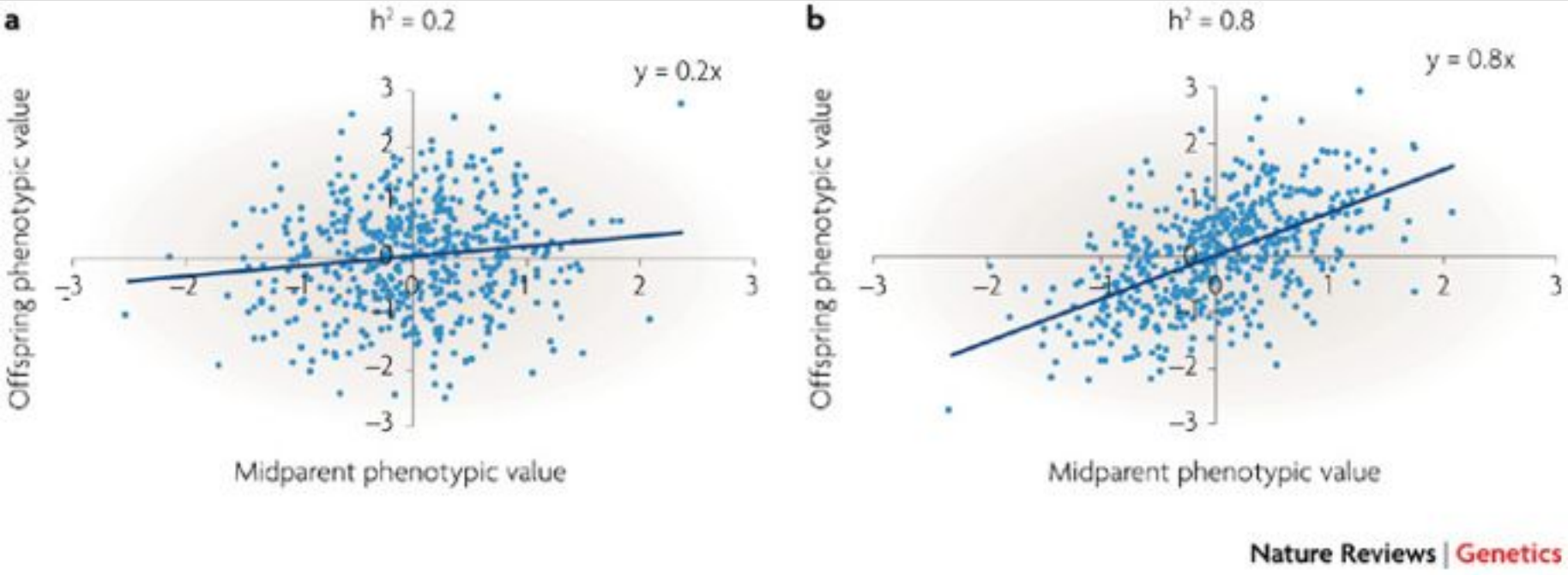
GWAS 8

Matti Pirinen
University of Helsinki
25.11.2020

HERITABILITY

- Proportion of phenotypic variance explained by variation in genome
- Depends on the population and point of time because environmental variance can vary
- Narrow sense heritability h^2: variance explained by the additive effects of the variants
- Broad sense heritability H^2 : variance explained by all genetic variation

