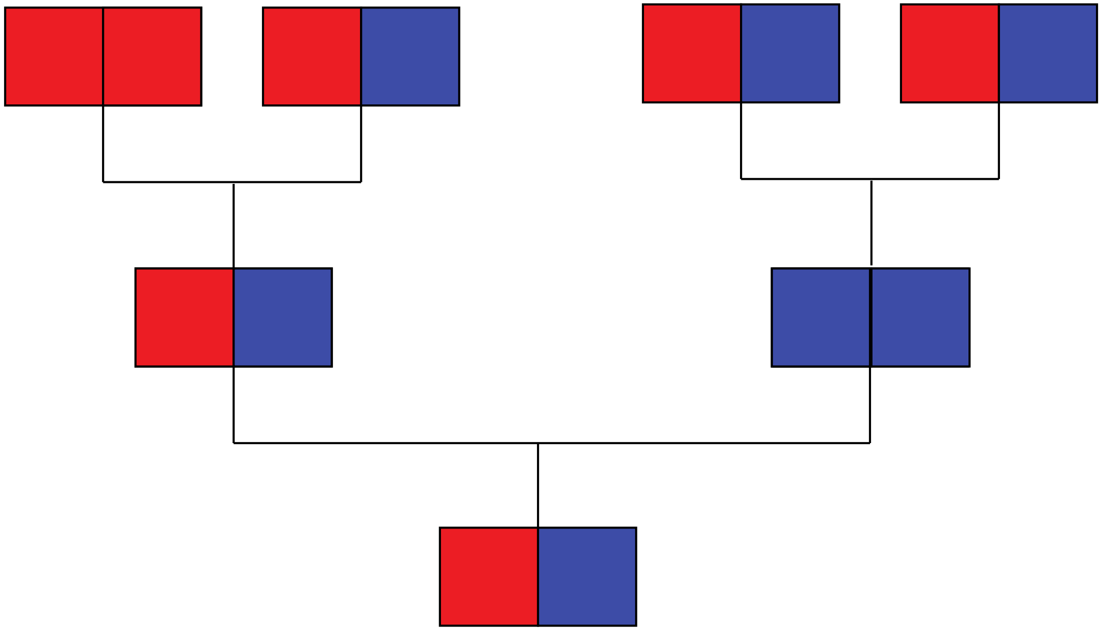
Disentangling Nature and Nurture Using Genomic Family Data



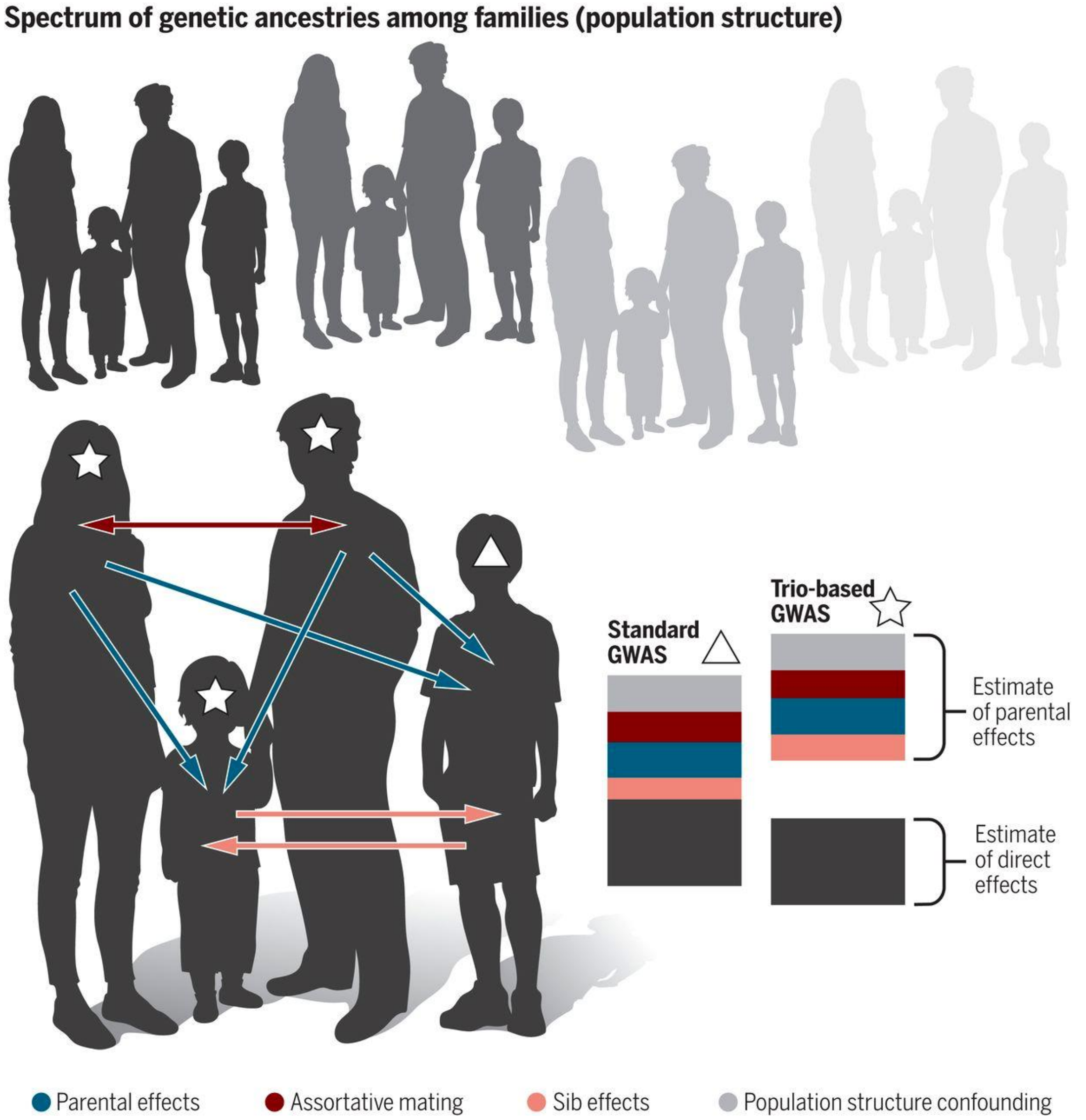
Alexander Young

SSGAC & USC

Blog: https://geneticvariance.wordpress.com/

@alexTISyoung

Deconstructing the sources of genotype-phenotype associations



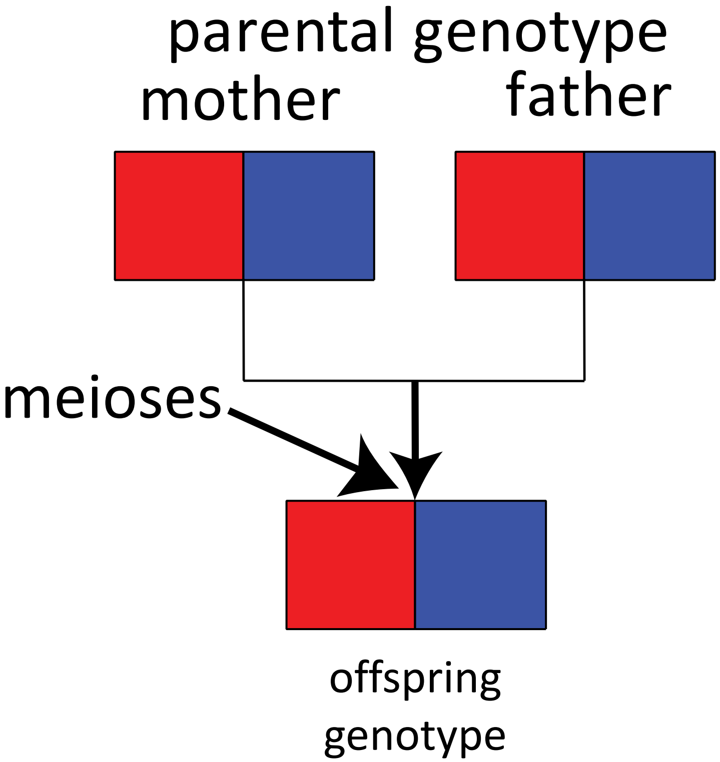
Young et al., Deconstructing the sources of genotype-phenotype associations in humans. *Science* (2020)

Separating out direct genetic effects

The transmission disequilibrium test (TDT):

Parental alleles	**A**	**T**
Expected proportion in offspring	50%	50%
Observed proportion	60%	40%

