# Genetic interactions

Yurii Aulchenko
yurii [dot] aulchenko [at] gmail [dot] com

# Loci identified for complex traits

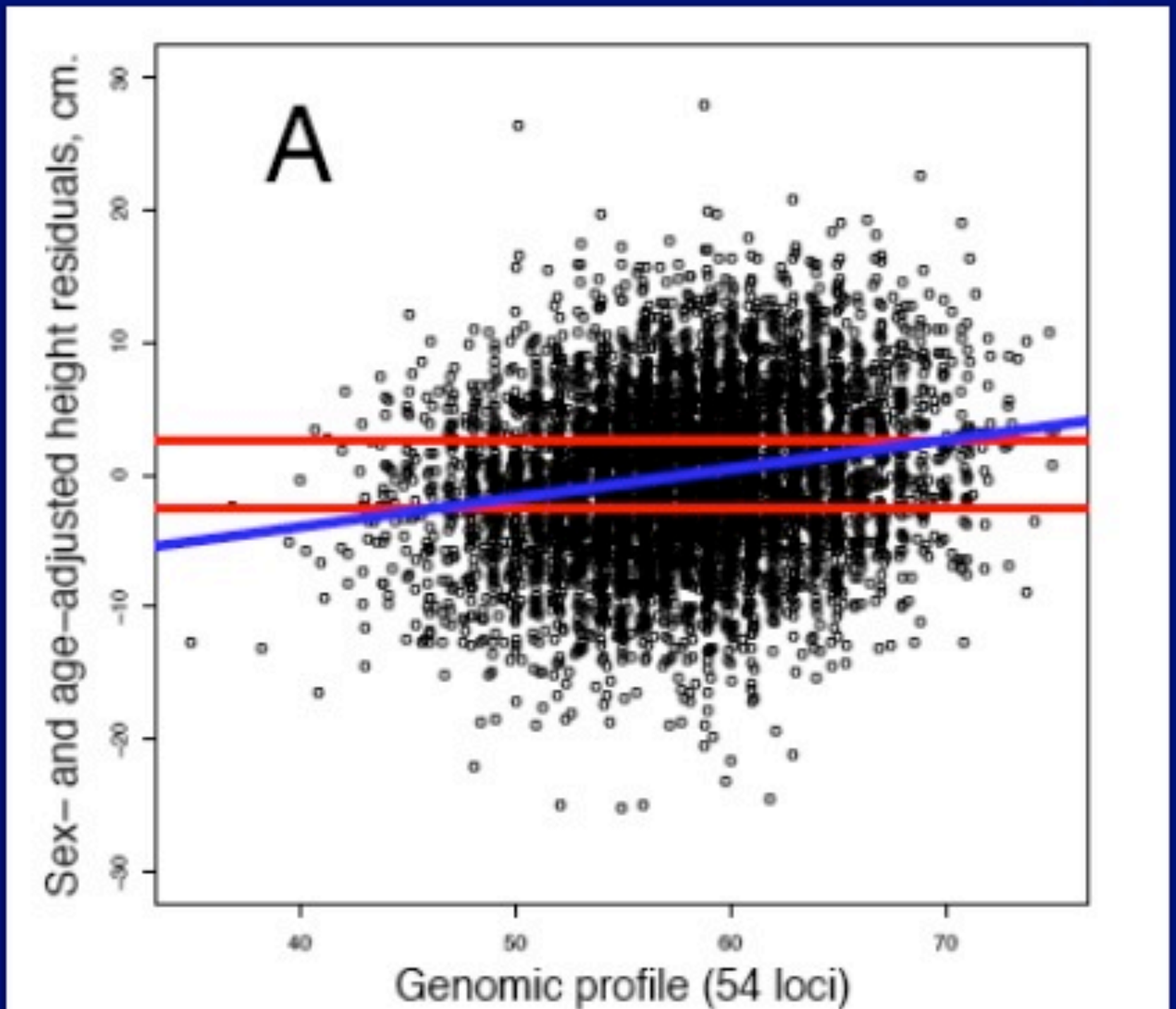
| | # Loci | | | |
|---|---|---|---|---|
| | <2005 | 2008 | 2010 | 2012 |
| Lipids | few | ~30 | 95 | +200 |
| Height | 0 | ~50 | 100+ | +300 |

# Loci identified for complex traits

| | # Loci | | | |
|---|---|---|---|---|
| | <2005 | 2008 | 2010 | 2012 |
| Lipids | few | ~30 | 95 | +200 |
| Height | 0 | ~50 | 100+ | +300 |

| | %Var | | | |
|---|---|---|---|---|
| | <2005 | 2008 | 2010 | 2012 |
| Lipids | ~2% | 5% | 10% | +15% |
| Height | 0 | 4% | 8% | +10% |

