

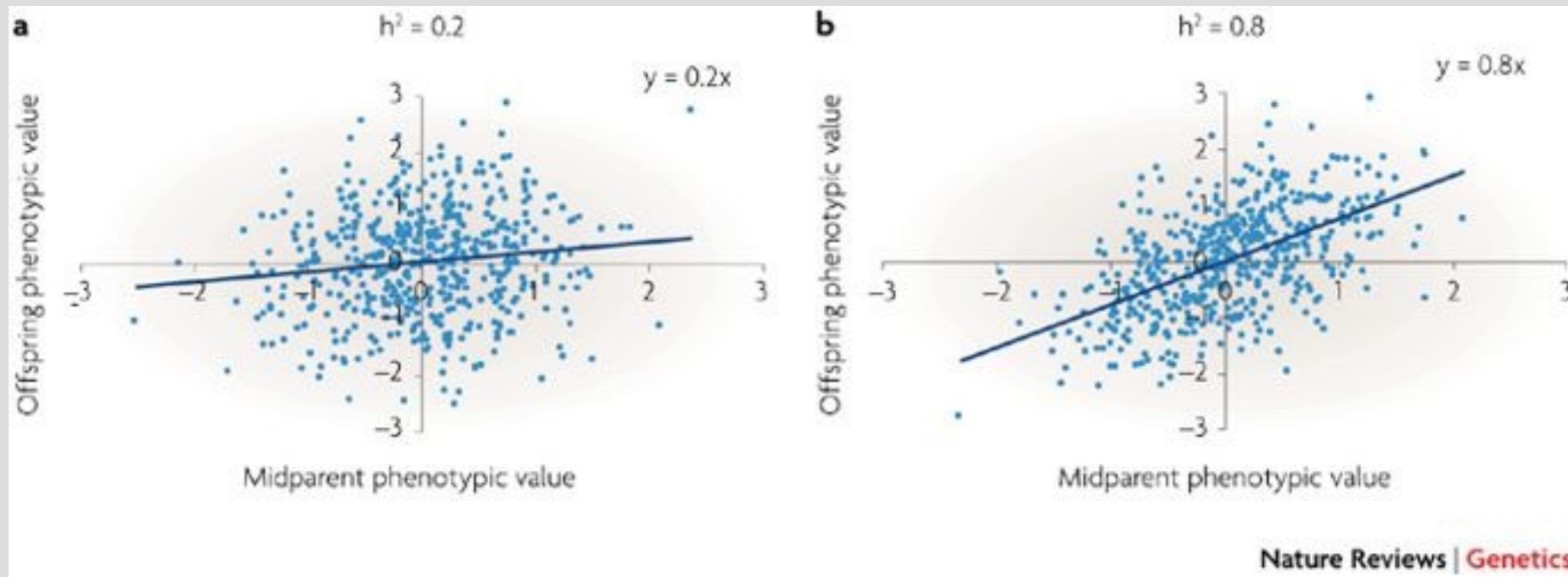
GWAS 8

Matti Pirinen
University of Helsinki
Week 5

HERITABILITY

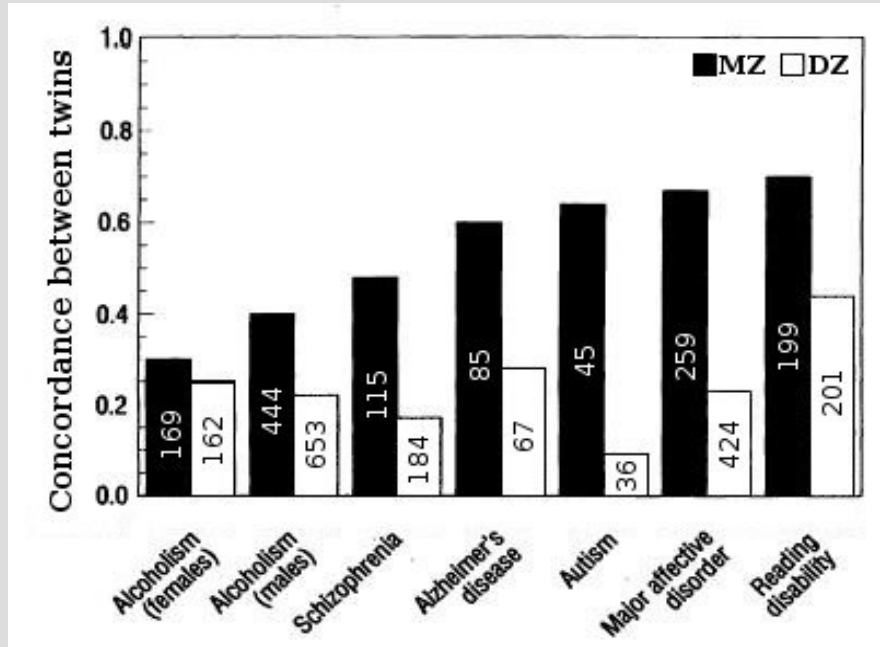
- Proportion of phenotypic variance explained by variation in genome
- Depends on the population and point of time because environmental variance can vary
- Narrow sense heritability h^2 : variance explained by the additive effects of the variants
- Broad sense heritability H^2 : variance explained by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