Dependencies between genes and environment



parental genotype

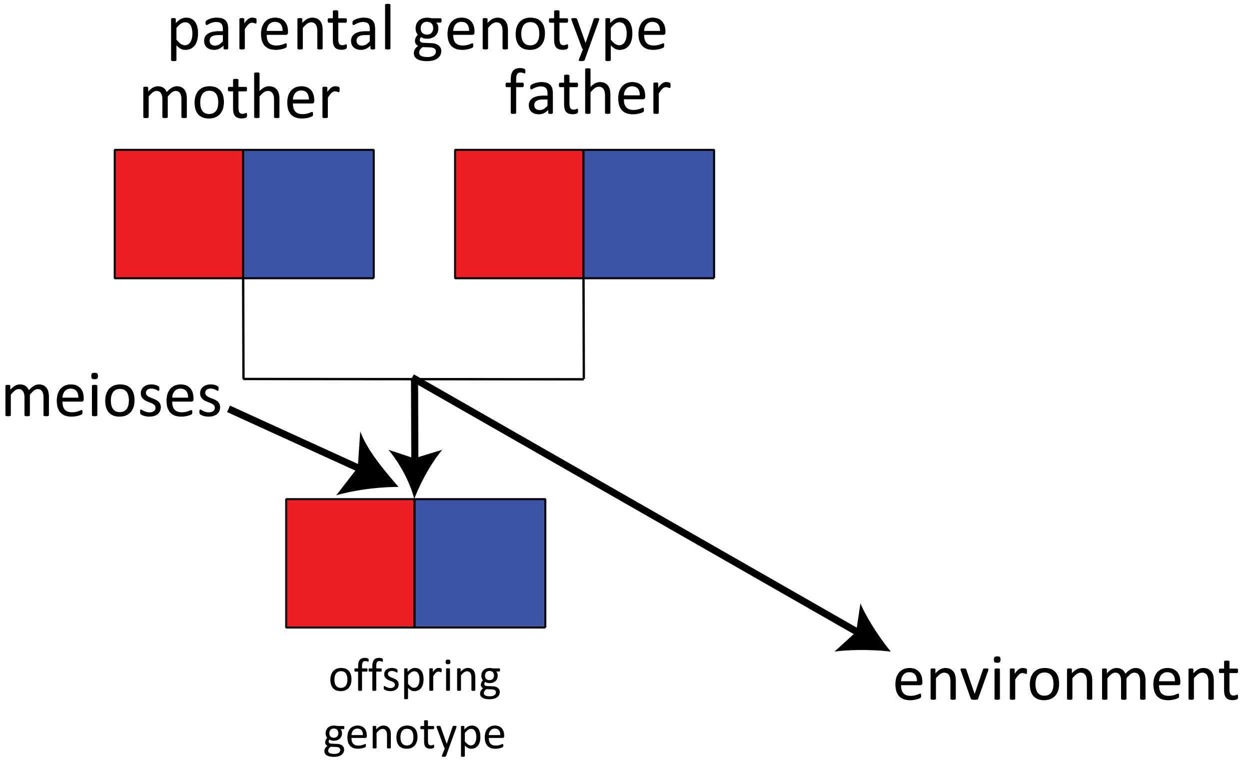
mother

father

meioses

offspring genotype

Dependencies between genes and environment



parental genotype

mother

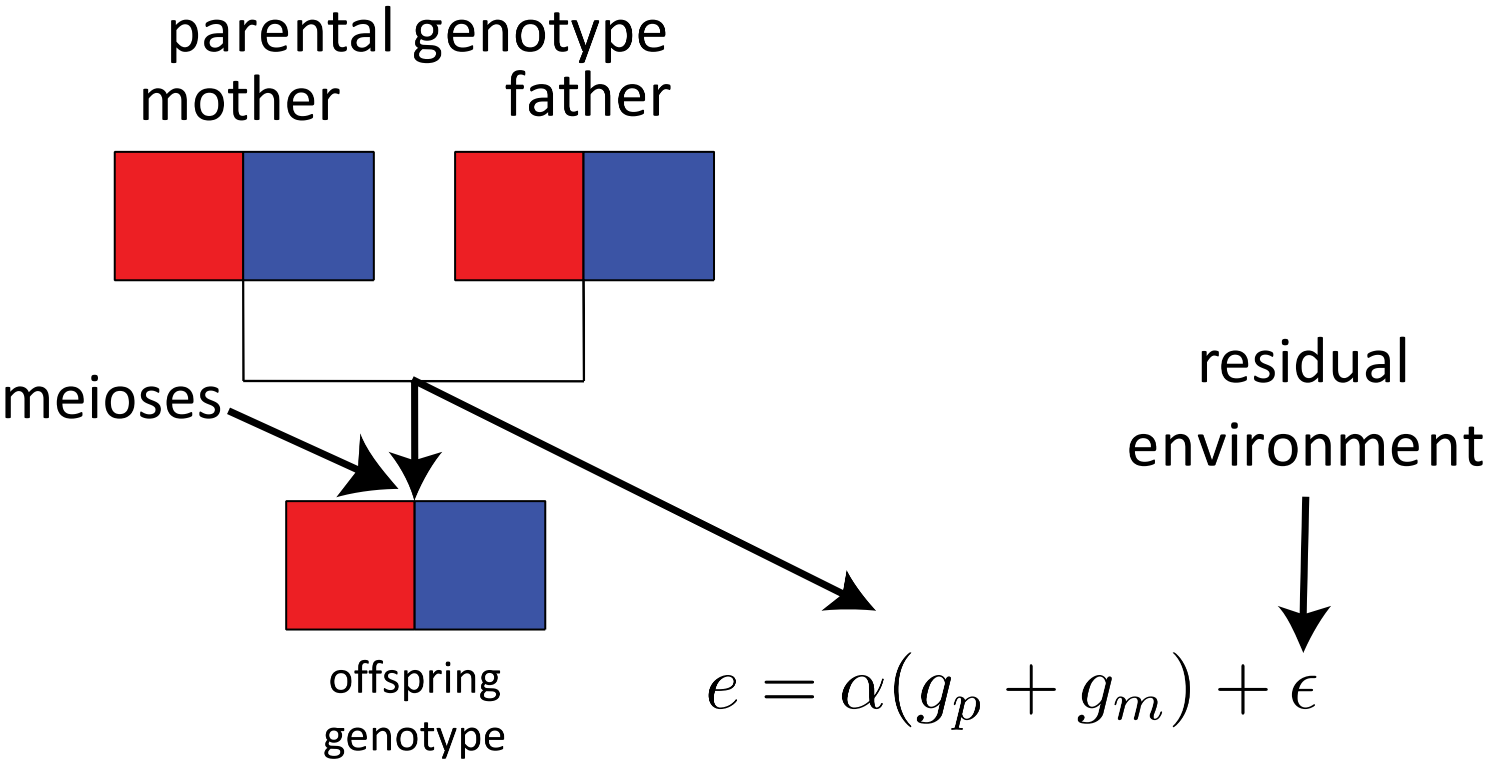
father

meioses

offspring genotype

environment

Dependencies between genes and environment



parental genotype

mother

father

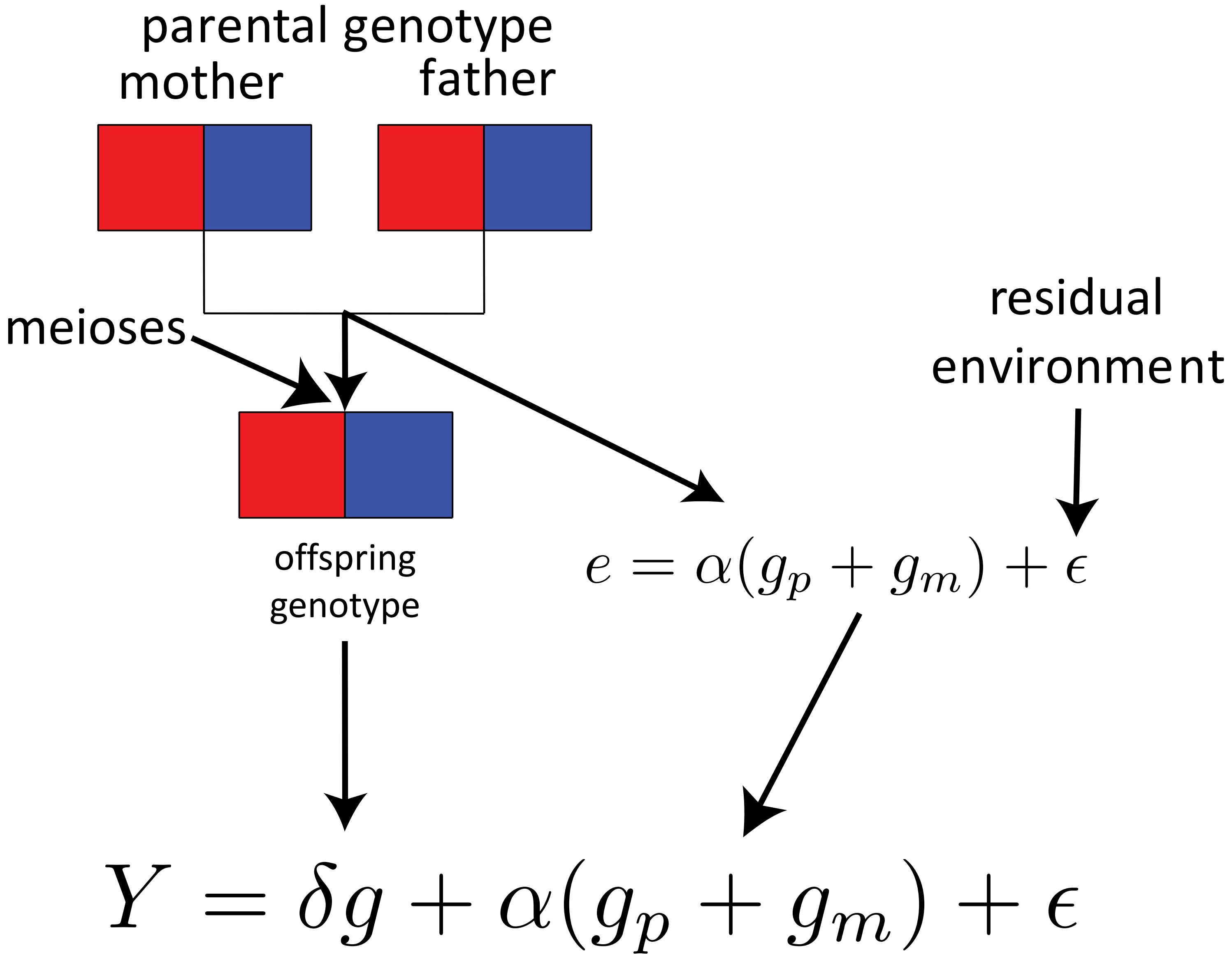
meioses

offspring genotype

residual environment

$$e = \alpha(g_p + g_m) + \epsilon$$

Dependencies between genes and environment



parental genotype

mother

father

meioses

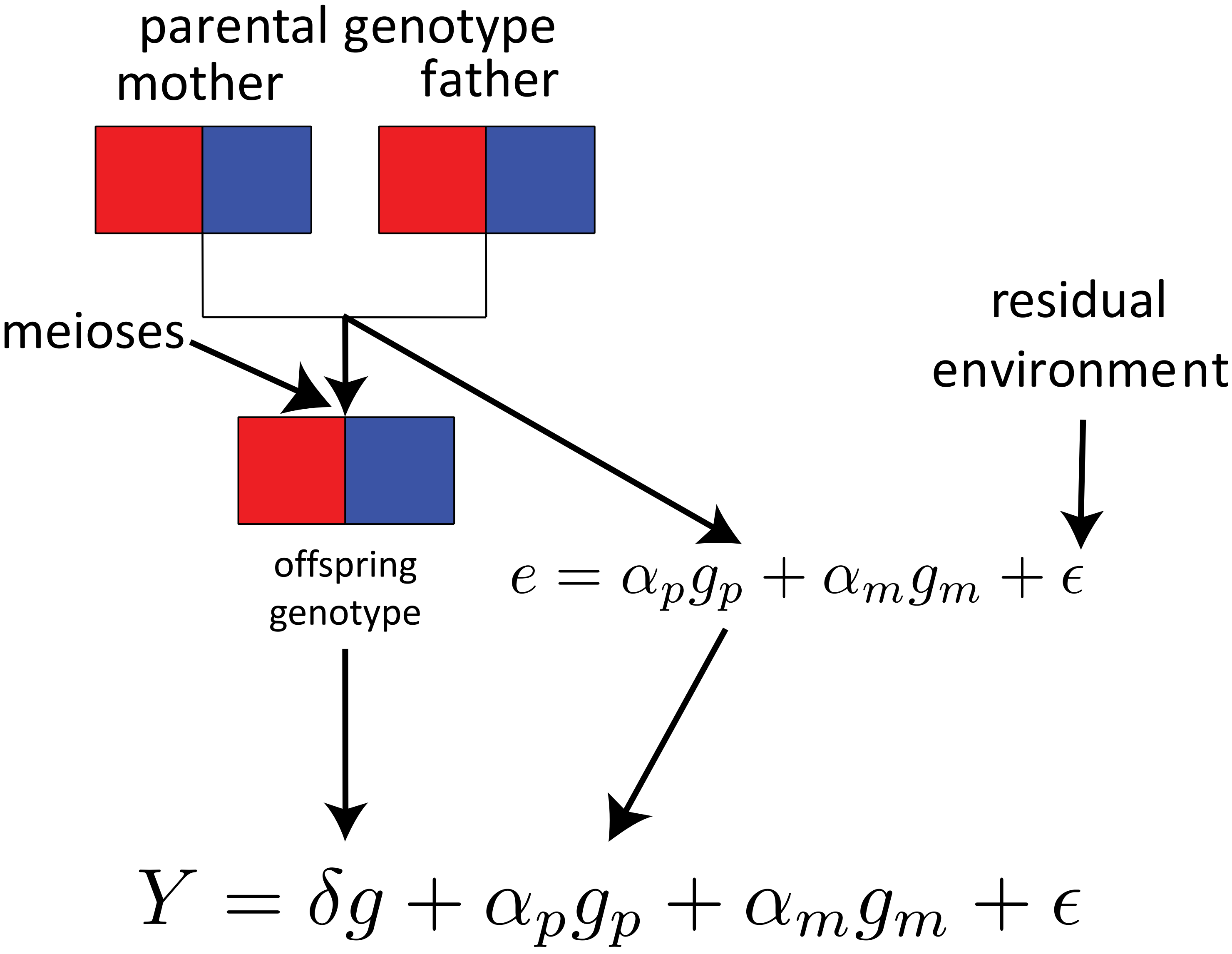
offspring genotype

residual environment

$$e = \alpha(g_p + g_m) + \epsilon$$

$$Y = \delta g + \alpha(g_p + g_m) + \epsilon$$

Dependencies between genes and environment



parental genotype

mother

father

meioses

residual environment

offspring genotype

$$e = \alpha_p g_p + \alpha_m g_m + \epsilon$$

$$Y = \delta g + \alpha_p g_p + \alpha_m g_m + \epsilon$$

RESEARCH

HUMAN GENOMICS

The nature of nurture: Effects of parental genotypes

Augustine Kong,[1,2,3]* Gudmar Thorleifsson,[1] Michael L. Frigge,[1]
Bjarni J. Vilhjalmsson,[4,5] Alexander I. Young,[1,2,6] Thorgeir E. Thorgeirsson,[1]
Stefania Benonisdottir,[1] Asmundur Oddsson,[1] Bjarni V. Halldorsson,[1] Gisli Masson,[1]
Daniel F. Gudbjartsson,[1,3] Agnar Helgason,[1,7] Gyda Bjornsdottir,[1]
Unnur Thorsteinsdottir,[1,8] Kari Stefansson[1,8]*

Sequence variants in the parental genomes that are not transmitted to a child (the proband) are often ignored in genetic studies. Here we show that nontransmitted alleles can affect a child through their impacts on the parents and other relatives, a phenomenon we call "genetic nurture." Using results from a meta-analysis of educational attainment, we find that the polygenic score computed for the nontransmitted alleles of 21,637 probands with at least one parent genotyped has an estimated effect on the educational attainment of the proband that is 29.9% ($P = 1.6 \times 10^{-14}$) of that of the transmitted polygenic score. Genetic nurturing effects of this polygenic score extend to other traits. Paternal and maternal polygenic scores have similar effects on educational attainment, but mothers contribute more than fathers to nutrition- and heath-related traits.

Direct and indirect effects of educational attainment polygenic score

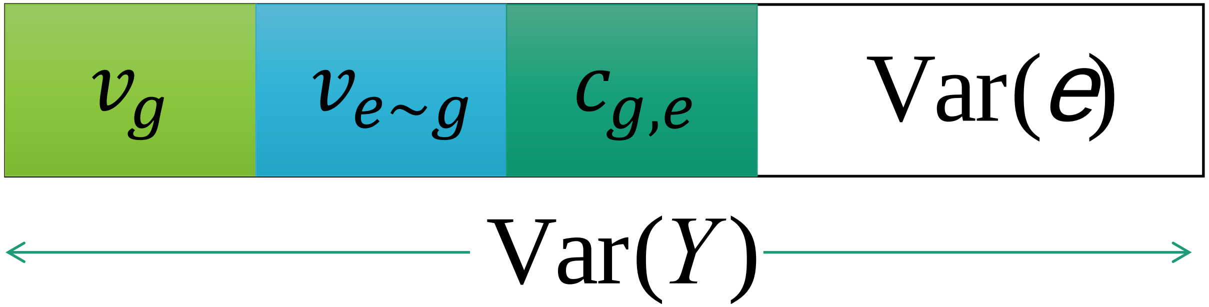
(Score constructed from Okbay et al. 2016)

Trait	R^2 (%)	R^2_δ (%)
Educational attainment	4.98	2.45
Age at first child	1.17	0.48
Health (composite)	0.67	0.23

Kong et al., The Nature of Nurture: effects of parental genotypes. *Science* (2018)