# Post-genomic prediction of human height

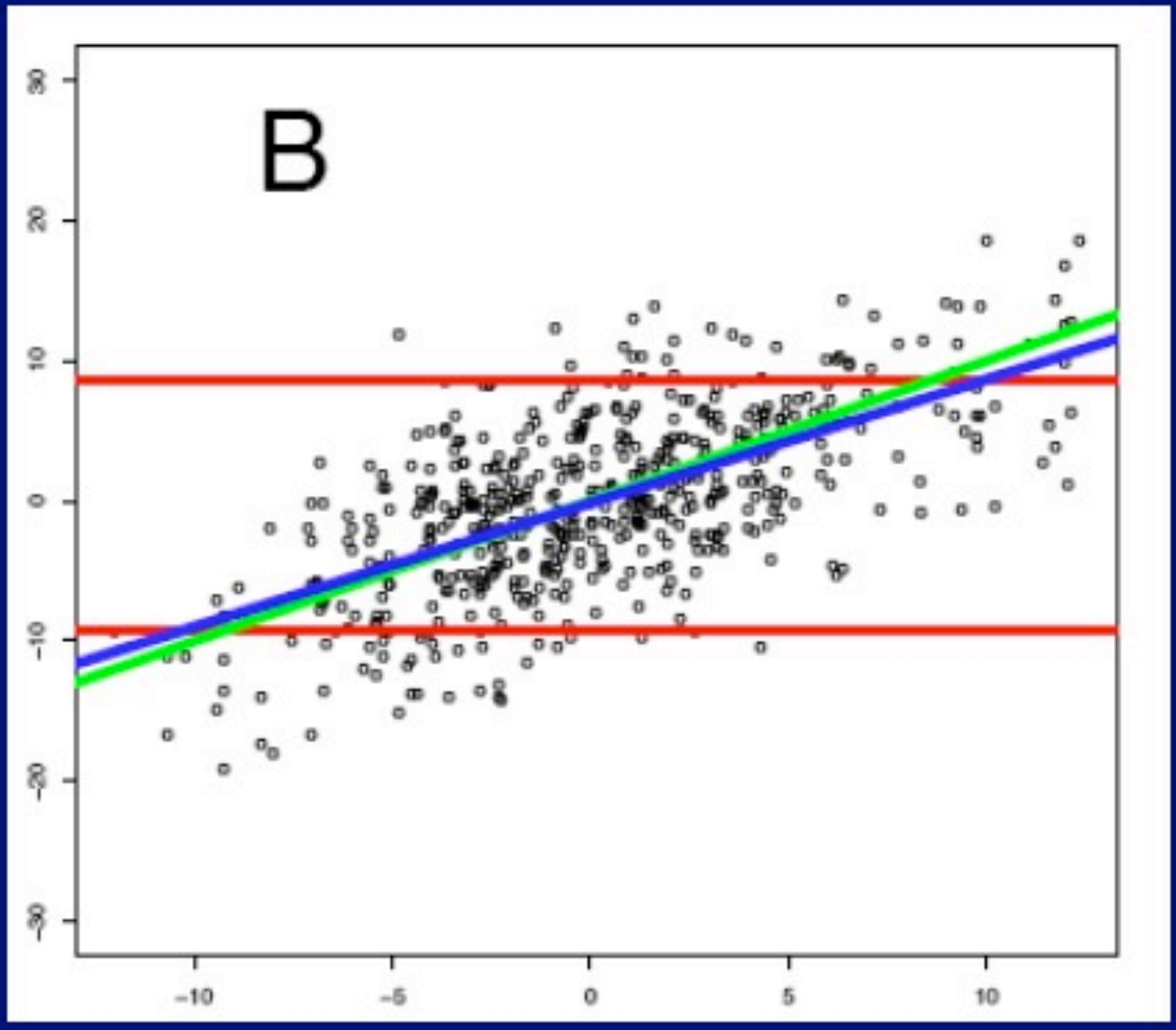



54 loci (2008)
explain 4%
of height variance

Recent height
paper: 8% (?)