HERITABILITY CAUSES PHENOTYPIC SIMILARITY AMONG RELATIVES

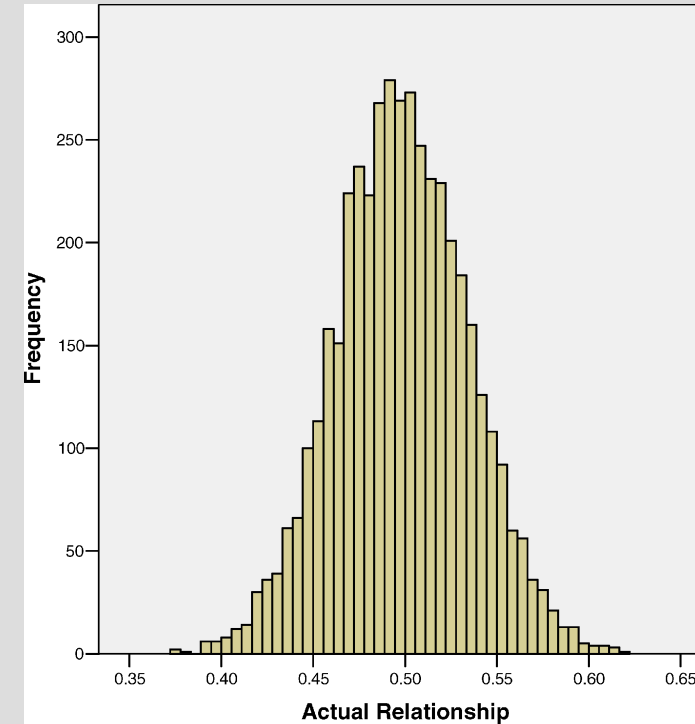


The slope of the regression line is an estimate of the narrow-sense heritability for traits with a heritability of 0.2 (**a**) and 0.8 (**b**) and phenotypic variance of 1. The variances of the observations about the regression line are 0.98 (**a**) and 0.68 (**b**), demonstrating that the average phenotypic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

HOW HAS HERITABILITY BEEN ESTIMATED ?



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



Do the 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 – 0.85). (Visscher et al. 2006 PLoS Genetics)

GWAS RESULTS

Variance explained (causally) by a variant:

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

Variance explained (marginally) by variant:

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

Variance explained by a region

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

x is genotype

f is minor allele frequency

λ is causal effect

β is marginal effect

* means standardized effect
or standardized genotype

MISSING HERITABILITY PROBLEM

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

POLYGENICITY

MPV

Mean platelet volume

MCH

Mean corpuscular hemoglobin

BMD:

Bone mineral density

HLSRC:

High light scatt. retinoc. count

FEV

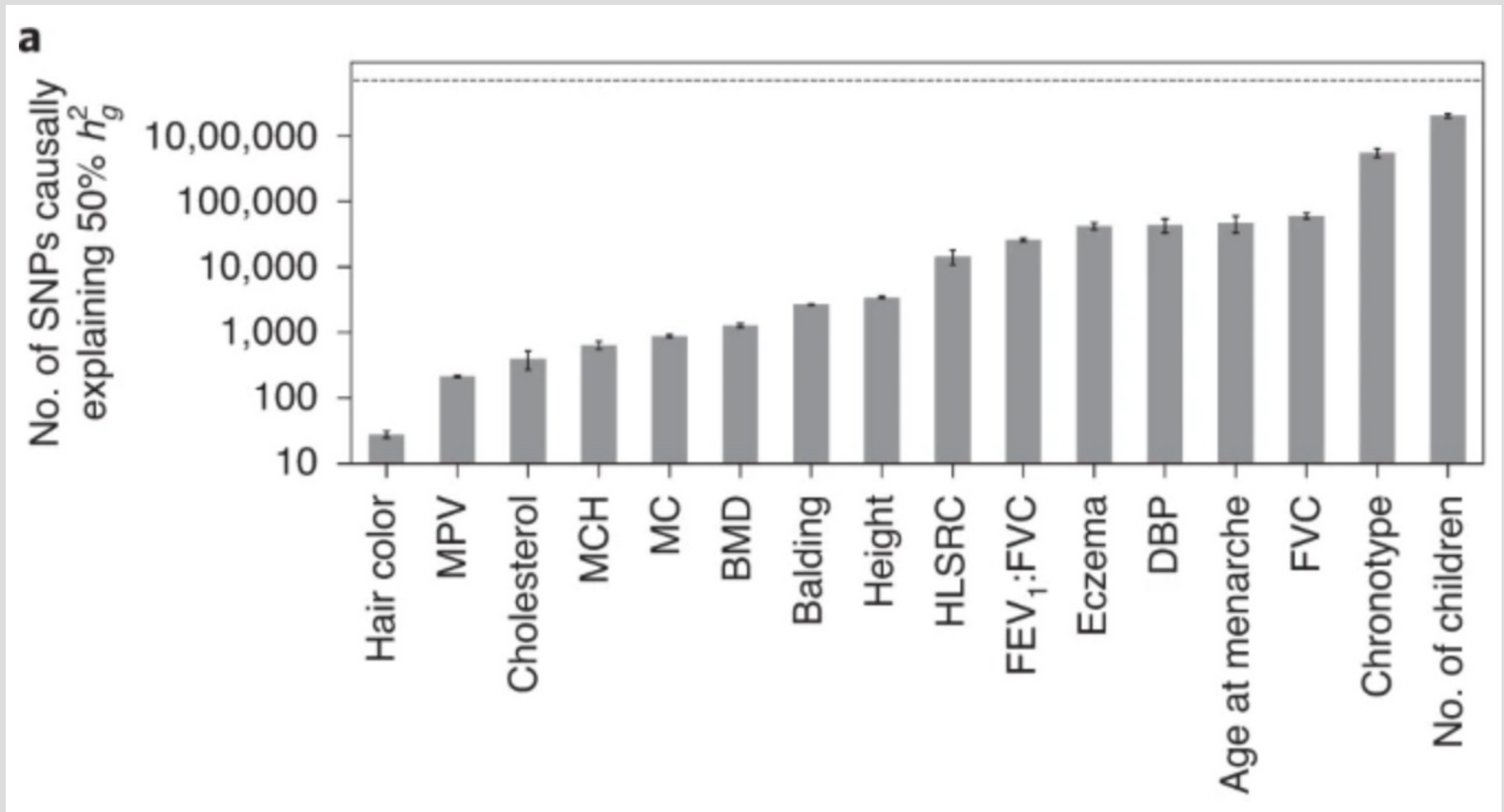
Forced expiratory volume

FVC

Forced vital capacity

DBP

Diastolic blood pressure

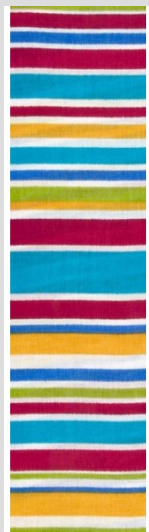


Weissbrod et al. 2020 Nat Gen

LINEAR MIXED MODEL TO ESTIMATE VARIANCE COMPONENT

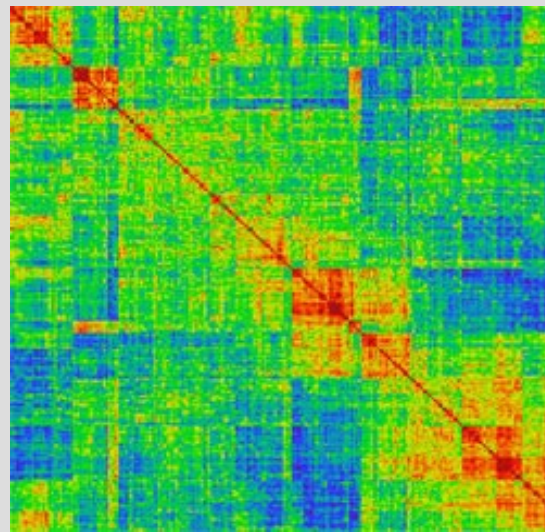
$$\mathbf{Y} \sim \mathcal{N}(0, h\mathbf{R} + (1-h)\mathbf{I})$$