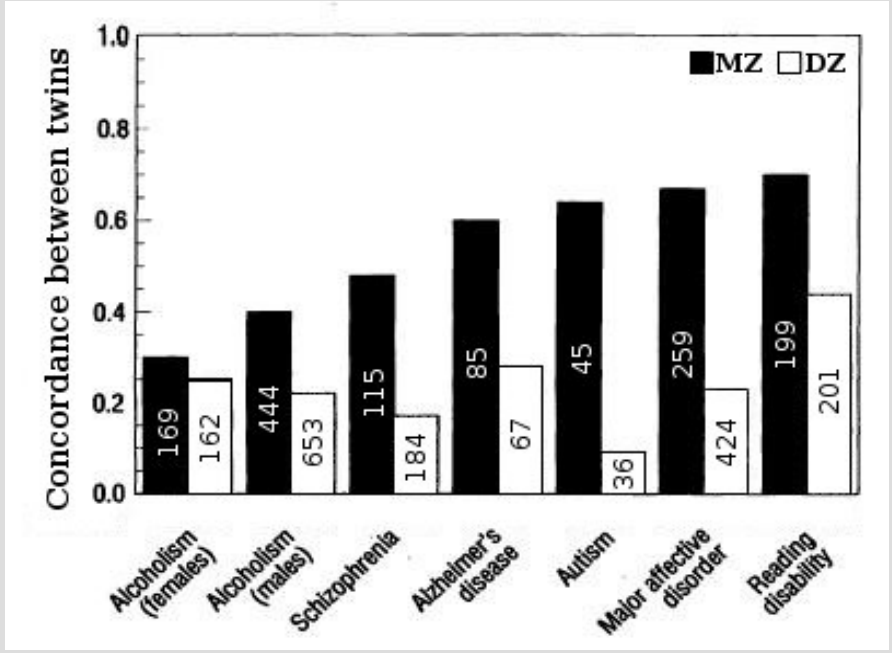
HERITABILITY CAUSES PHENOTYPIC SIMILARITY AMONG RELATIVES



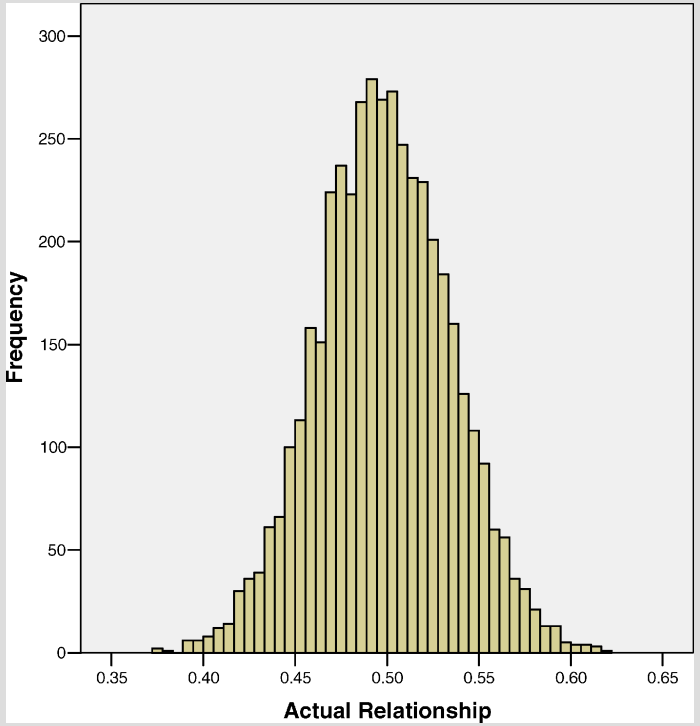
The slope of the regression line is an estimate of the narrow-sense heritability for traits with a heritability of 0.2 (**a**) and 0.8 (**b**) and phenotypic variance of 1. The variances of the observations about the regression line are 0.98 (**a**) and 0.68 (**b**), demonstrating that the average phenotypic value of the parents (midparent phenotypic value) is a better predictor of the offspring phenotypic value if heritability is high.

Visscher et al. 2008

HOW HERITABILITY HAS BEEN ESTIMATED ?



Compare concordance in monozygotic twins (share full genome) to that of dizygotic twins (share ~50%). Under (strong) assumptions the difference estimates heritability.