Variance decomposition

$$Y = \delta g + \eta(g_p + g_m) + \epsilon$$

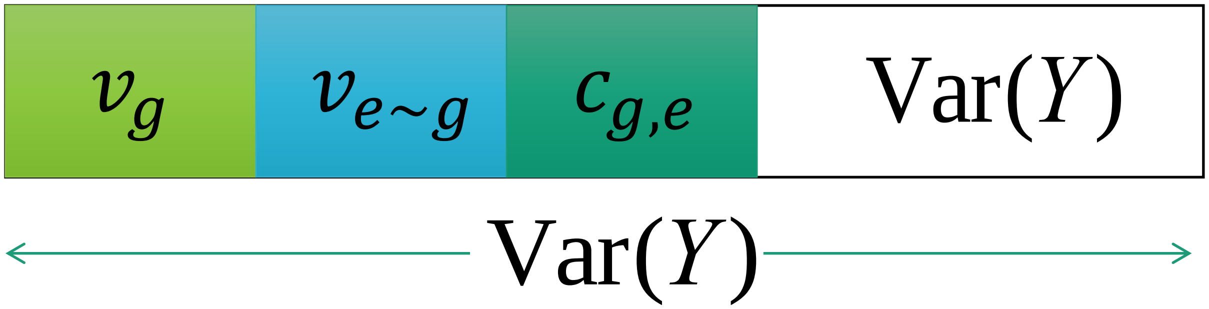


v_g $v_{e\sim g}$ $c_{g,e}$ $\mathrm{Var}(e)$

$\mathrm{Var}(Y)$

Variance decomposition

$$\mathrm{Var}(e)$$


v_g

$v_{e\sim g}$

$c_{g,e}$

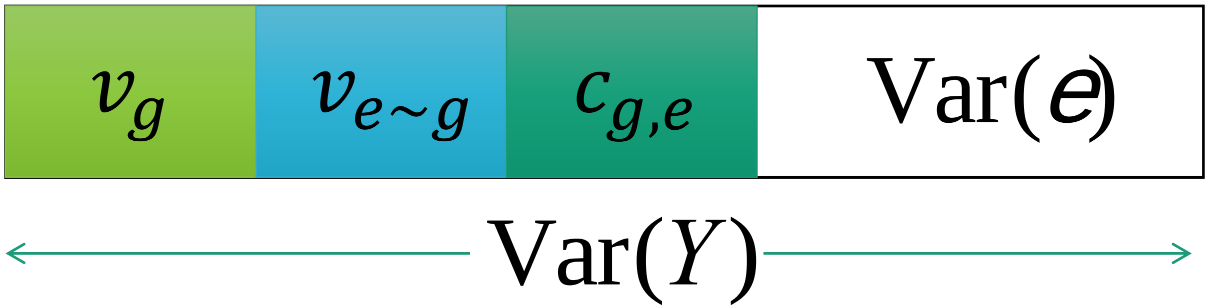
$\mathrm{Var}(Y)$

$\mathrm{Var}(Y)$

v_g Additive genetic variance: $h^2 = v_g/\mathrm{Var(Y)}$

Variance decomposition

$$Y = \delta g + \eta(g_p + g_m) + \epsilon$$

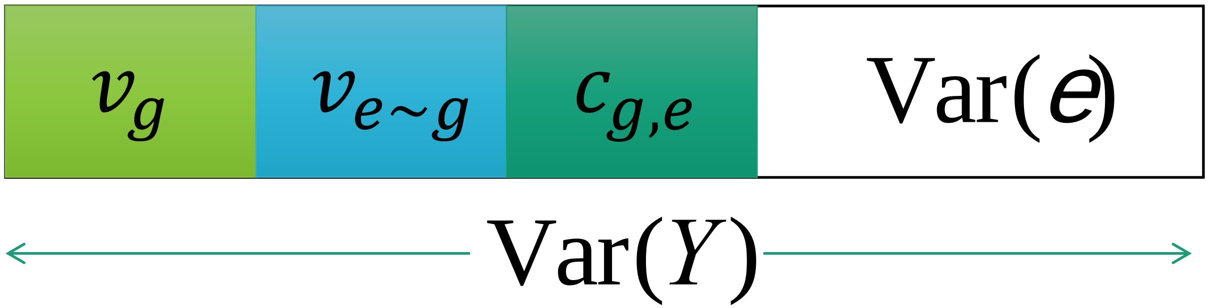


v_g Additive genetic variance: $h^2 = v_g/\mathrm{Var(Y)}$

$v_{e\sim g}$ Variance of environmental component correlated with parental genotype

Variance decomposition

$$Y = \delta g + \eta(g_p + g_m) + \epsilon$$



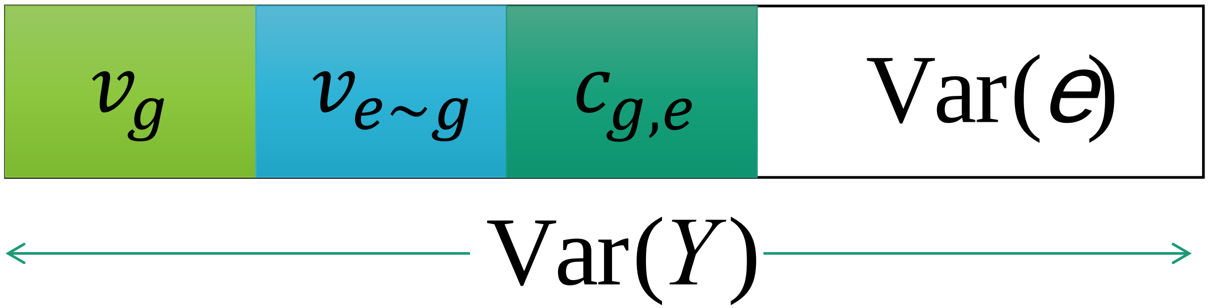
v_g $v_{e\sim g}$ $c_{g,e}$ $\mathrm{Var}(e)$

$\mathrm{Var}(Y)$

v_g	Additive genetic variance: $h^2 = v_g/\mathrm{Var(Y)}$
$v_{e\sim g}$	Variance of environmental component correlated with parental genotype
$c_{g,e}$	Covariance between genetic and environmental effects

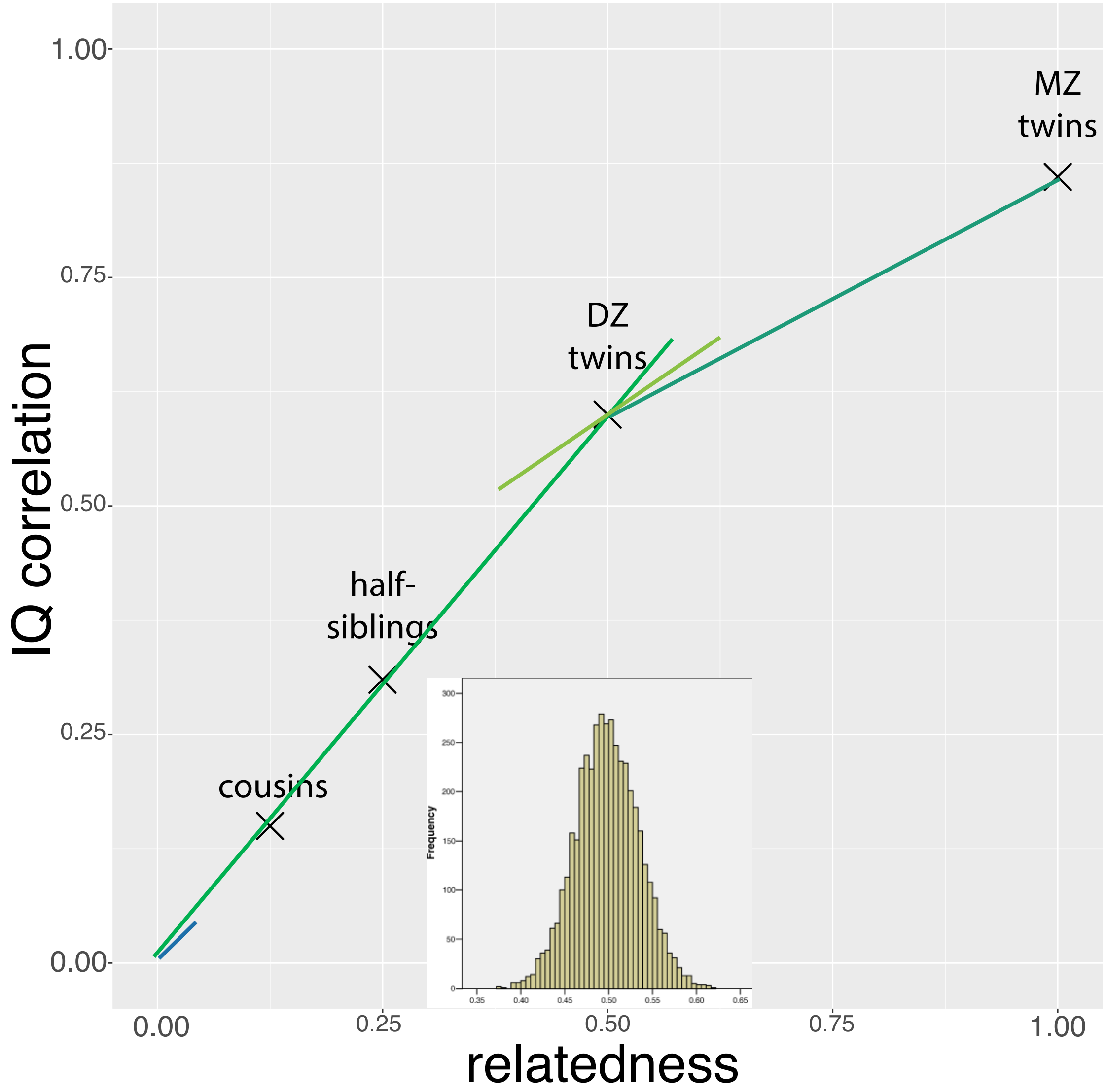
Variance decomposition

$$Y = \delta g + \eta(g_p + g_m) + \epsilon$$



v_g	Additive genetic variance: $h^2 = v_g/\mathrm{Var(Y)}$
$v_{e\sim g}$	Variance of environmental component correlated with parental genotype
$c_{g,e}$	Covariance between genetic and environmental effects
$\mathrm{Var}(e)$	Residual environmental variance

Heritability Estimation



Twin

GREML
(Yang et al. 2010)

Kinship
(Zaitlen et al. 2013,
Hill et al. 2018)

Sib-Regression
(Visscher et al. 2006)

Data from: Familial studies of intelligence: a review. TJ Bouchard JR and M McGue, Science, 1981.

ARTICLES
https://doi.org/10.1038/s41588-018-0178-9

nature genetics

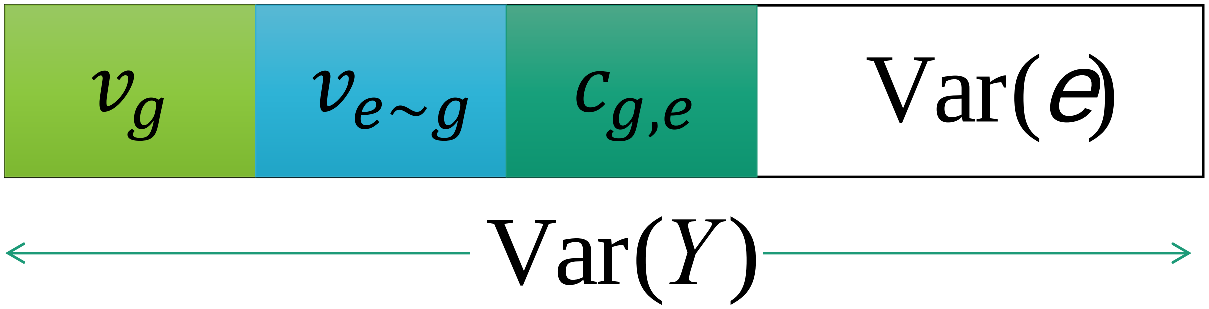
Relatedness disequilibrium regression estimates heritability without environmental bias

Alexander I. Young[1,2,3*], Michael L. Frigge[1], Daniel F. Gudbjartsson[1,4], Gudmar Thorleifsson[1], Gyda Bjornsdottir[1], Patrick Sulem[1], Gisli Masson[1], Unnur Thorsteinsdottir[1,5], Kari Stefansson[1,5] and Augustine Kong[1,3,4*]

Heritability measures the proportion of trait variation that is due to genetic inheritance. Measurement of heritability is important in the nature-versus-nurture debate. However, existing estimates of heritability may be biased by environmental effects. Here, we introduce relatedness disequilibrium regression (RDR), a novel method for estimating heritability. RDR avoids most sources of environmental bias by exploiting variation in relatedness due to random Mendelian segregation. We used a sample of 54,888 Icelanders who had both parents genotyped to estimate the heritability of 14 traits, including height (55.4%, s.e. 4.4%) and educational attainment (17.0%, s.e. 9.4%). Our results suggest that some other estimates of heritability may be inflated by environmental effects.

Variance decomposition

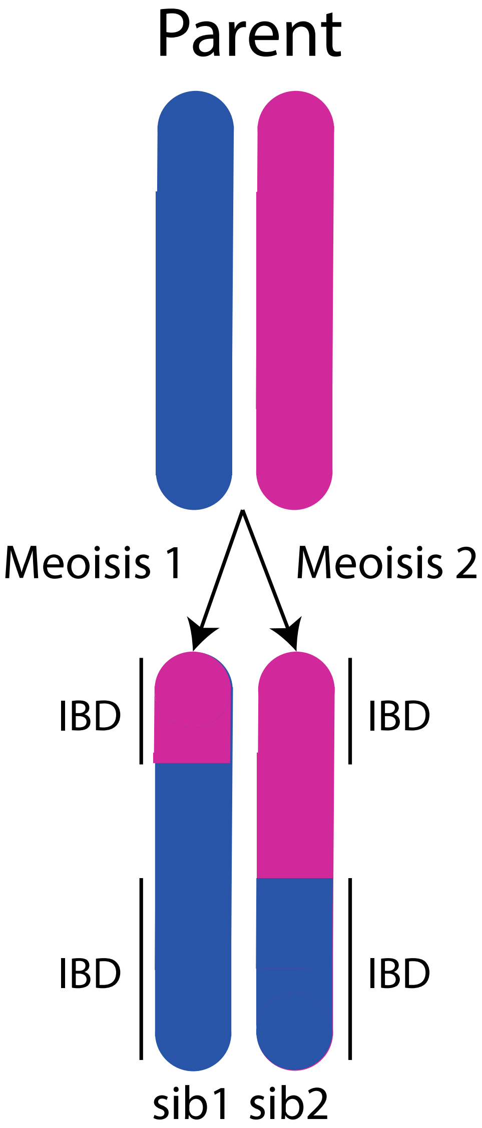
$$Y = \delta g + \eta(g_p + g_m) + \epsilon$$


v_g $v_{e\sim g}$ $c_{g,e}$ $\mathrm{Var}(e)$

$\mathrm{Var}(Y)$

v_g	Additive genetic variance: $h^2 = v_g/\mathrm{Var(Y)}$
$v_{e\sim g}$	Variance of environmental component correlated with parental genotype
$c_{g,e}$	Covariance between genetic and environmental effects
$\mathrm{Var}(e)$	Residual environmental variance

Identity-by-descent (IBD)



Expect ½ of genome inherited from father/mother to be IBD for siblings

Genotypic covariance between individuals

$$\mathrm{Cov}(g_i, g_j) = ?$$

Assume that $g_i = g_j$ when both inherited from recent common ancestor (identical-by-descent, IBD)