For height in Finns
we estimate $h \sim 50\%$



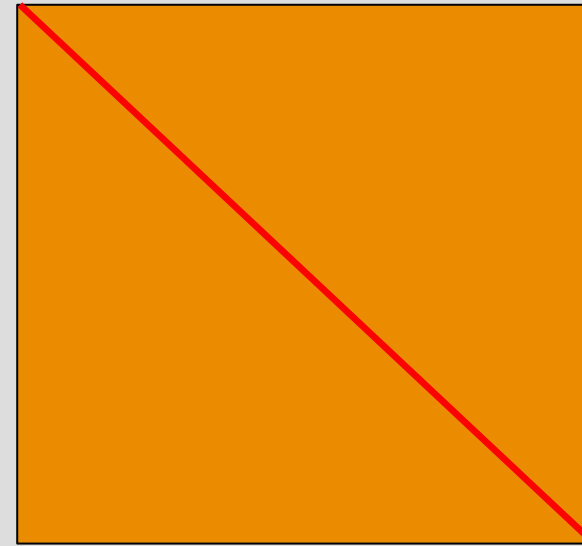
\mathbf{Y}

$\sim h$



SNP relatedness

$+ (1-h)$

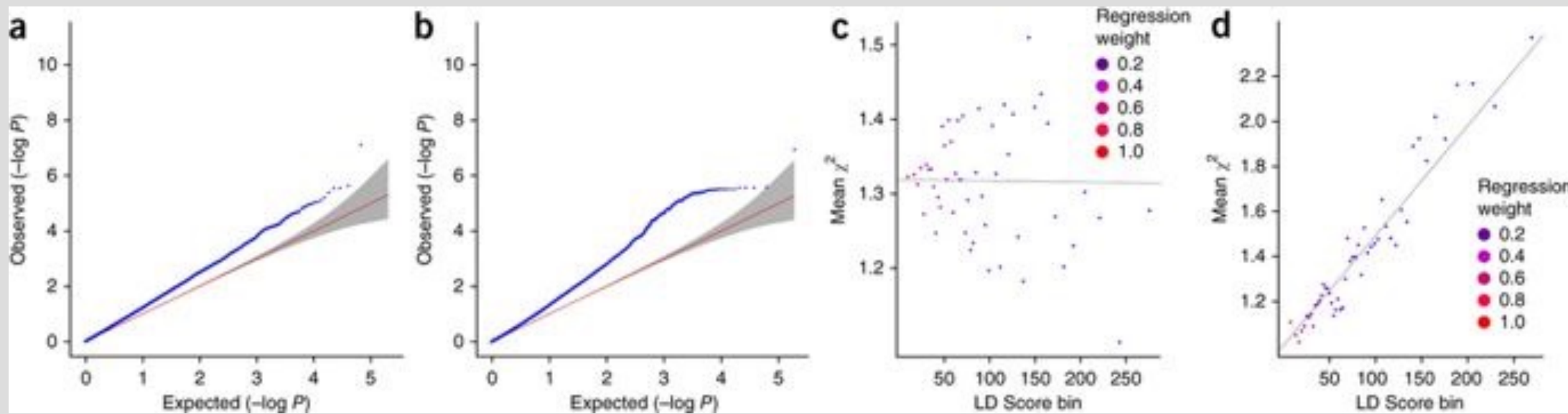


Independent environment

variation in phenotype \sim pattern of genetic similarity $+$ random noise uncorrelated between individuals