Whether full-sib pairs that share more of their genomes also have more similar phenotypes?
Heritability estimate for height from 3375 pairs of sibs was 0.80 (0.46 – 085). (Visscher et al. 2006 PLoS Genetics)

GWAS RESULTS

Variance explained (causally) by variant:

$$\mathrm{Var}(x\lambda) = \lambda^2 \mathrm{Var}(x) \approx 2f(1-f)\lambda^2 = (\lambda^*)^2$$

Variance explained (marginally) by variant:

$$\mathrm{Var}(x\beta) = \beta^2 \mathrm{Var}(x) \approx 2f(1-f)\beta^2 = (\beta^*)^2$$

Variance explained by a region

$$\begin{aligned} h^2_{\mathrm{reg}} &= \mathrm{Var}(\boldsymbol{x}^{*T}\boldsymbol{\lambda}^*) = \boldsymbol{\lambda}^{*T}\mathrm{Var}(\boldsymbol{x}^*)\boldsymbol{\lambda}^* = \boldsymbol{\lambda}^{*T}\boldsymbol{R}\boldsymbol{\lambda}^* \\ &= (\boldsymbol{R}^{-1}\boldsymbol{\beta}^*)^T\boldsymbol{R}(\boldsymbol{R}^{-1}\boldsymbol{\beta}^*) = \boldsymbol{\beta}^{*T}\boldsymbol{R}^{-1}\boldsymbol{\beta}^*, \end{aligned}$$

MISSING HERITABILITY PROBLEM

- When summing up heritability from GWAS loci, it accounts for only a small part of heritability estimated in twins or siblings
- Where is the rest?
  - Twin estimates are too high?
  - Traits are highly polygenic with so small effects that they have not been picked by GWAS (yet)?