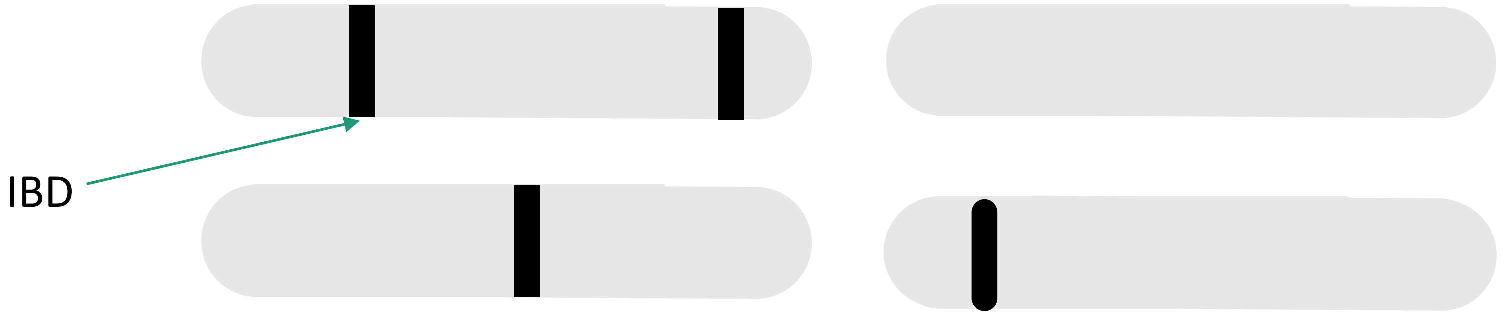
Assume that $g_i \perp g_j$ when not identical by descent

$$\mathrm{Cov}(g_i, g_j) = 2f(1-f)R_{i,j}$$

Relatedness between pair

Relatedness coefficient

$$R_{ij} = 2K_{i,j} = \frac{1}{2}(K_{i,j}^{pp} + K_{i,j}^{pm} + K_{i,j}^{mp} + K_{i,j}^{mm})$$



K^{pp}

paternal-paternal

K^{pm}

paternal-maternal

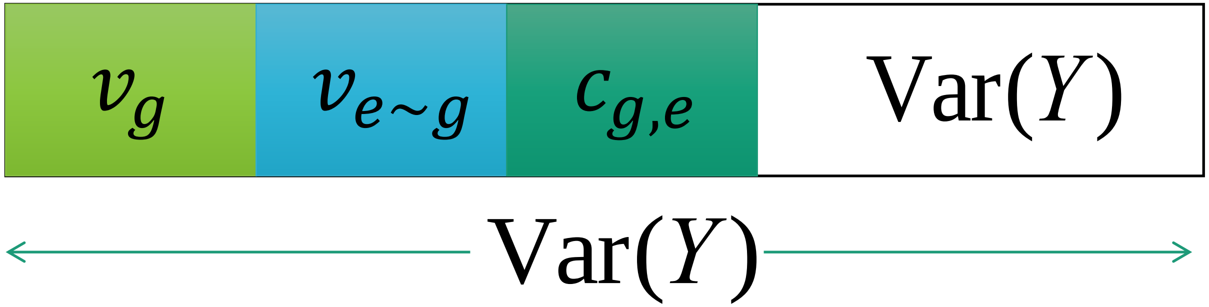
maternal-paternal

K^{mp}

maternal-maternal

K^{mm}

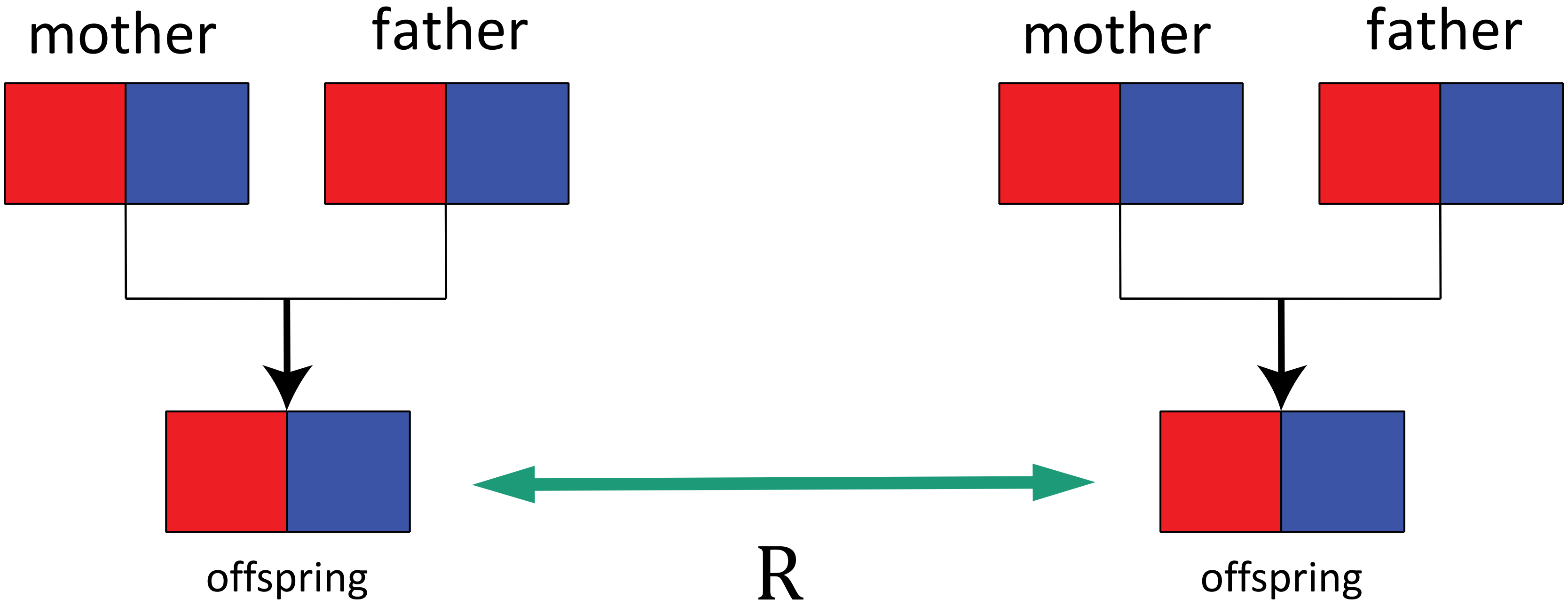
Covariance decomposition into inherited and non-inherited components



v_g $v_{e\sim g}$ $c_{g,e}$ $\mathrm{Var}(Y)$

$\mathrm{Var}(Y)$

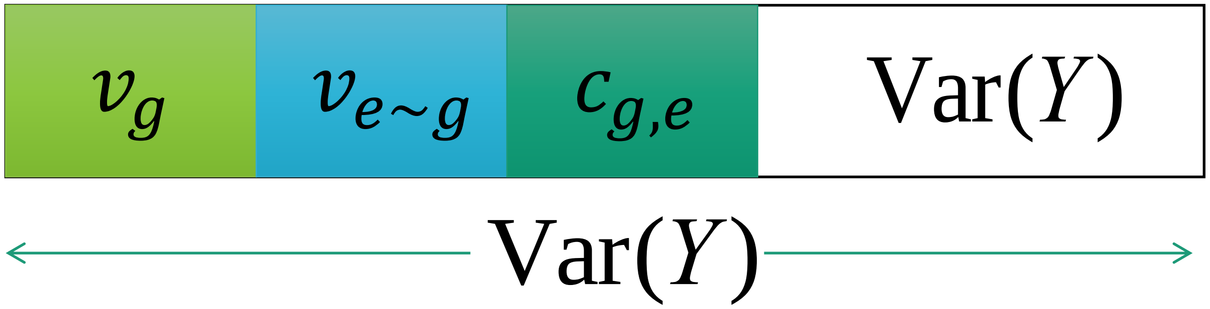
$$\text{Cov}(\mathbf{Y}) = v_g\mathrm{R} + v_{e\sim g}\mathrm{R}_{\text{par}} + c_{g,e}\mathrm{R}_{\text{o,par}} + \sigma^2\text{Cov}(\boldsymbol{\varepsilon})$$



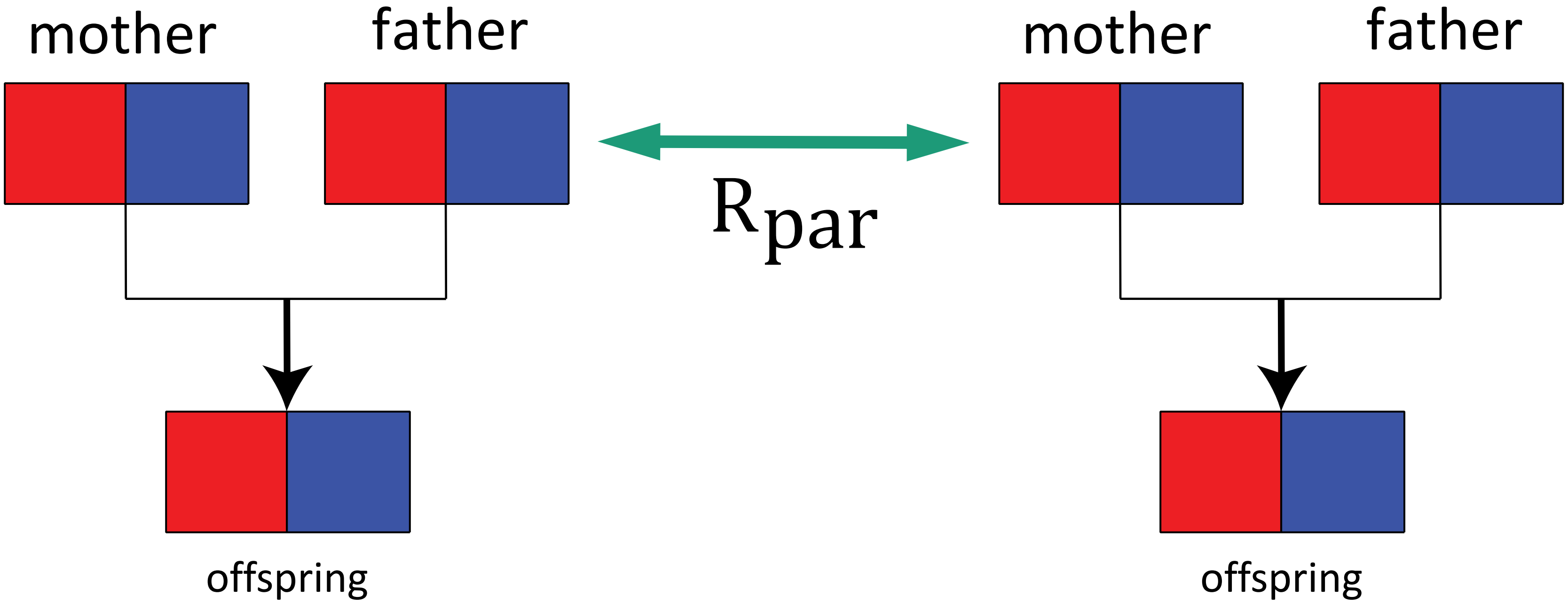
mother father mother father

offspring R offspring

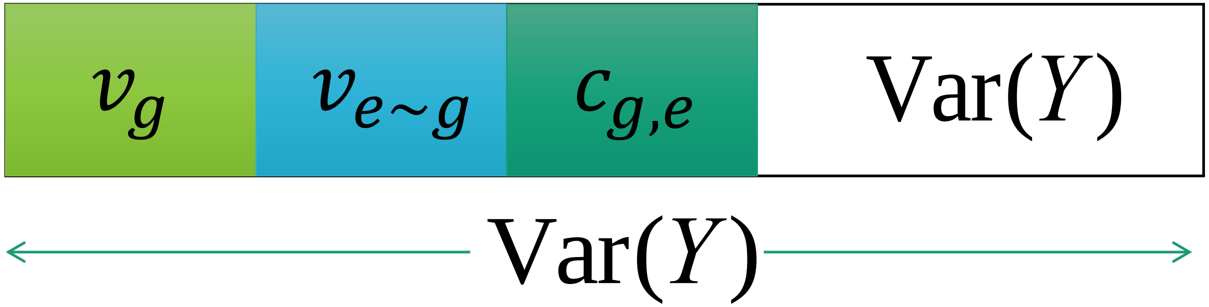
Covariance decomposition into inherited and non-inherited components



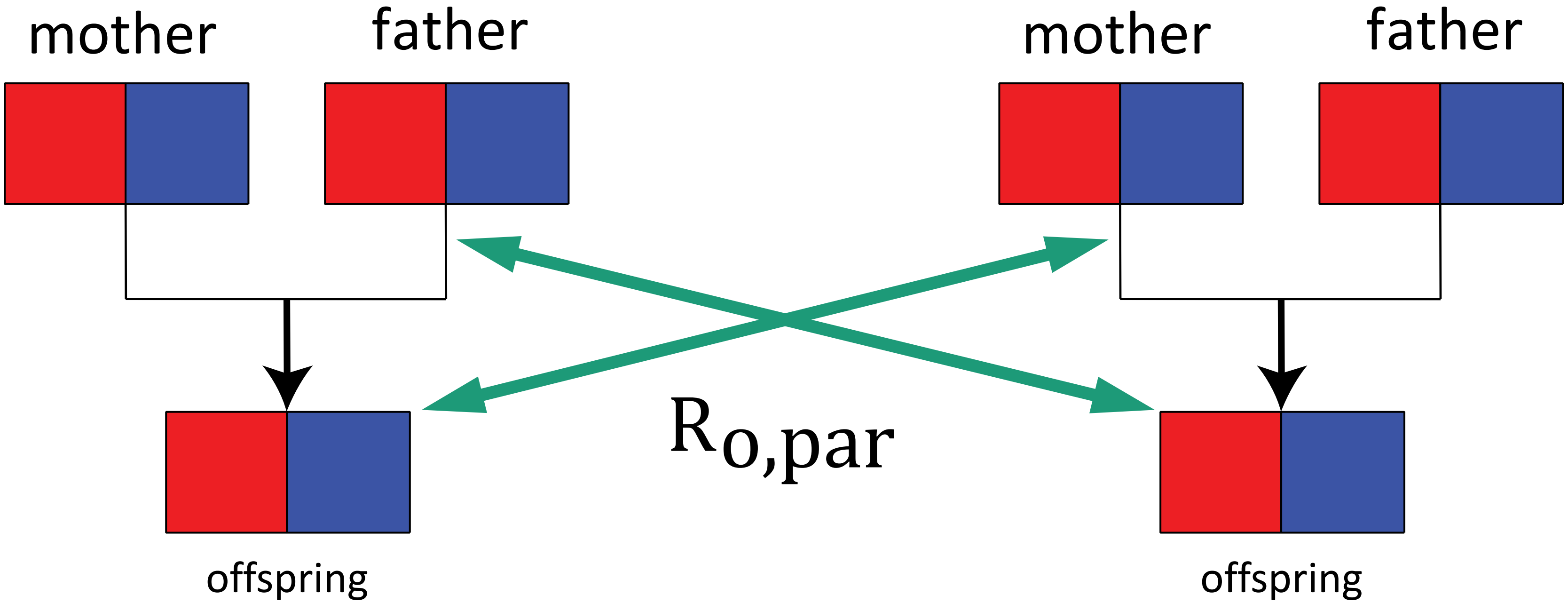
$$\mathrm{Cov}(\mathbf{Y}) = v_g \mathrm{R} + v_{e\sim g}\mathrm{R}_{\mathrm{par}} + c_{g,e}\mathrm{R}_{\mathrm{o,par}} + \sigma^2 \mathrm{Cov}(\boldsymbol{\varepsilon})$$