*Aulchenko et al., 2009, EJHG*

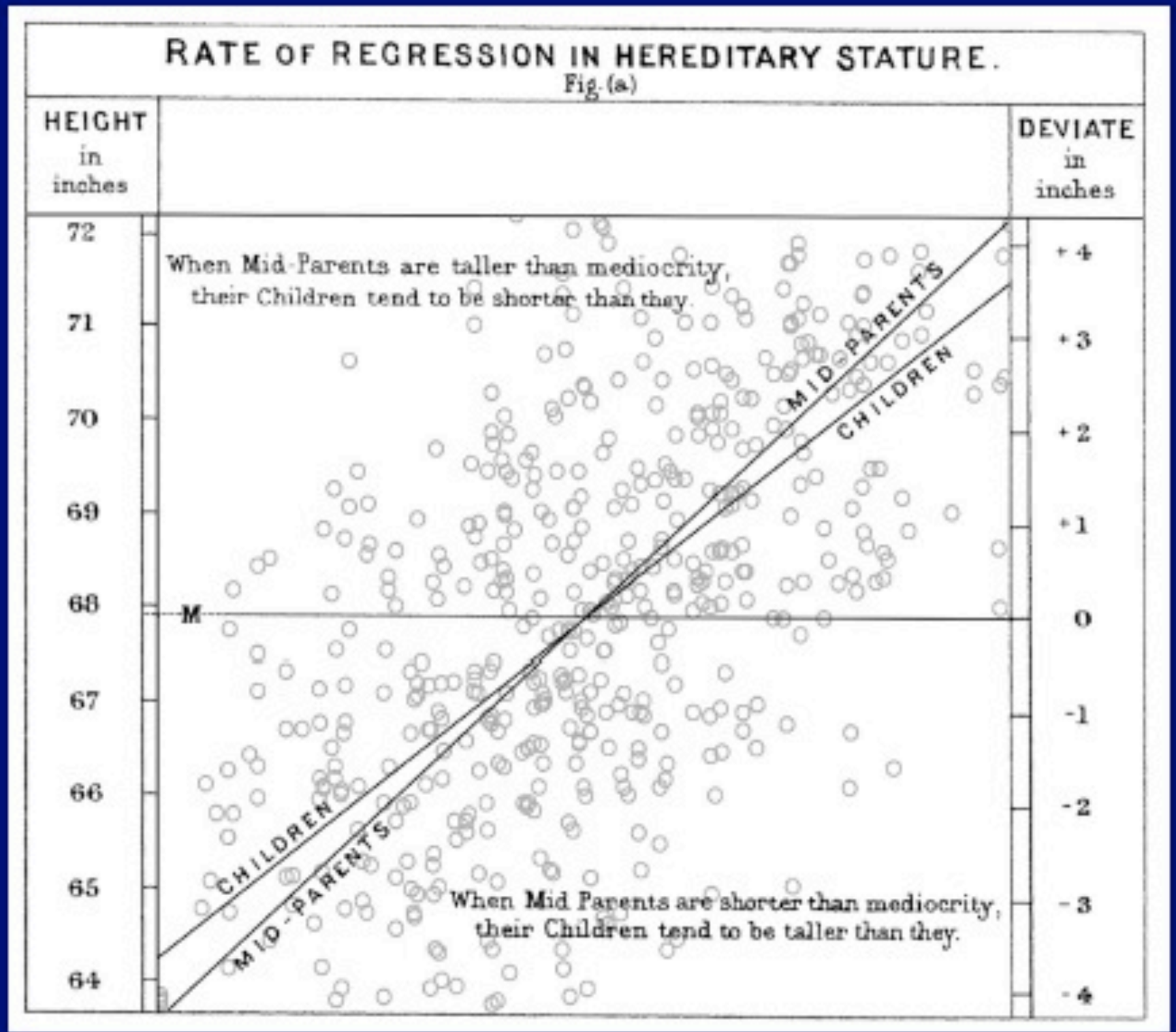
# Yet another prediction of human height




This profile explains 40% of height variance

*Aulchenko et al., 2009, EJHG*

# Galton, 1886, “Regression towards mediocrity in hereditary stature”



# Where to move next?

Alleles of small effects

Things we do not see/check

- Missing genome: X, mt, Y
- True causative variants (not tags!)
- Chromosomal re-arrangements
- Rare point mutations

More complex models (all kind of **interactions**)

- Inter-locus (e.g. dominance)
- Intra-locus (GxG)
- Gene-environment (GxE)
- Parent-of-origin (POE)

**The case of the missing heritability**

# Where to move next?

Bigger studies

Alleles of small effects

Things we do not see/check

- Missing genome: X, mt, Y
- True causative variants (not tags!)
- Chromosomal re-arrangements
- Rare point mutations

More complex models (all kind of **interactions**)

- Inter-locus (e.g. dominance)
- Intra-locus (GxG)
- Gene-environment (GxE)
- Parent-of-origin (POE)

**The case of the missing heritability**

# Where to move next?

Alleles of small effects

Bigger studies

Things we do not see/check

- Missing genome: X, mt, Y
- True causative variants (not tags!)
- Chromosomal re-arrangements
- Rare point mutations

Need to look at that

More complex models (all kind of **interactions**)

- Inter-locus (e.g. dominance)
- Intra-locus (GxG)
- Gene-environment (GxE)
- Parent-of-origin (POE)

**The case of the missing heritability**

# Where to move next?

Alleles of small effects

Bigger studies

Things we do not see/check

- Missing genome: X, mt, Y
- True causative variants (not tags!)
- Chromosomal re-arrangements
- Rare point mutations

Need to look at that

Re-sequencing (technically)

More complex models (all kind of **interactions**)

- Inter-locus (e.g. dominance)
- Intra-locus (GxG)
- Gene-environment (GxE)
- Parent-of-origin (POE)

**The case of the missing heritability**

# Where to move next?

Alleles of small effects

Bigger studies

Things we do not see/check

- Missing genome: X, mt, Y
- True causative variants (not tags!)
- Chromosomal re-arrangements
- Rare point mutations

Need to look at that

Re-sequencing (technically)

More complex models (all kind of **interactions**)

- Inter-locus (e.g. dominance)
- Intra-locus (GxG)
- Gene-environment (GxE)
- Parent-of-origin (POE)

**Statistical modeling**

**The case of the missing heritability**

# The mystery of missing heritability: Genetic interactions create phantom heritability

**Or Zuk[a], Eliana Hechter[a], Shamil R. Sunyaev[a,b], and Eric S. Lander[a,1]**

Broad Institute of MIT and Harvard, Cambridge, MA 02142; and [b]Genetics Division, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115

Contributed by Eric S. Lander, December 5, 2011 (sent for review October 9, 2011)