POLYGENICITY

MPV
Mean platelet volume

MCH
Mean corpuscular hemoglobin

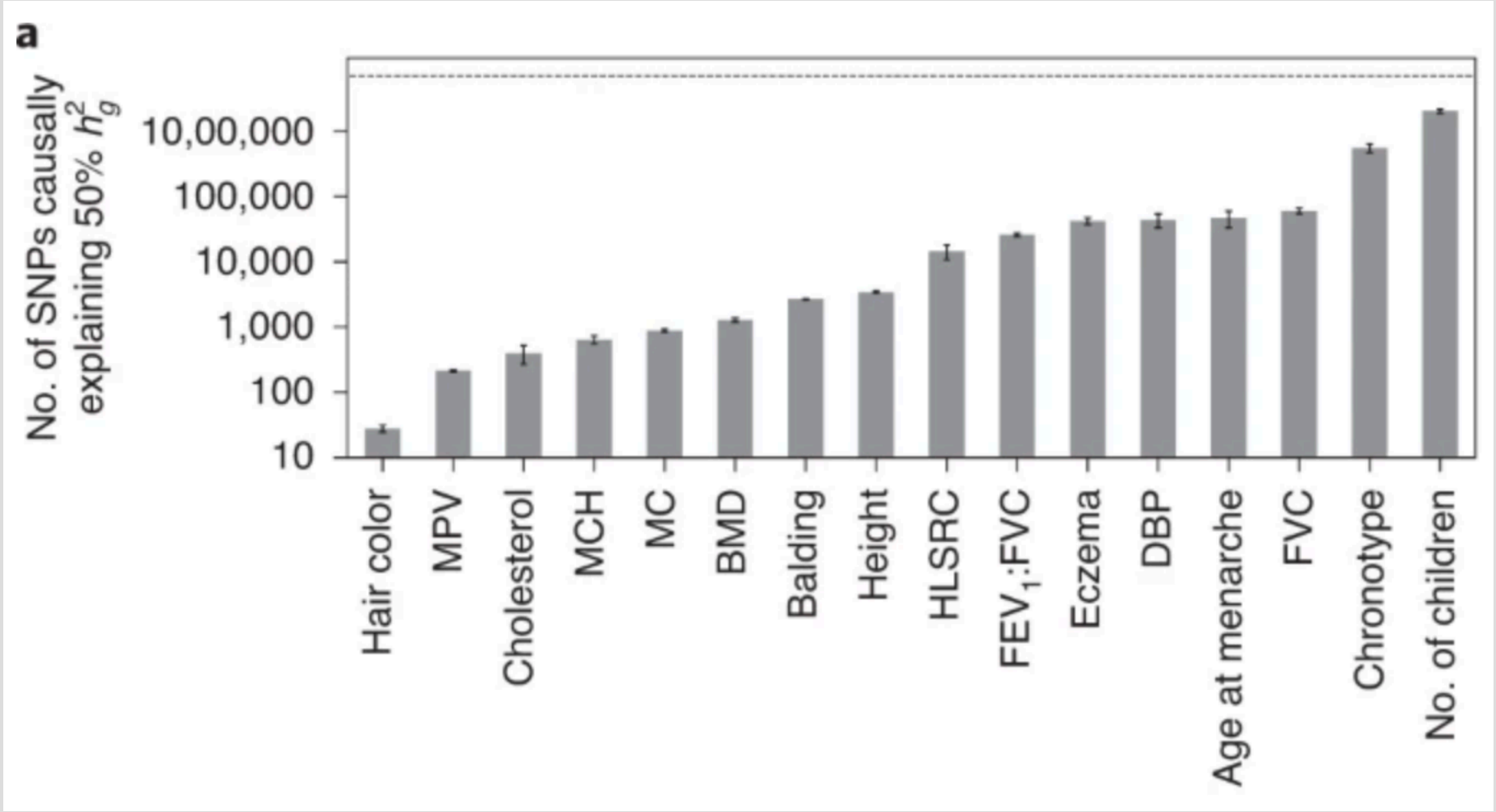
BMD:
Bone mineral density

HLSRC:
High light scatt. retinoc. count

FEV
Forced expiratory volume

FVC
Forced vital capacity

DBP
Diastolic blood pressure

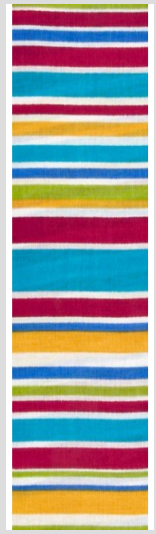


Weissbrod et al. 2020 Nat Gen

LINEAR MIXED MODEL TO ESTIMATE VARIANCE COMPONENT

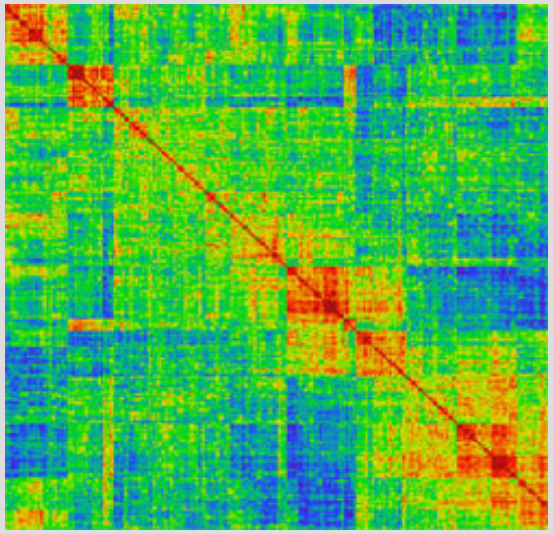
$$\mathbf{Y} \sim \mathcal{N}(0, h\mathbf{R} + (1-h)\mathbf{I})$$

For height in Finns
we estimate h ~ 50%



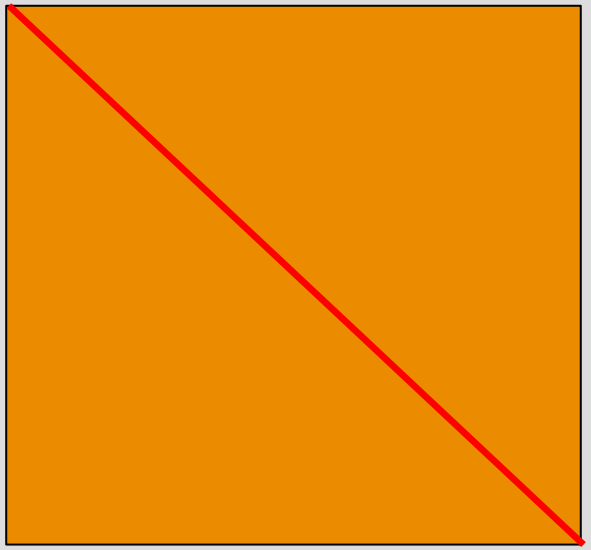
Y

~ h



SNP relatedness

+(1-h)