Covariance decomposition into inherited and non-inherited components



$$\mathrm{Cov}(\mathbf{Y}) = v_g \mathrm{R} + v_{e\sim g} \mathrm{R}_{\mathrm{par}} + c_{g,e} \mathrm{R}_{\mathrm{o,par}} + \sigma^2 \mathrm{Cov}(\boldsymbol{\varepsilon})$$



Relatedness Disequilibrium Regression

‘True’ covariance model:

$$\text{Cov}(\mathbf{Y}) = v_g \mathrm{R} + v_{e\sim g} \mathrm{R}_{\mathrm{par}} + c_{g,e} \mathrm{R}_{\mathrm{o,par}} + \text{Cov}(\boldsymbol{\varepsilon})$$

RDR covariance model

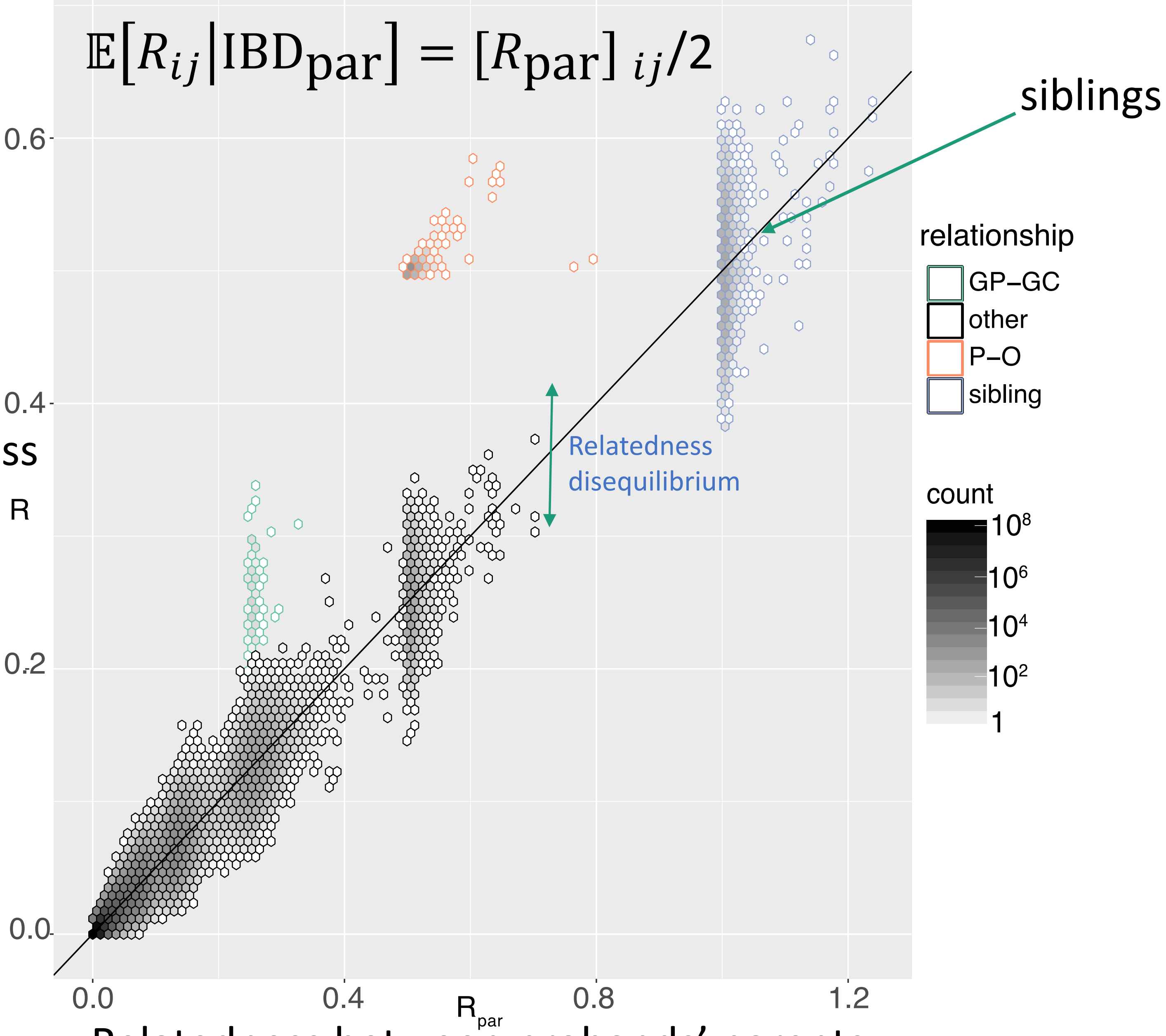
$$\text{Cov}(\mathbf{Y}) = v_g \mathrm{R} + v_{e\sim g} \mathrm{R}_{\mathrm{par}} + c_{g,e} \mathrm{R}_{\mathrm{o,par}} + \sigma^2 \mathrm{I}$$

Theorem

The estimator of the genetic variance from the RDR model converges to the true heritability

$$\widehat{v_g} \longrightarrow v_g$$

Relatedness Disequilibrium



$$\mathbb{E}[R_{ij}|\mathrm{IBD}_{\mathrm{par}}] = [R_{\mathrm{par}}]_{ij}/2$$

siblings

Relatedness disequilibrium

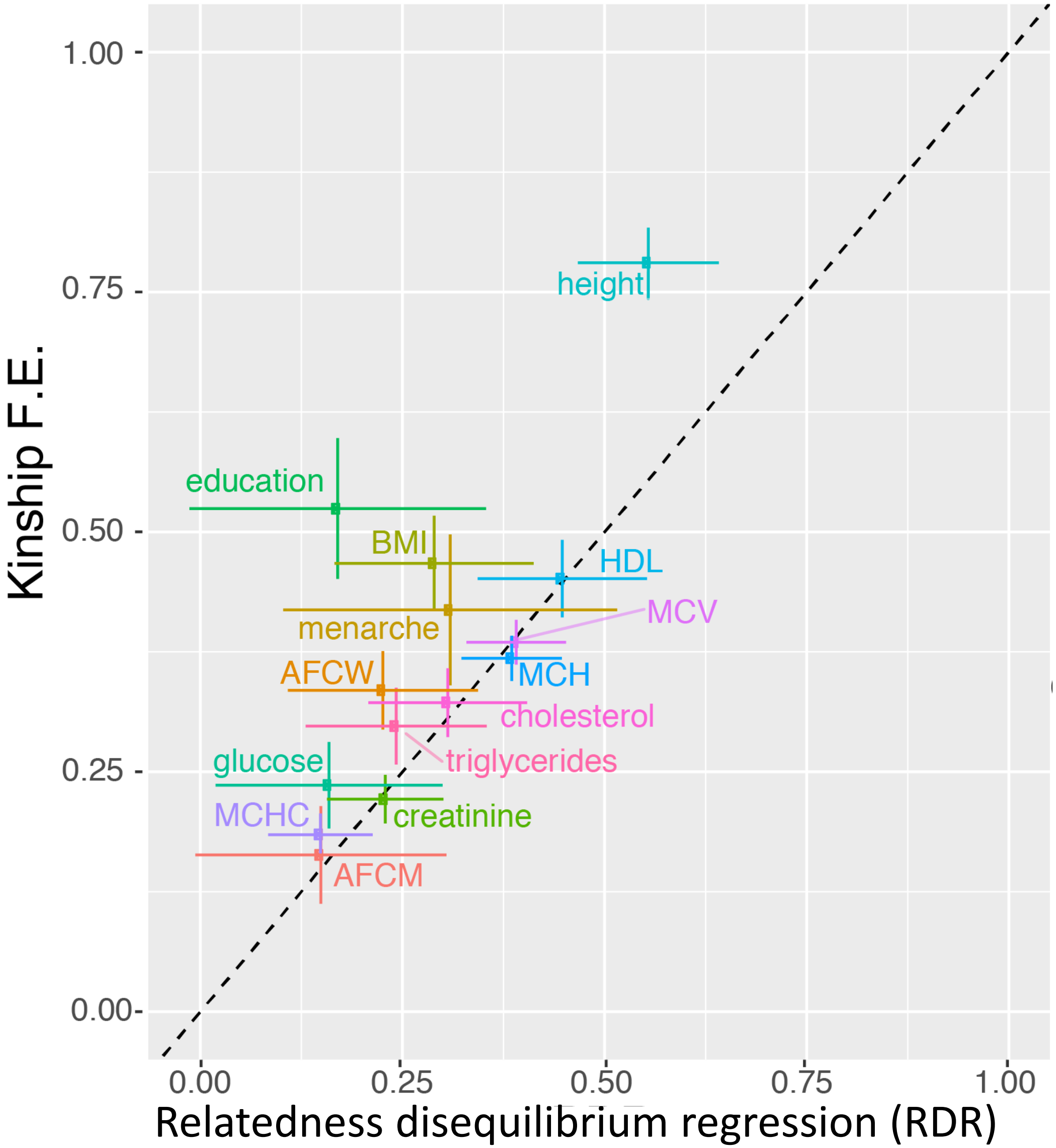
Relatedness between probands

Relatedness between probands' parents

Selected estimates

Trait	RDR h^2 (S.E.) (%)
BMI	28.9 (6.3)
height	55.4 (4.4)
education	17.0 (9.4)

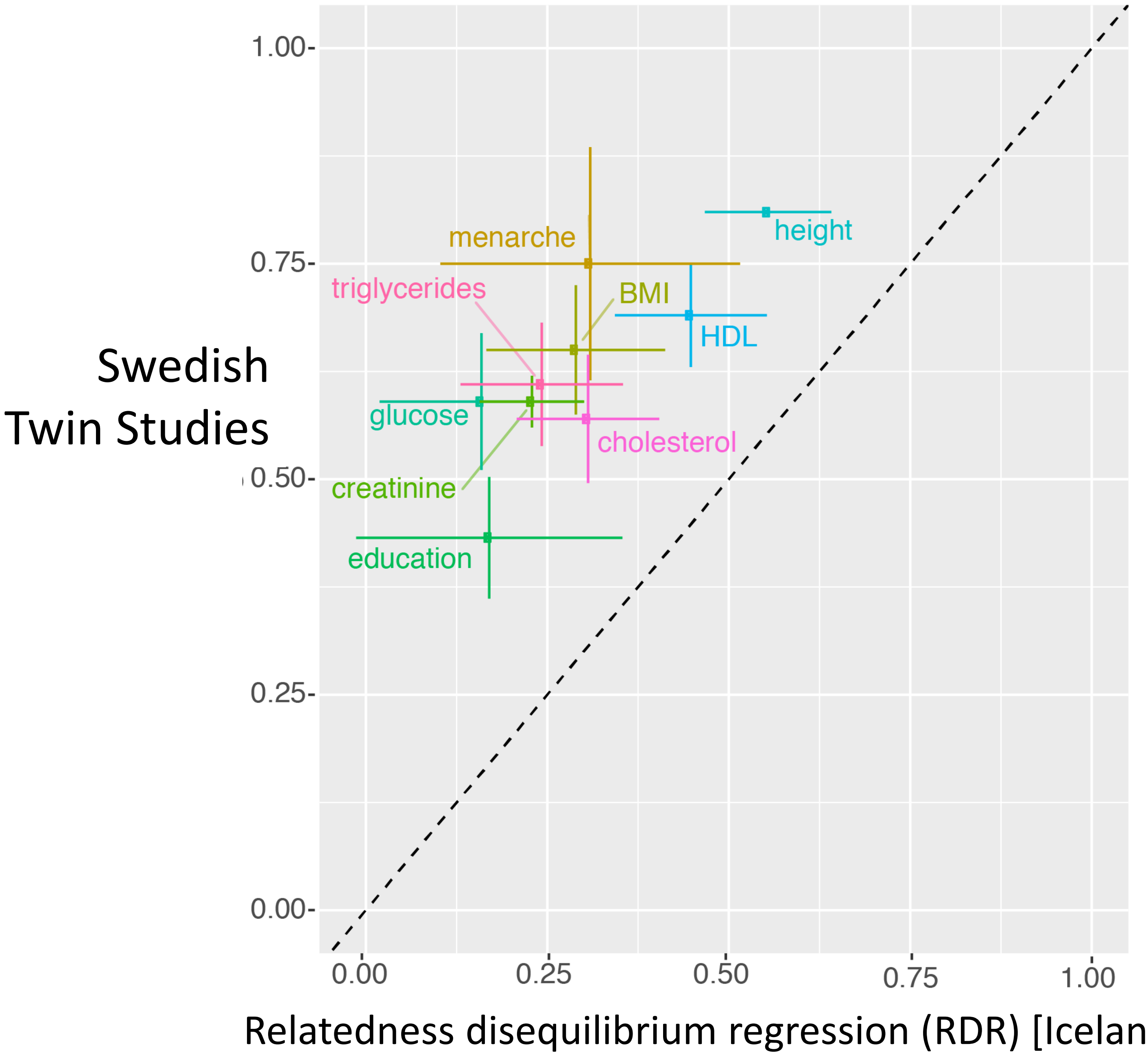
Comparison to Kinship (F.E.) Estimator



Significant differences for:
EA, height, BMI, AAFCW

Mean Kin F.E.-RDR = 12.1%.