# Interaction models

The value of the trait in $i$-th individual is assumed to follow linear model

$$Y_i = m + b_f F_i + b_g g_i + b_{fg} F_i g_i + e_i$$

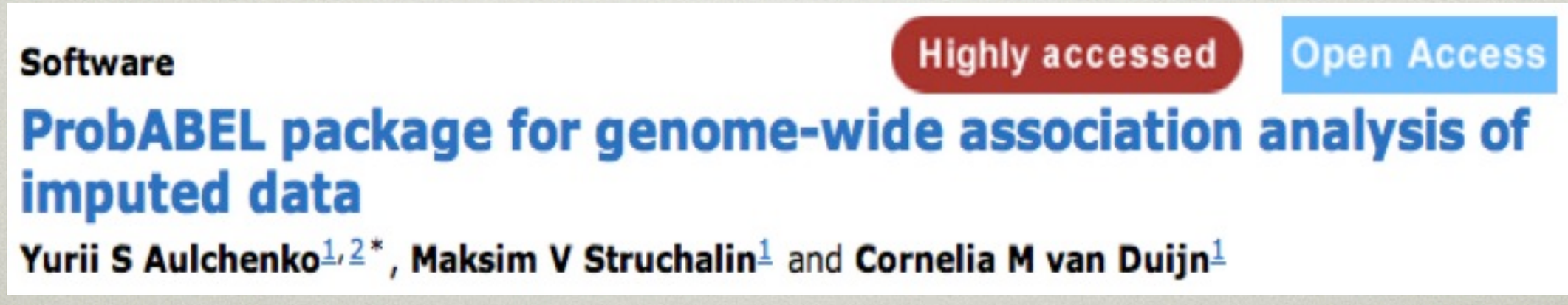
where $m$ is intercept, $F_i$ is the value of some "factor", $g_i$ is the genotypic value, and $e_i$ is random residual error

# What could "F" be?

- Alleles at other locus (GxG)
- An environment (GxE) - with E being external or internal (e.g. sex)
- Methylation status
- Indicator of transmitting parent (parent of origin models)
- Other allele at the same locus
- ... etc.

# GxE

- We thought that modeling was trivial
- ... but initial results were strange, so some methodological work had to be done ...


Software

Highly accessed

Open Access

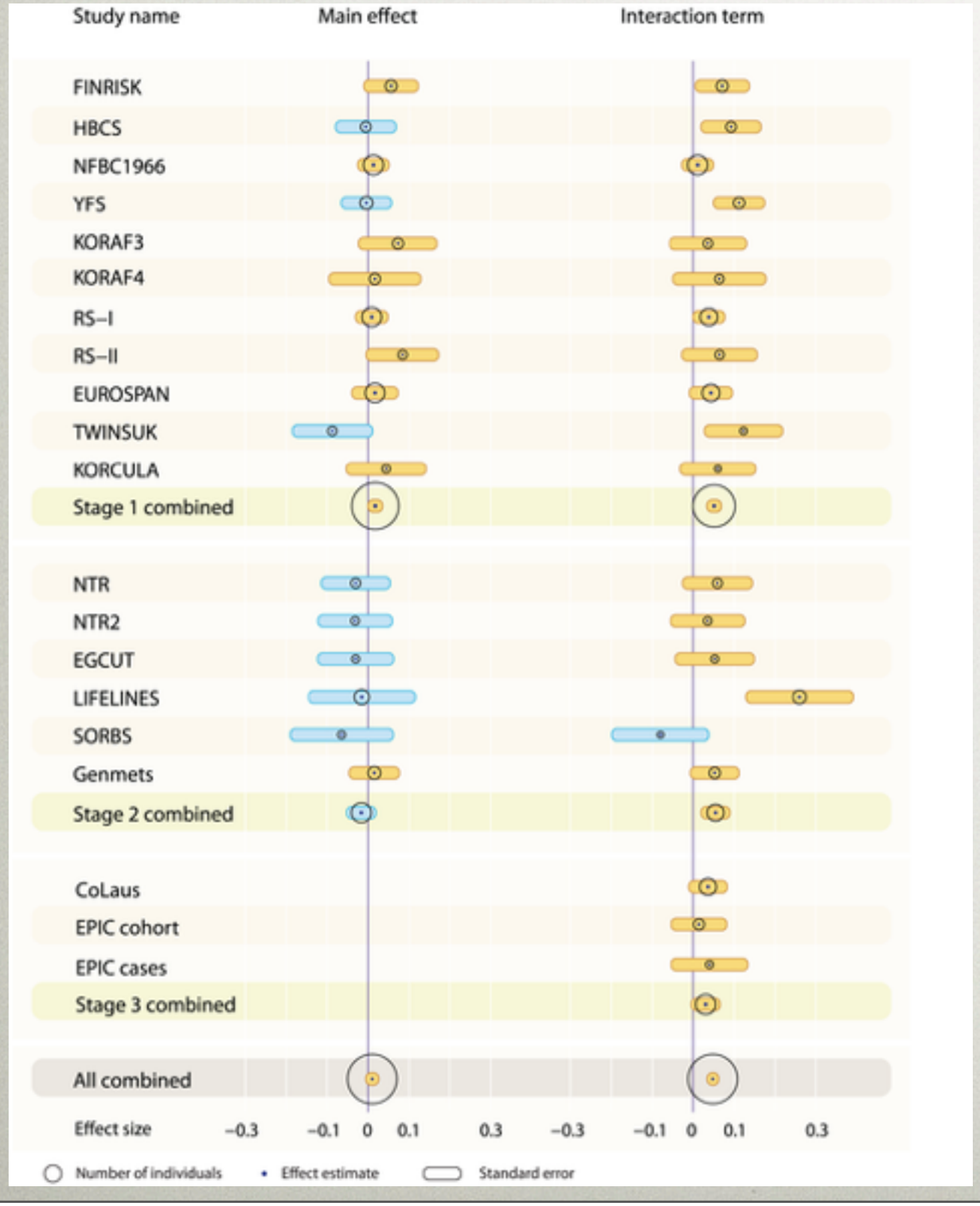
**ProbABEL package for genome-wide association analysis of imputed data**