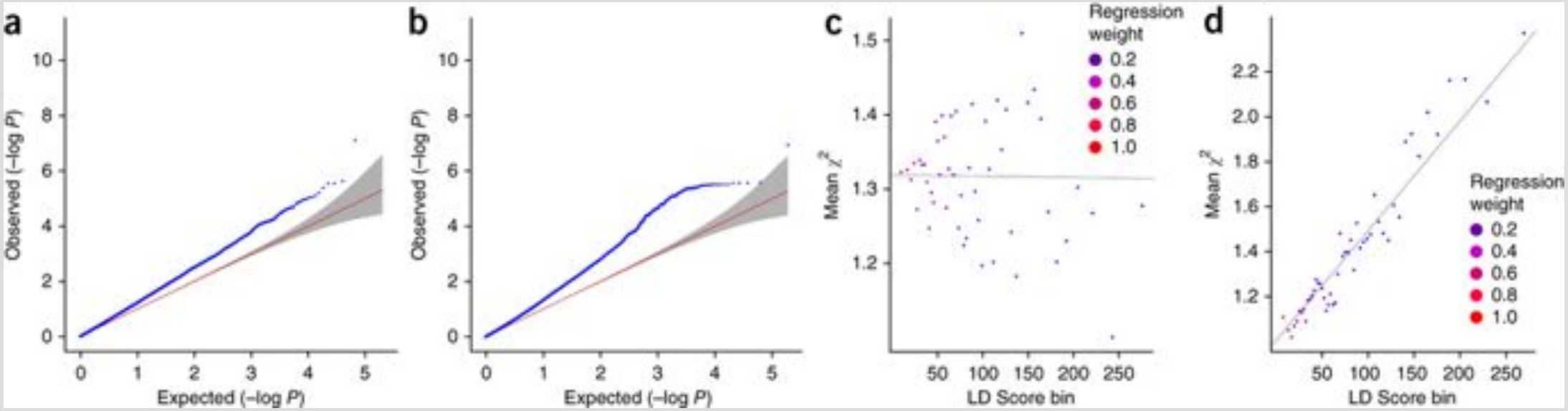
Independent environment

Variation in phenotype ~ Pattern of genetic similarity + random noise uncorrelated between individuals