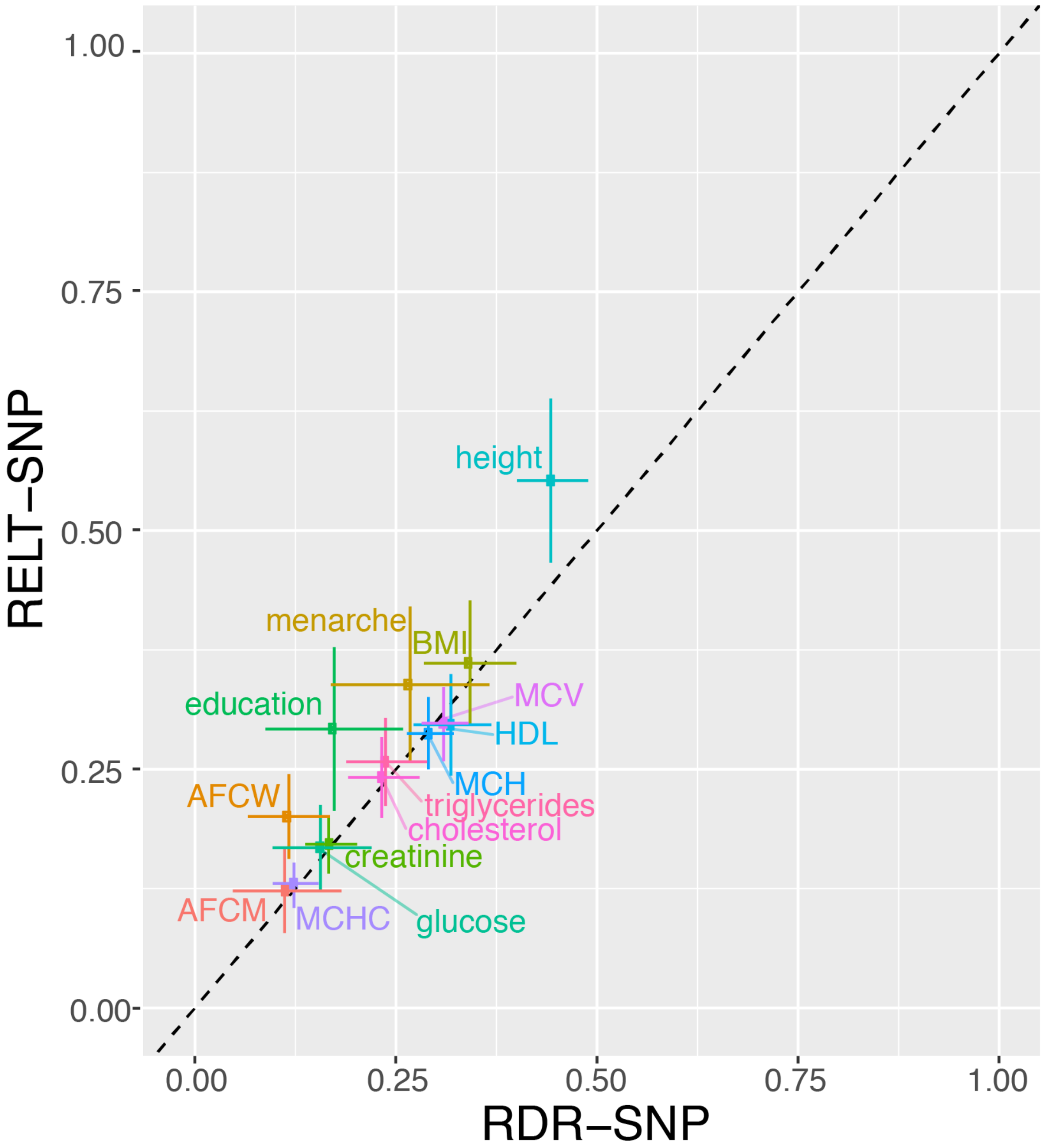
Comparison to Swedish Twin Studies



Mean Twin-RDR: 33%.

Twin estimate higher ($p<0.05$) for all traits.

Overestimation of heritability by GREML methods



Educational attainment

$$\frac{h^2_{\text{RELT−SNP}}}{h^2_{\text{RDR−SNP}}} = 1.69,\ p<0.027$$

Consistent with parental genetic nurturing 30% of direct

Consistent with Kong et al., Science 2018, and EA3 (Lee et al. Nature Genetics 2018)

AFCW

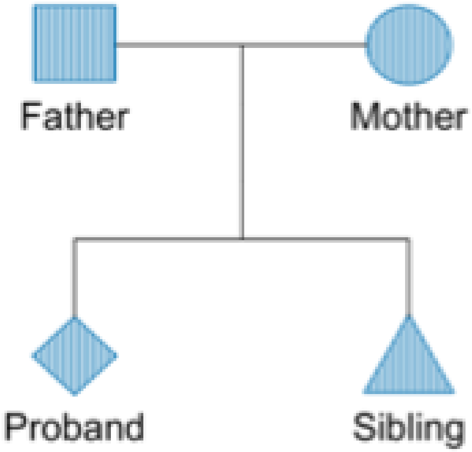
$$\frac{h^2_{\text{RELT−SNP}}}{h^2_{\text{RDR−SNP}}} = 1.72,\ p<7.6\text{x}10^{-3}$$

Height

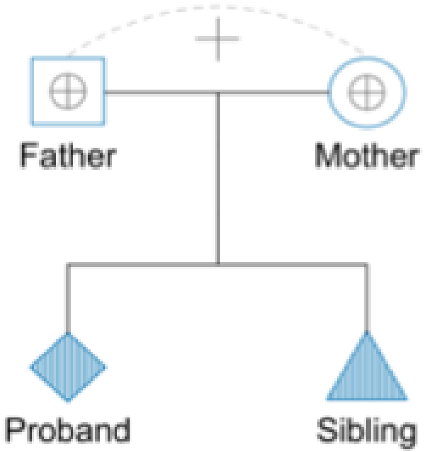
$$\frac{h^2_{\text{RELT−SNP}}}{h^2_{\text{RDR−SNP}}} = 1.24,\ p<0.015$$

Problem: parental genotypes often missing

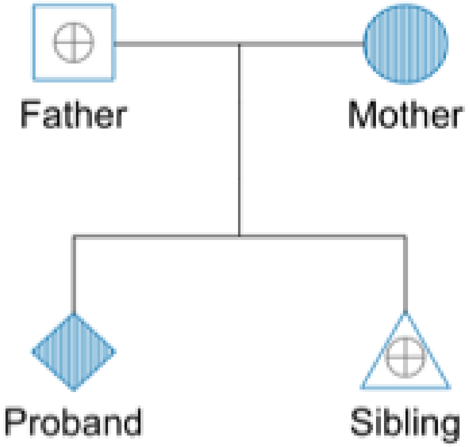
Complete Data



Sibs only

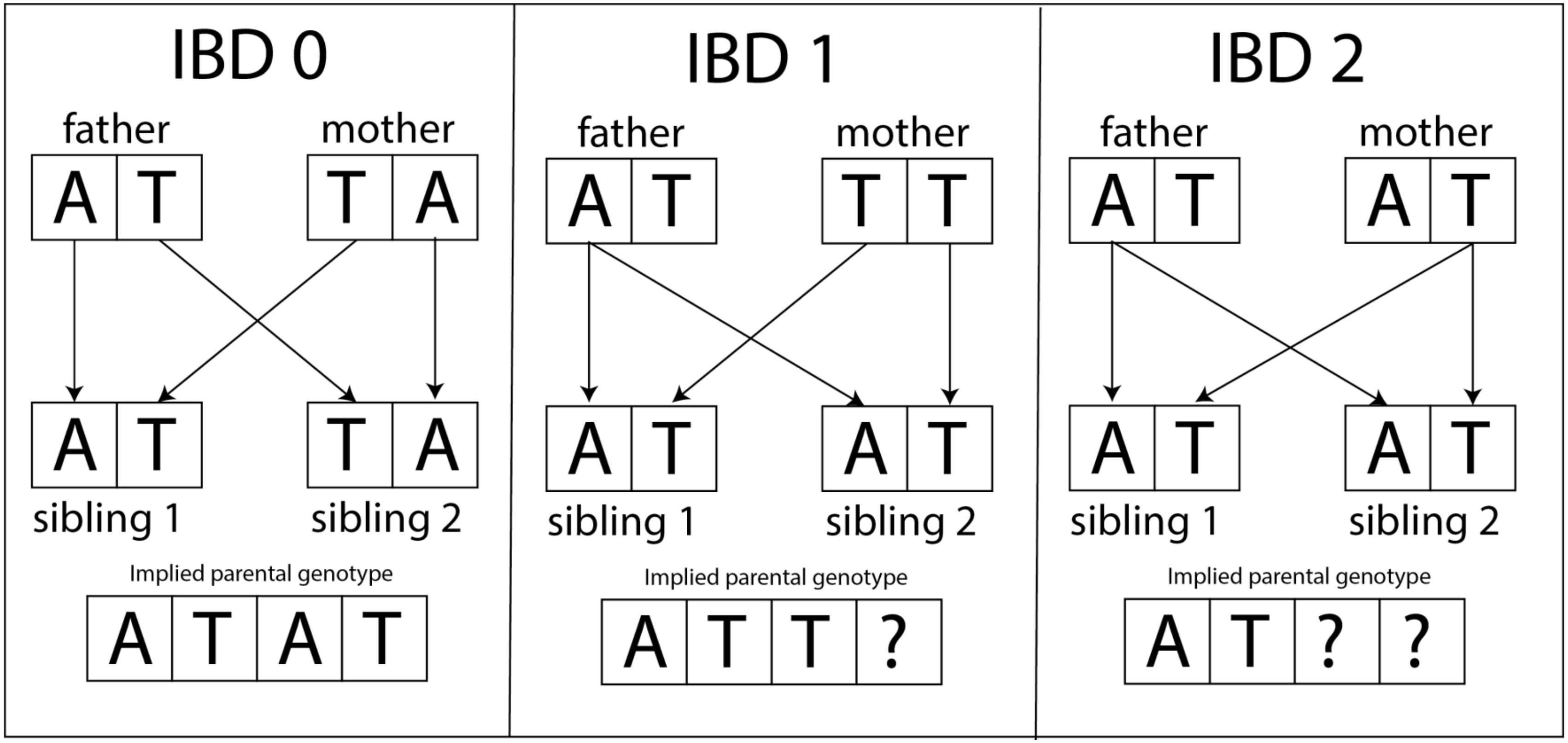


Parent-offspring



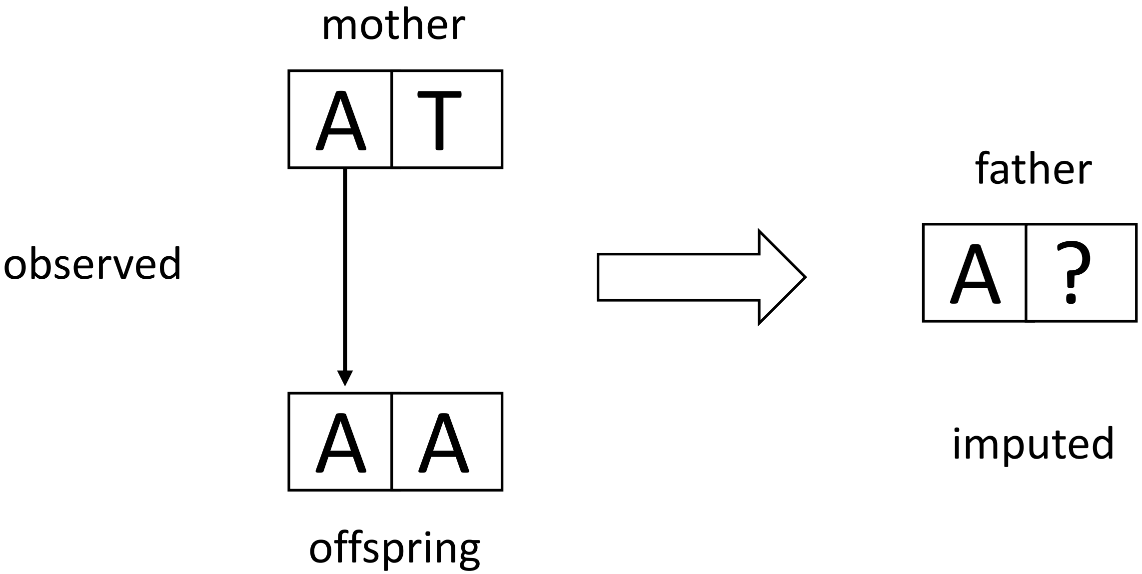
Kong et al., Family analysis with Mendelian Imputations, biorXiv 2020

Imputation from siblings using IBD



$$\mathrm{Var}(\hat{g}_{par(i)}) = 3f(1-f) = (3/4)\mathrm{Var}(g_{par(i)}).$$

Imputation from parent-offspring pairs



mother

A T

observed

A A

offspring

father

A ?