Yurii S Aulchenko[1,2*], Maksim V Struchalin[1] and Cornelia M van Duijn[1]

OPEN ACCESS Freely available online

PLoS GENETICS

# A Genome-Wide Screen for Interactions Reveals a New Locus on 4p15 Modifying the Effect of Waist-to-Hip Ratio on Total Cholesterol




- A meta-analysis of genome-wide association (GWA) data from 18 population-based cohorts with European ancestry (maximum *N* = 32,225).
- Eight further cohorts (*N* = 17,102) for replication
- SNP *rs6448771* demonstrated genome-wide significant interaction with waist-to-hip-ratio (WHR) on total cholesterol (TC) with a combined *P*-value of $4.79 \times 10^{-9}$

# Chasing the unknown: Gx*

- Let assume that the model is the same

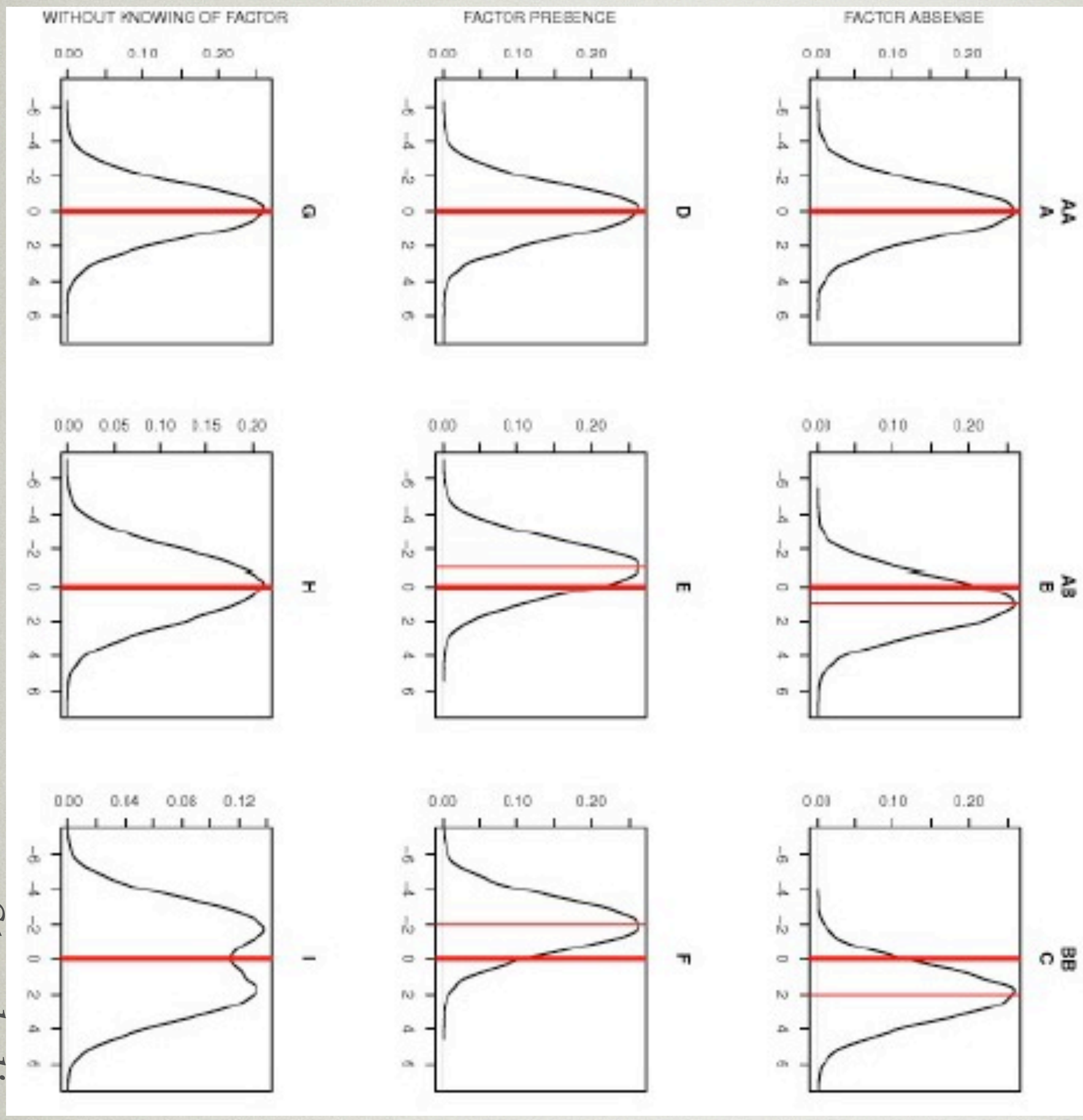
$$y_i \sim \mu + \beta_g g_i + \beta_F F_i + \beta_{gF} \cdot g_i F_i + \epsilon_i,$$

- but assume WE DO NOT KNOW F

- ? Can we work out a method which tell us what SNPs are interacting (with some unknown F) ? … then we can look for F – e.g. test a number of 'environmental covariates, other G's showing Gx*…