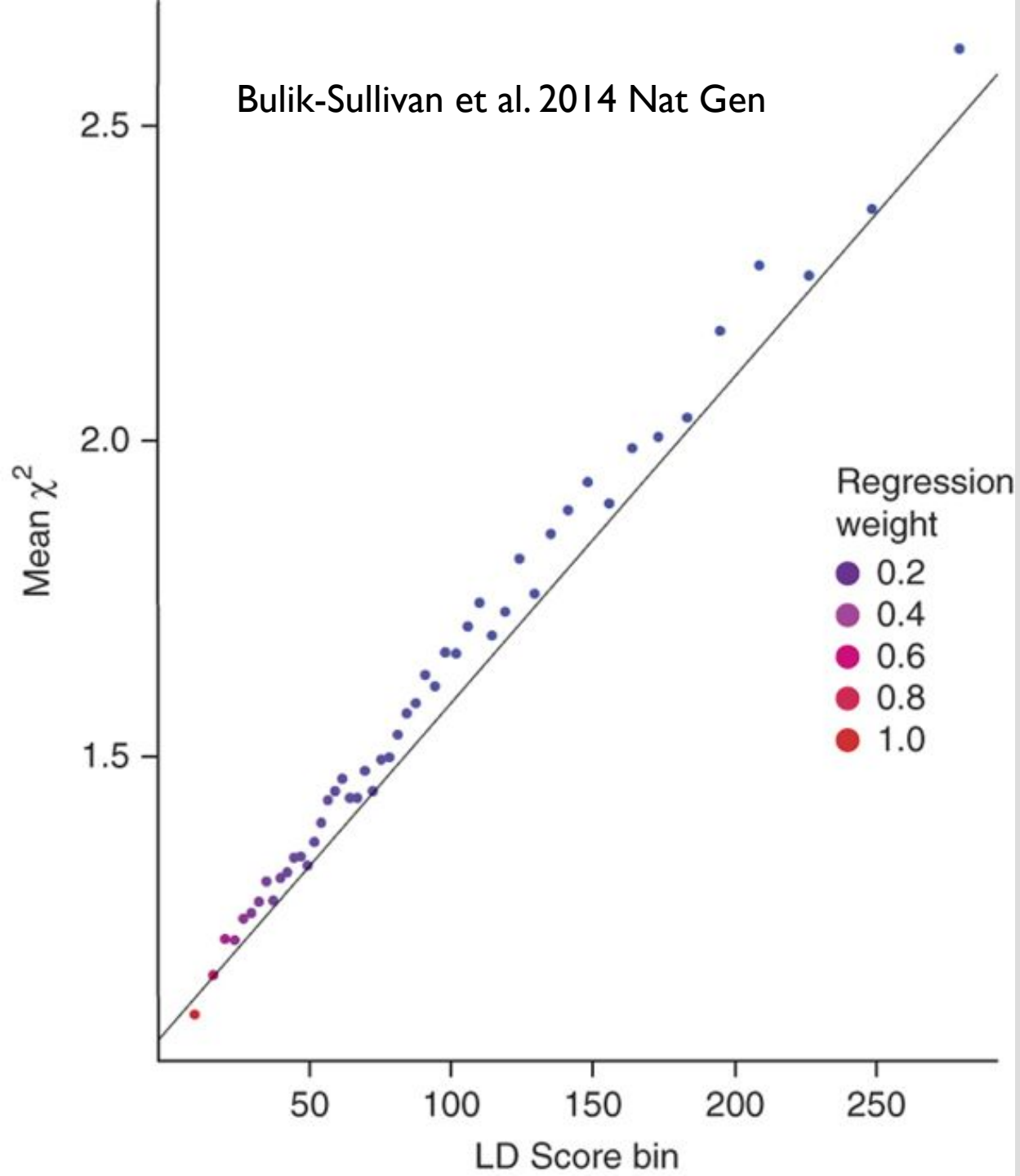
Parameter h measures how well phenotypic variation is explainable by pattern of genetic similarity