imputed

With double het resolved by phase:

$$\mathrm{Var}(\hat{g}_{p(i)}) = f(1-f) = \mathrm{Var}(g_{p(i)})/2$$

Imputation for multiple regression

Model

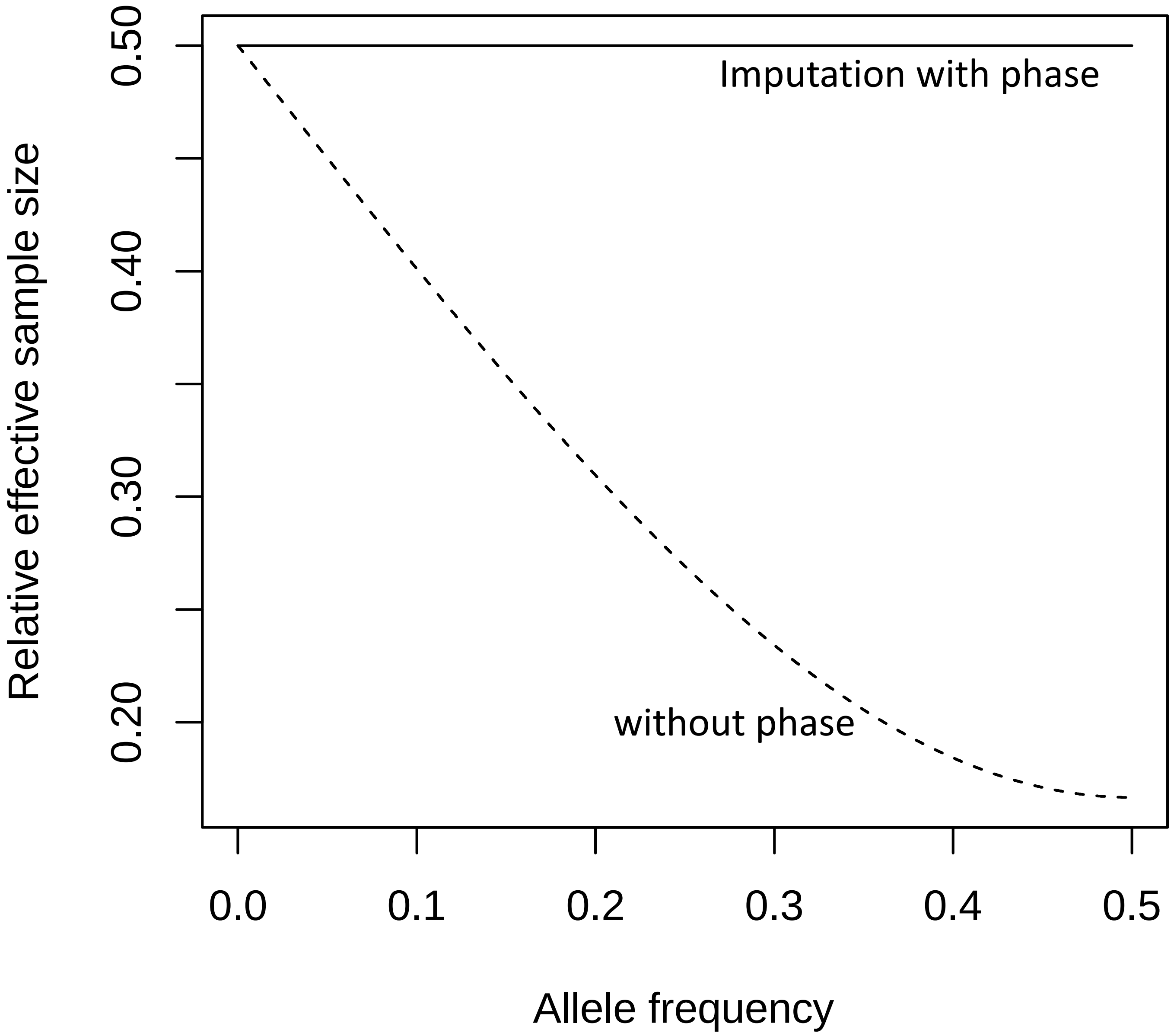
$$y_i = \delta g_i + \alpha_m g_{m(i)} + \alpha_p g_{p(i)} + \epsilon_i$$

Regression with imputed paternal genotype

$$y_i \sim g_i + g_{m(i)} + \hat{g}_{p(i)} \;\rightarrow\; \delta,\, \alpha_m,\, \alpha_p$$

Imputation gives unbiased estimates!

Efficiency relative to both parents observed (direct effects)



Estimating effects from siblings

Correlated with parental genotype

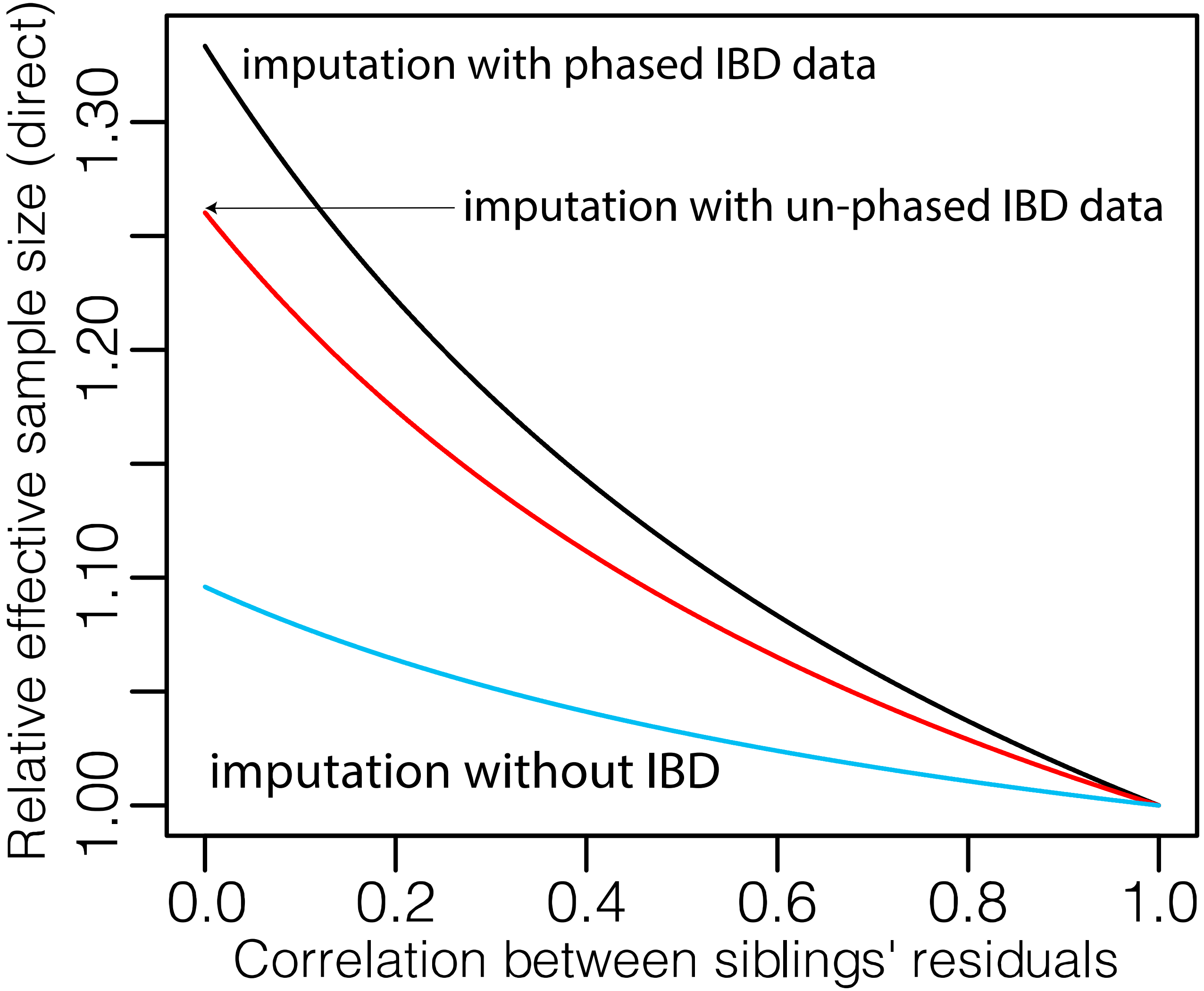
Between family genetic differences

Uncorrelated with parental genotype

Within-family genetic differences

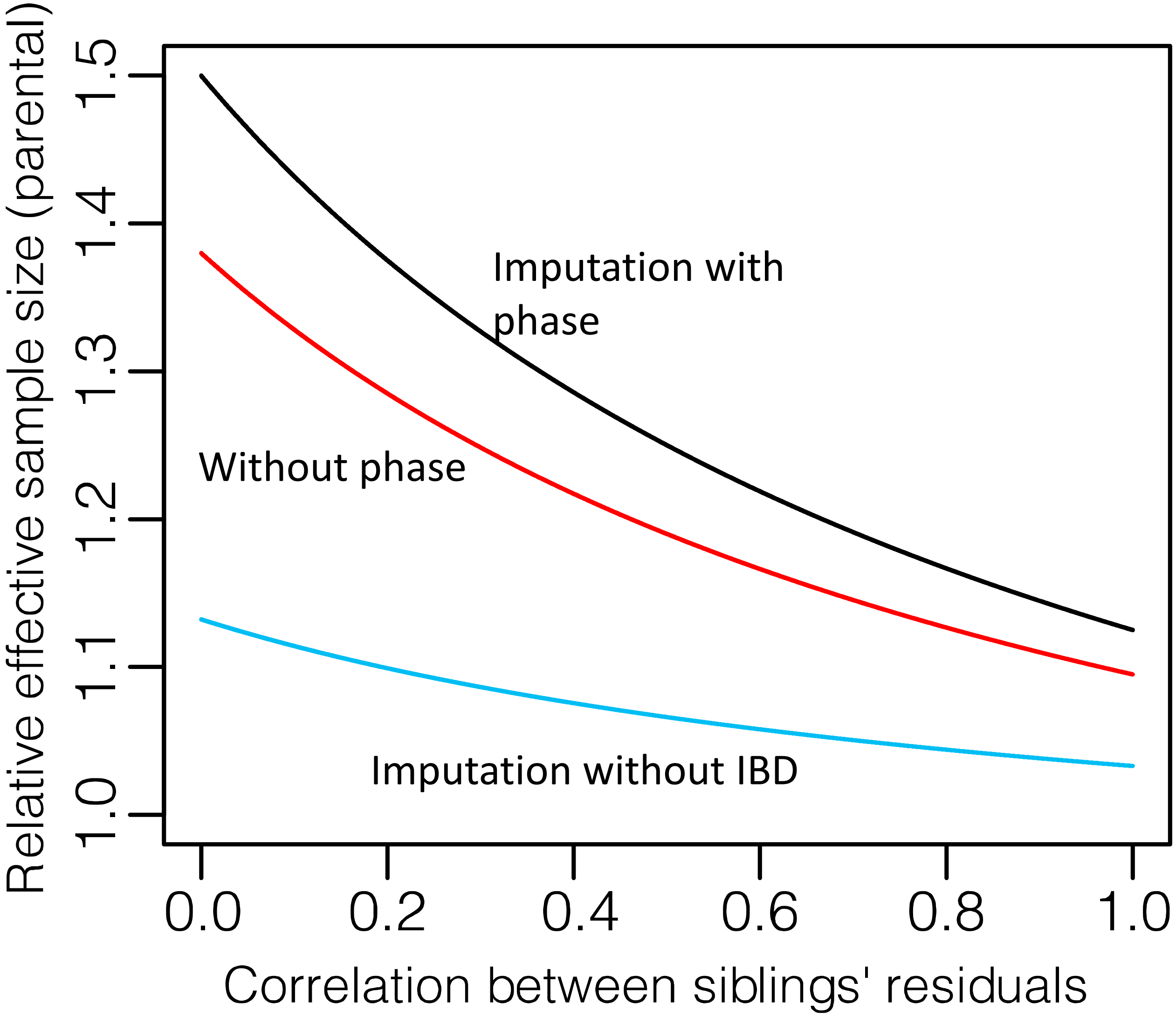
Orthogonal axes of information for estimation of direct genetic effects

Estimating effects from siblings



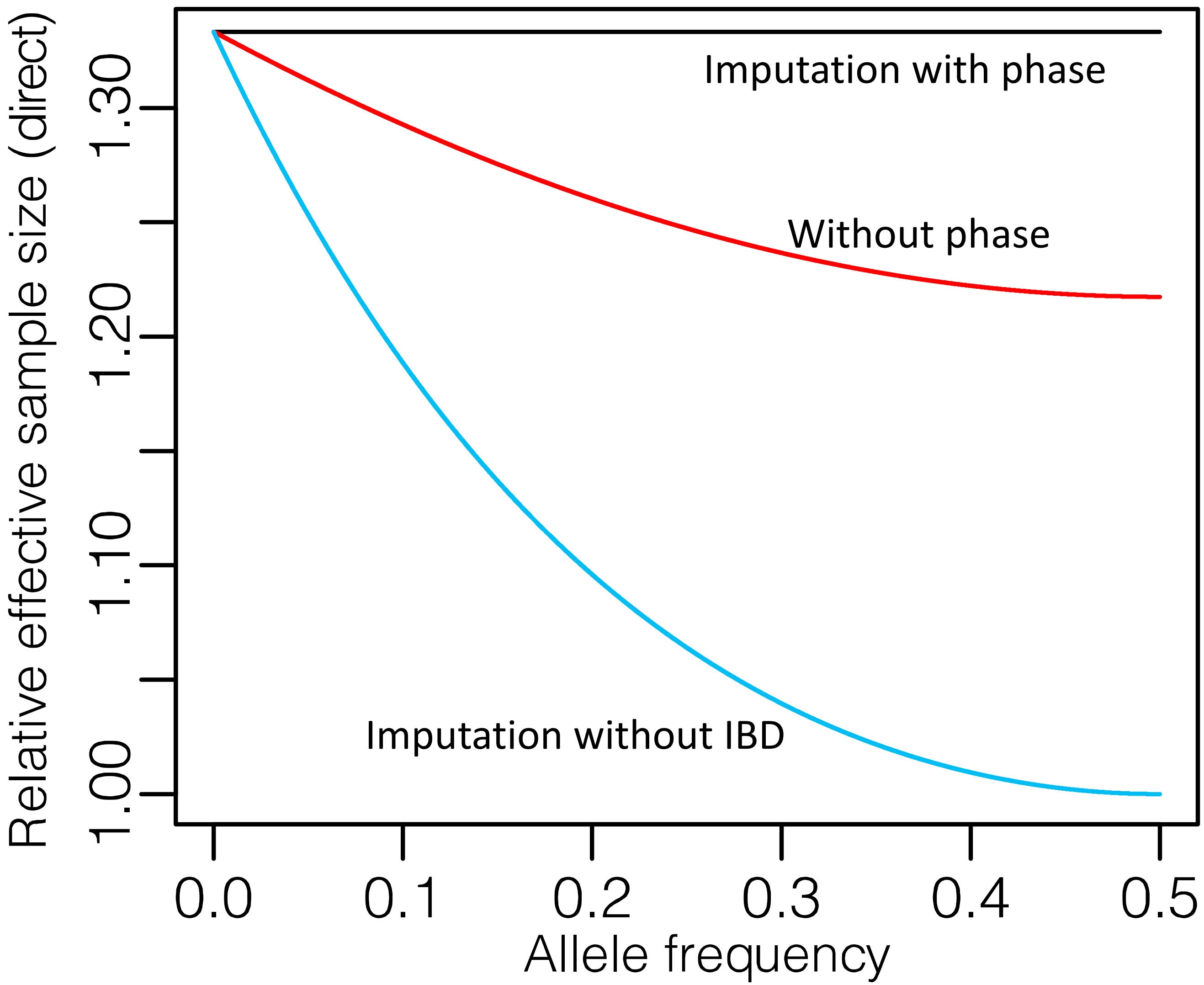
Efficiency for estimation of direct effects relative to no imputation