# True model

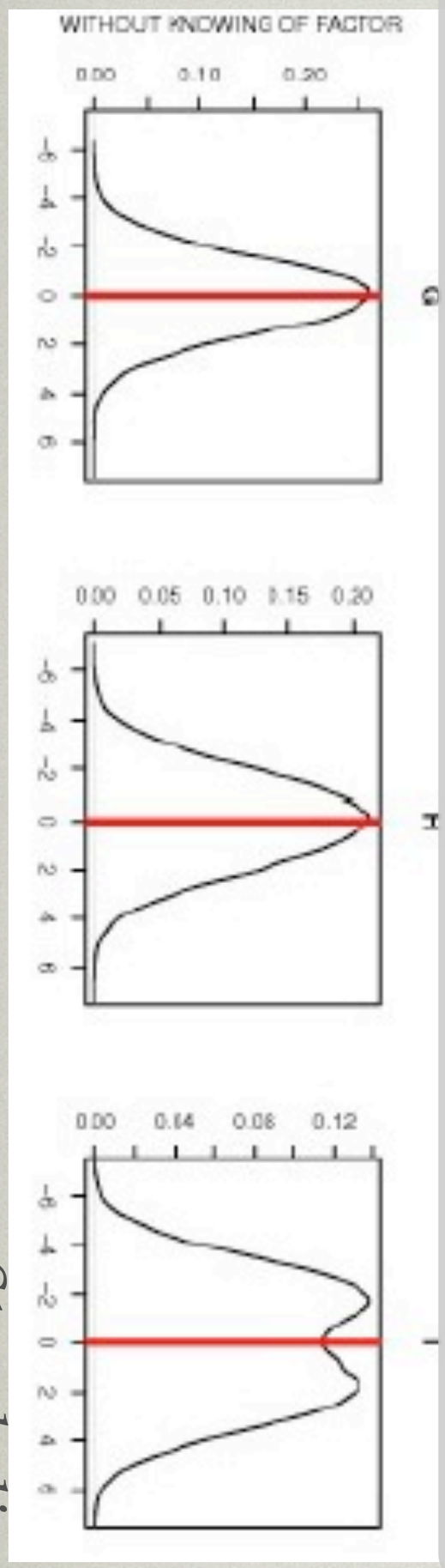
F=1 F=0




*Struchalin et al., 2010*

# True model

F=1 F=0

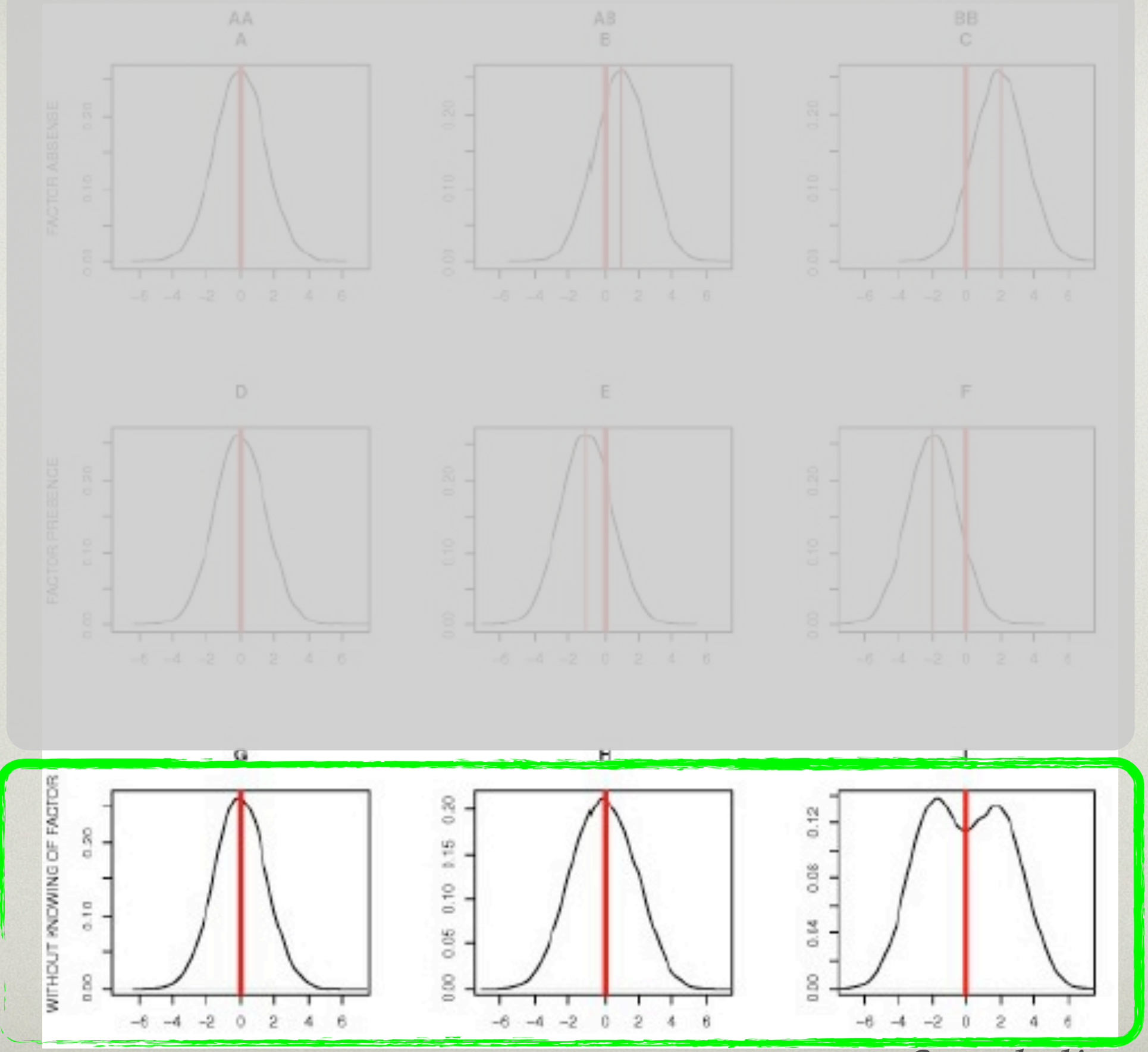



*Struchalin et al., 2010*

True model

F=1

F=0

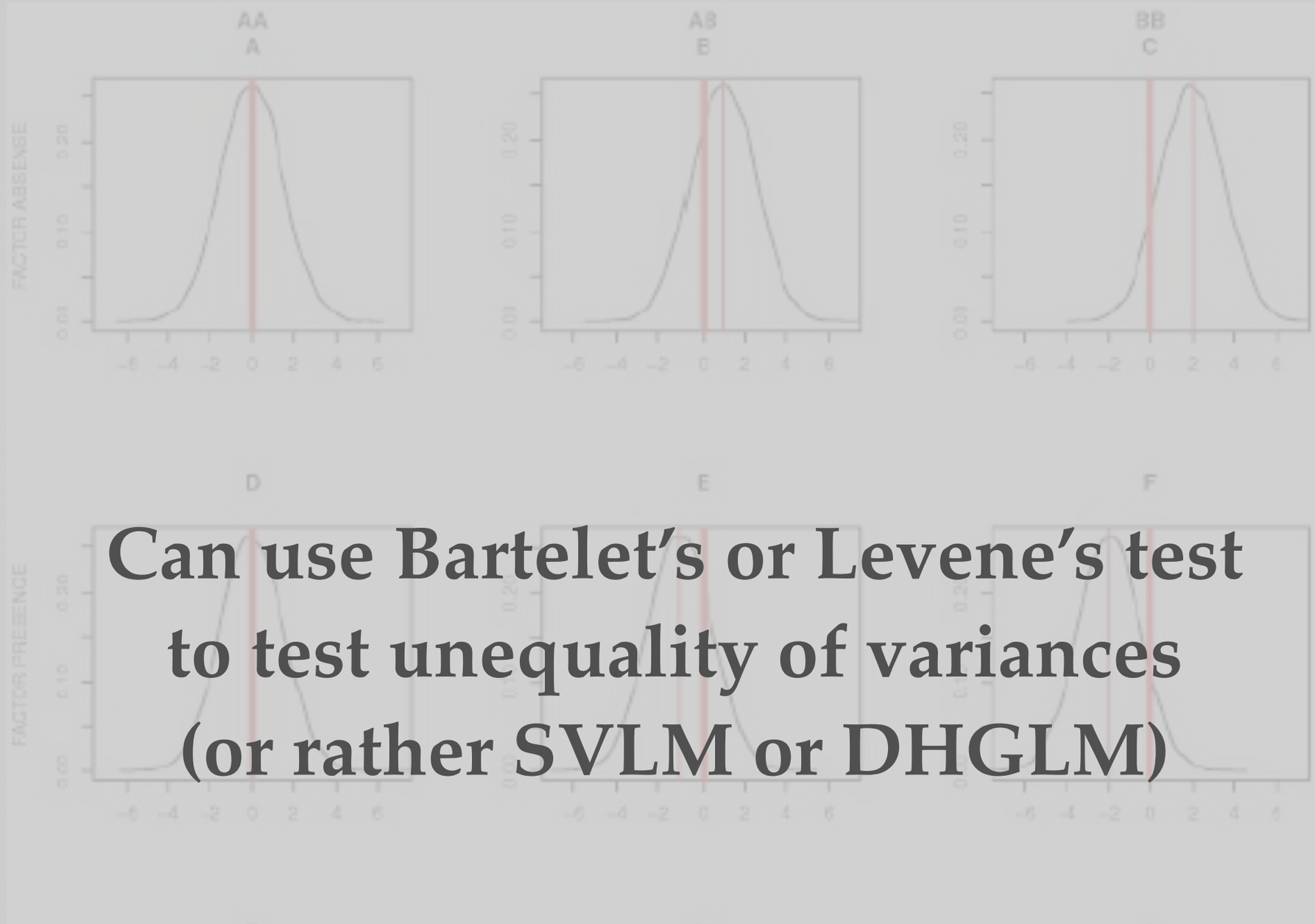



*Struchalin et al., 2010*

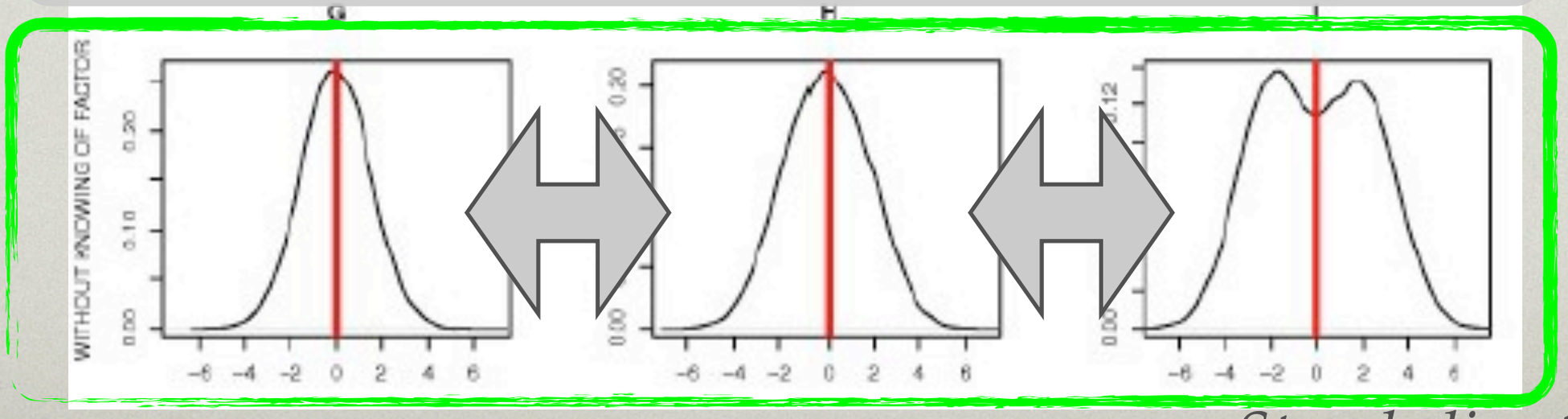
True model

F=0

F=1




**Can use Bartelet's or Levene's test to test unequality of variances (or rather SVLM or DHGLM)**




*Struchalin et al., 2010*

# Gx* method indeed works!

| Trait | Interacting SNP | MAF | Chr | Position (Kb) | Nearest Gene | Type | Covariable | Variance of A1A1* (N) | Variance of A1A2* (N) | Variance of A2A2* (N) | Levene's P-value | Interaction P-value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| CRP | | | | | | | | | | | | |
| | rs12753193 | 0.38 | 1 | 65942.3 | *LEPR* | - | BMI | 1.27 (8491) | 1.47 (10126) | 1.68 (3167) | 1.6E-29 | 7.2E-10 |
| sICAM-1 | | | | | | | | | | | | |
| | rs17999[illegible]9 | 0.11 | 19 | 10255.8 | *ICAM1* | Missense | Smoking | 6621 (17063) | 5316 (4421) | 4104 (300) | 2.1E-09 | 4.8E-09 |
| | rs738409 | 0.22 | 22 | 42656.1 | *PNPLA3* | Missense | BMI | 6087 (13098) | 6743 (6965) | 9205 (1110) | 1.9E-10 | 1.6E-07 |

*A1A1: Homozygous Major Allele; A1A2: Heterozygous; A2A2: Homozygous Minor Allele.
