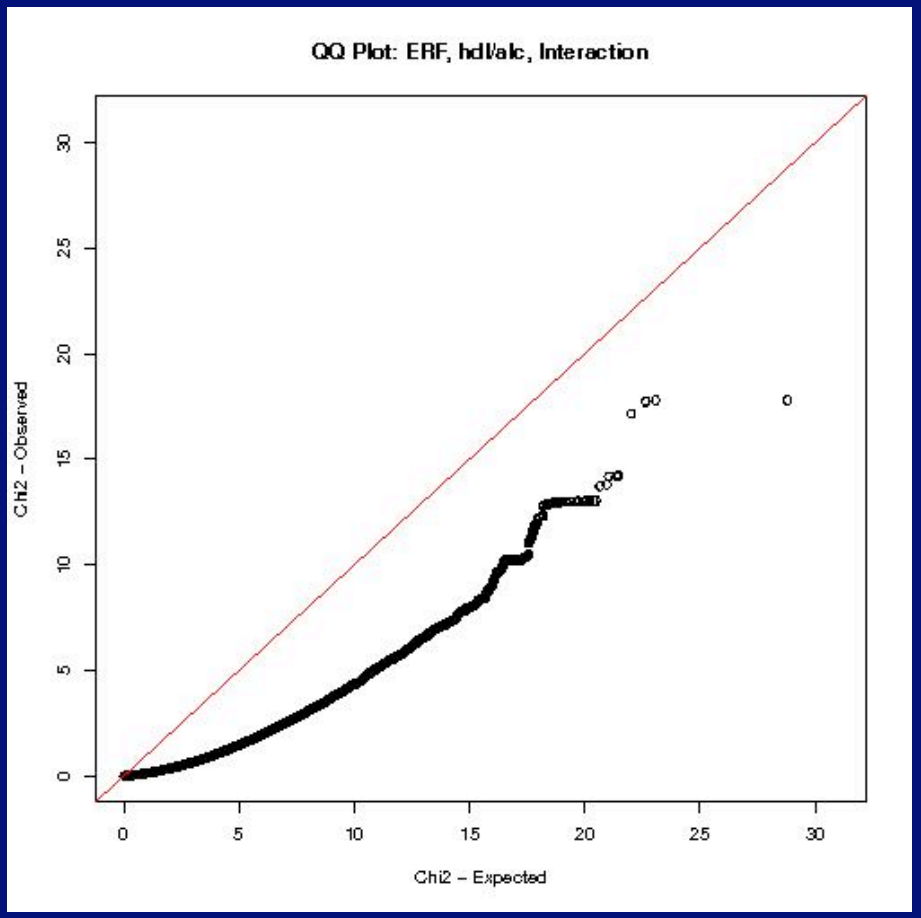
- GxE methodology is not perfect?

Too small λ's for GxE in pedigrees: problem definition

- FASTA/mmscore approximation: two-step procedure
  - estimate parameters of polygenic model
  - use score test on residuals, accounting for estimated matrix of variances and covariances between phenotypes
- Logic
  - any covariates, especially 'big' ones, should be included in step 1, otherwise score test assumptions are violated
  - put the main effect in the first step, and interaction-only in the second step
- What we see: $\lambda << 1$ for GxE term

What happens to λ's: en example





$\lambda = 0.918$

$\lambda = 0.281$

Solution & remaining problem

- Explanation:
  - there is covariance between main and interaction effect… Should never separate these two!
- Solution:
  - ProbABELv0.0-9 implements GLS estimator in 2nd step
- Problem to remain:
  - the covariates brought into the 2nd step are ‘big’ ones, violating score test assumption and leading to conservativity
- Solution to the problem remaining after solution …

λ's going all the way around 1

- Rotterdam study: **population-based** cohort used for genetic research for over 15 years
- In GWAS performed over many traits, always λ < 1.05
- GxE results for some traits:

	Environmental factor			
	cov 1	cov 2	cov 3	cov 4
trait 1	1.13	1.13	1	1.14
trait 2	0.98	1.04	1.02	1.04
trait 3	1.12	**1.22**	1	1.09
trait 4	1.05	1.01	1.03	0.97
trait 5	1.1	1.09	1.07	1.01
trait 6	1.02	1.01	0.92	1.03
trait 7	0.94	0.95	**0.89**	1

Solution: use robust (co)variances?

- Implemented in ProbABELv0.1-1

	Environmental factor			
	cov 1	cov 2	cov 3	cov 4
trait 1	1.03	1.04	1.03	1.02
trait 2	1.03	1.01	1.03	1.02
trait 3	1.02	1.04	1.03	1.02
trait 4	1.04	1.03	1.03	1.01
trait 5	1	1.02	1.03	1.01
trait 6	1.03	1.01	1.02	1.01
trait 7	1.02	1.03	1.03	1.01

Are we facing 'detectability limit'?

- Holmans 'detectability limit' hypothesis: in a study showing residual inflation of test statistics, there is a limit on detectable effect size, whatever the sample size is
- Power to detect locus explaining 0.1% using sample of 30,000 people is 50%
- If λ_{1000} = 1.03 only, the power is 2% (!)
- Detectability limit with λ_{1000} =1.03 is a locus explaining 0.09% of variance (whatever sample size!)
- Read it other way: we hoped that brute force approach will always work (by big N's we can compensate for imperfect methodology). THIS IS NOT TRUE

Variance analysis

- Conducting Bartlett-RN-transform test
  - Not correct if main effect is present
- Effects of using ‘best guess’ genotypes
  - This leads to biased effect estimates, hence possibly major drop in power
- Levene’s for meta-analysis?
  - have a formula
- Incorporation of uncertainty into analysis? Additive, etc. models?
  - likelihood, have a formula
- **Levene AND uncertainity?!**
  - Wow! Weighted Levene? A lot of work to do…

Conclusions

- We learned a lot from ENGAGE lipids GxE project
- Common Variants X (sex, smoking, alcohol, …) do not explain a lot: under any assumptions, we should have detected any interaction explaining ≥ 1% of phenotypic variance
- With GWAS we are facing fundamental statistical limits. The role of methodology & software developments is very important