Bulik-Sullivan et al.
2014 Nat Gen

LD SCORE REGRESSION



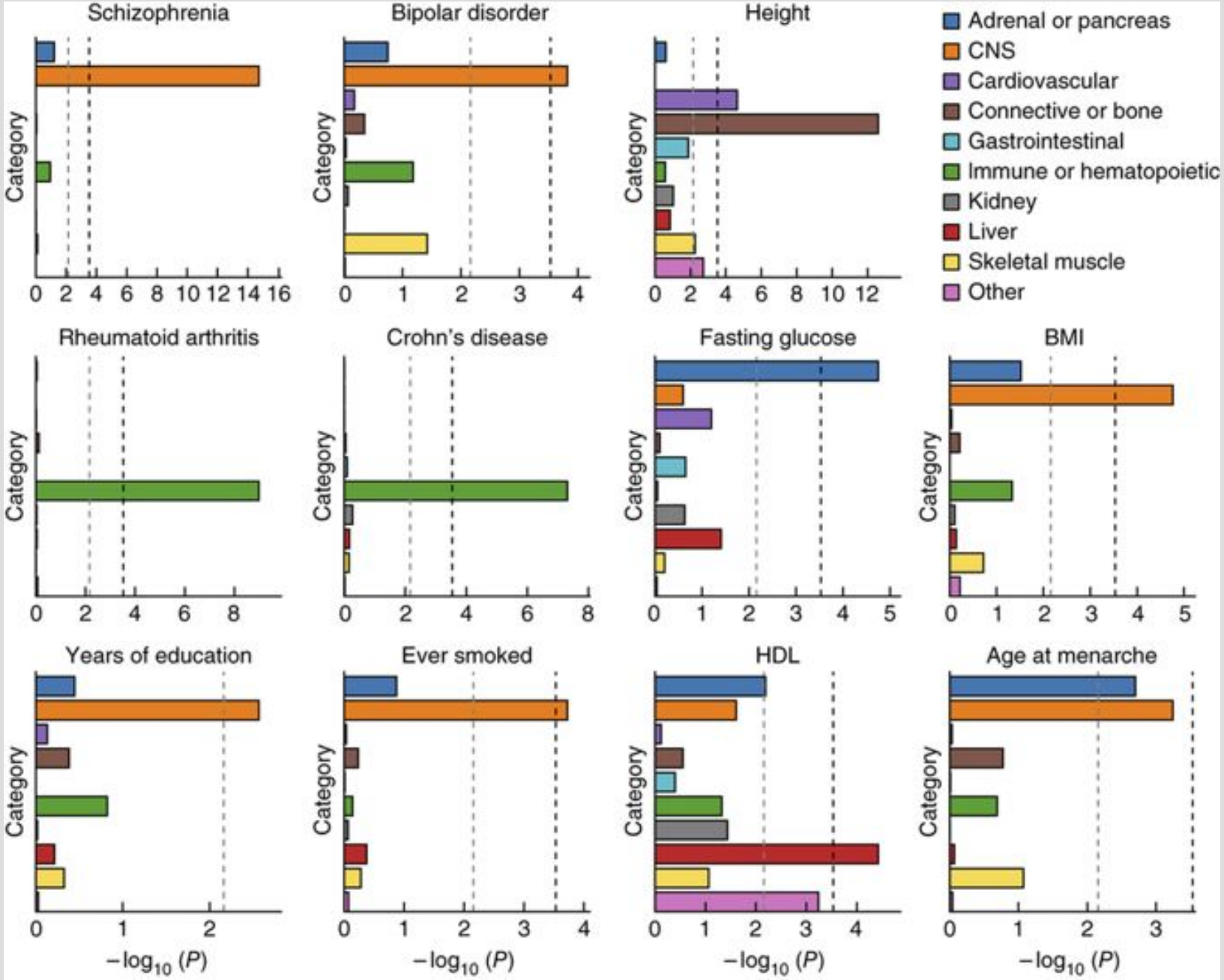
(**a**) Quantile-quantile plot with population stratification (λ_{GC} = 1.32, LD Score regression intercept = 1.30). (**b**) Quantile-quantile plot with a polygenic genetic architecture where 0.1% of SNPs are causal (λ_{GC} = 1.32, LD Score regression intercept = 1.006). (**c**) LD Score plot with population stratification. Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ^2 statistic of variants in that quantile. Colors correspond to regression weights, with red indicating large weight. The black line is the LD Score regression line. (**d**) LD Score plot as in **c** but with polygenic genetic architecture.



LDSC ON SCHIZOPHRENIA GWAS RESULTS

Each point represents an LD Score quantile, where the *x* coordinate of the point is the mean LD Score of variants in that quantile and the *y* coordinate is the mean χ^2 statistic of variants in that quantile in the most recent schizophrenia meta-analysis. Colors correspond to regression weights, with red indicating large weight and blue indicating small weight. The black line is the LD Score regression line. The line appears to fall below the points on the right because this is a weighted regression in which the points on the left receive the largest weights

LDSC FOR PARTITIONING HERITABILITY



Finucane et al. 2015 Nat Gen

Multiple regression of GWAS statistics on LD-scores with different categories of variants.

Here the groups are variants linked to certain cell types.