doi:10.1371/journal.pgen.1000981.t001

Pare *et al.*, PLoS Genet, 2010

Replicated by Struchalin *et al.*,
BMC Genet, 2010

SVLM method & VariABEL package:
Struchalin et al., BMC Genet., 2011

Methodology article

Highly accessed

Open Access

# An R package "VariABEL" for genome-wide searching of potentially interacting loci by testing genotypic variance heterogeneity

**Maksim V Struchalin**[1], **Najaf Amin**[1], **Paul HC Eilers**[2], **Cornelia M van Duijn**[1] and **Yurii S Aulchenko**[1,3]*

GenABEL.org

Home

## Primary links

- Packages
  - DatABEL
  - GenABEL
  - MetABEL
  - MixABEL
  - ParallABEL
  - PredictABEL
  - ProbABEL
  - VariABEL

## Packages

GenABEL, or *ABEL, is an umbrella name for a number of software packages aiming to facilitate statistical analyses of polymorphic genomes data. This is reach program set which now allows very flexible genome-wide association (GWA) analysis (GenABEL, ProbABEL, MixABEL), meta-analysis (MetABEL), parallelization of GWA analyses (ParallABEL), management of very large files (DatABEL), and facilitates evaluation of prediction (PredictABEL).

Most likely, you only need one of the packages for your specific task. Figure out one you need, install, and use! If you have questions, please refer to the port" section of this web-site.

# Rigorous treatment

Rönnegård *et al. Genetics Selection Evolution* 2010, **42**:8
http://www.gsejournal.org/content/42/1/8

GSE Genetics Selection Evolution

RESEARCH Open Access

## Genetic heterogeneity of residual variance - estimation of variance components using double hierarchical generalized linear models

Lars Rönnegård[1,2*], Majbritt Felleki[1,2], Freddy Fikse[2], Herman A Mulder[3], Erling Strandberg[2]

# Conclusions - variance heterogeneity

- Variance heterogeneity test is an interesting approach to prioritize markers for interaction testing
- Note that strictly speaking...
  - Negative results do not mean there is NO interaction (power, special scenarios)
  - Positive results should be interpreted with caution

# Conclusions - general

- Some examples of GxE in complex traits start appearing
- Genetic interactions are tough
- Genetic interactions appear to be less common and/or less strong than we have hoped for
- ??? Genetic interactions may be more pronounced for rare(r) variants

# Advanced topics (GEO3?)

- Inter-genic interactions
- Parent of origin effects and gene-gene interactions
- Variance heterogeneity - testing if identified interaction explains observed variance heterogeneity
- Variance heterogeneity - rigorous treatment of variance heterogeneity through DHGLM