Estimating effects from siblings



Efficiency for estimation of parental effects relative to no imputation

Estimating effects from siblings



Efficiency for estimation of direct effects relative to no imputation

Mixed model regression

$$Y_{ij} = X_{ij}\alpha + \mu_i + \epsilon_{ij};\ \ \mu_i \sim N(0, \sigma_F^2);\ \ \epsilon_{ij} \sim N(0, \sigma_\epsilon^2);$$

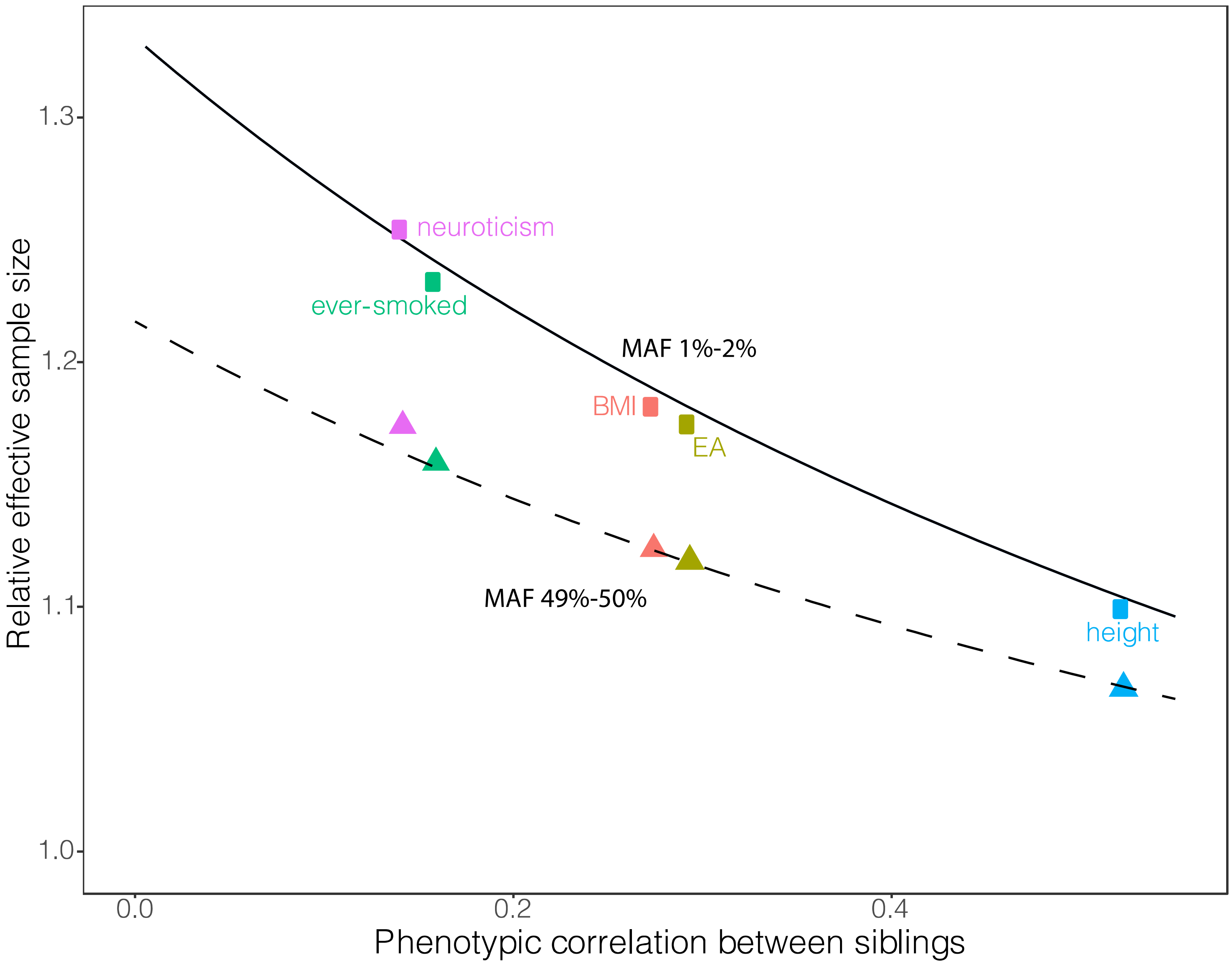
Genotype matrix with columns: Proband, (imputed) paternal, (imputed) maternal

Family level random effect

Individual error term

Fast imputation and mixed model regression code in SNIPar (single nucleotide imputation of parents) software package: https://github.com/AlexTISYoung/SNIPar

Gain in effective sample size (direct effects) over no imputation



How much bias does GWAS have?

Model

$$y_i = \delta g_i + \alpha_m g_{m(i)} + \alpha_p g_{p(i)} + \epsilon_i$$

Standard GWAS picks up ‘parental effects’

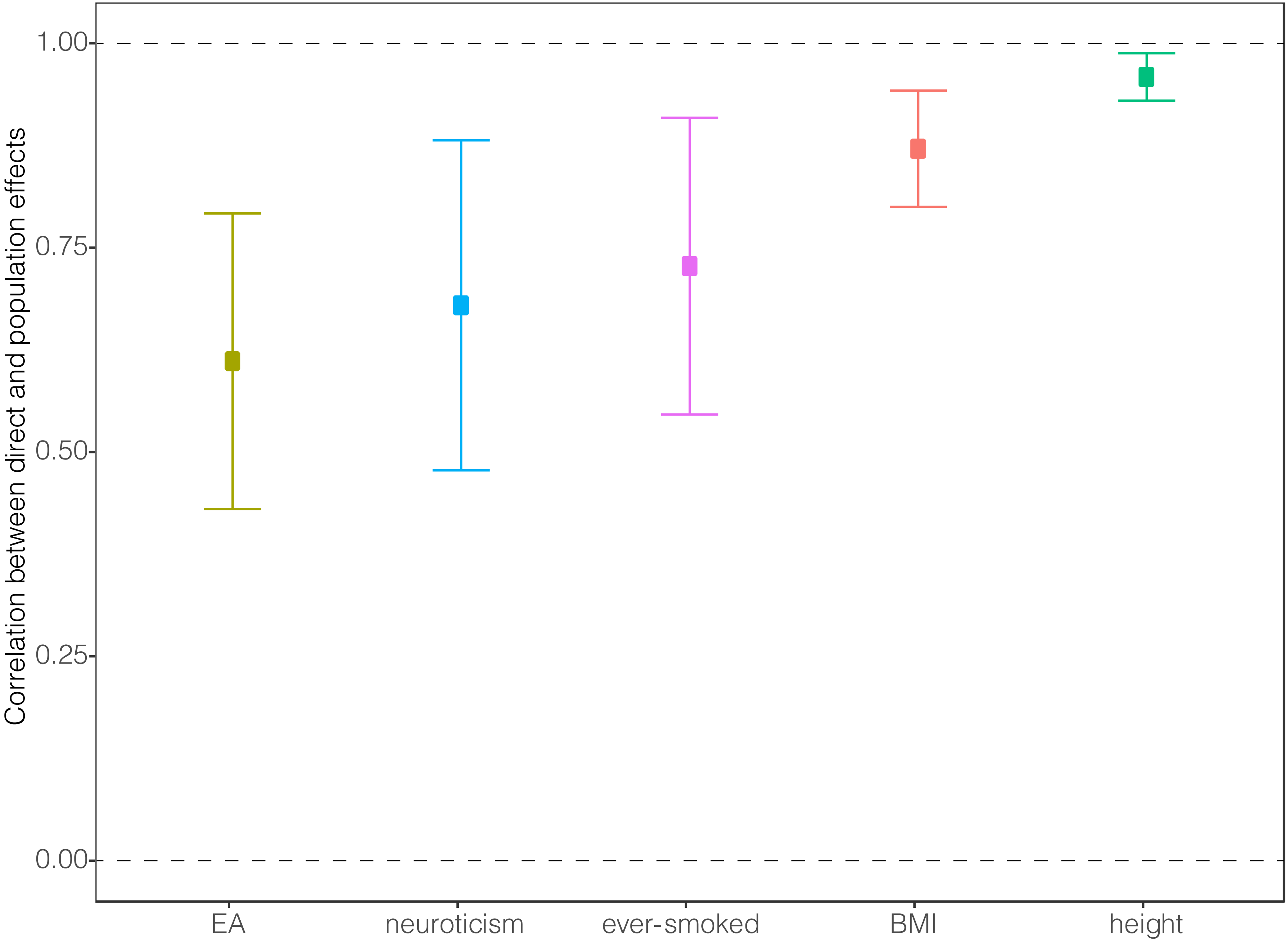
$$Y \sim g_i \;\Rightarrow \hat{\beta} \rightarrow \delta + (\alpha_p + \alpha_m)/2$$

‘Population effect’

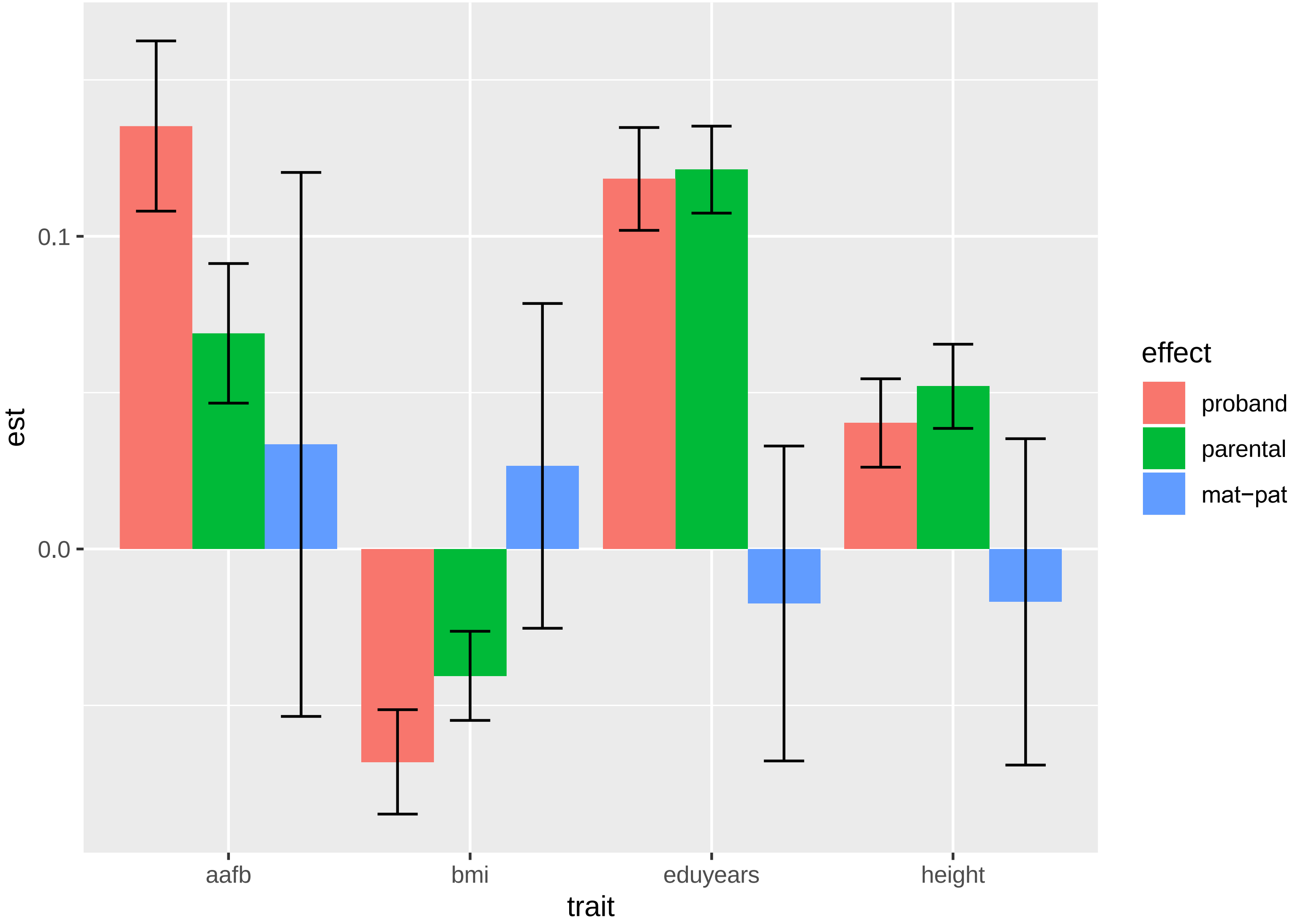
What is the correlation of direct and population effect?

$$\mathrm{Corr}(\delta, \delta + (\alpha_p + \alpha_m)/2) = ?$$

Correlation between direct and population effects



Application to PGS analyses



Using EA PGS. Note that assortative mating leads to slight bias in parental effect estimates. (Kong et al. Family Analysis with Mendelian Imputations, biorXiv 2020).

Acknowledgements

University of Oxford

Augustine Kong
Stefania Benonisdottir

SSGAC

Moeen Nehzati

Daniel Benjamin

Patrick Turley
Chanwook Lee

David Cesarini
Aysu Okbay
Hariharan Jayashakar

Funding

Li Ka Shing Foundation
Open Philanthropy
Ragnar Soderberg Foundation
NIA/NIH

Software: github.com/alextisyoung/SNIPar

@alexTISyoung