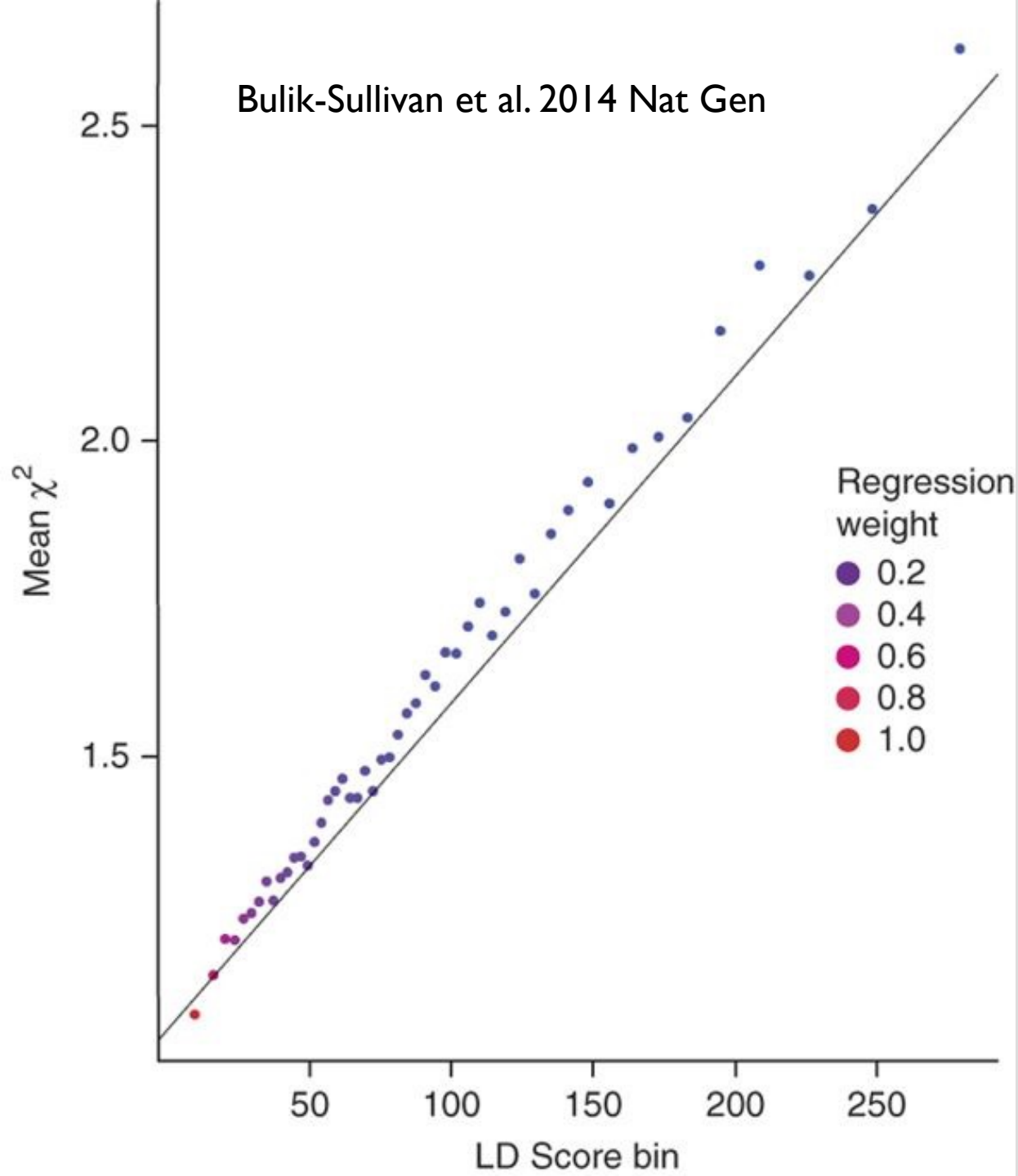
Parameter h measures how well phenotypic variation is explainable by pattern of genetic similarity

LD SCORE REGRESSION



- (a) Quantile-quantile plot with population stratification ($\lambda_{GC} = 1.32$, LD score regression intercept = 1.30). (b) Quantile-quantile plot with a polygenic genetic architecture where 0.1% of SNPs are causal ($\lambda_{GC} = 1.32$, 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.

Bulik-Sullivan et al. 2014 Nat Gen



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

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

Here the groups are variants linked to gene activity in certain cell types.

X-axis shows enrichment P-values, i.e., How much statistical evidence there is that the GWAS signal of a phenotype is enriched in a particular type of variants.

