Resemblance between relatives

Key concepts

Model phenotypes by fixed effects and random effects including genetic value (additive, dominance, epistatic)

Model covariance of genetic effects by relationship estimated from pedigree (or SNP genotypes)

Estimate genetic variance by REML

What do we mean by “resemble”?

Similar values of quantitative traits

Measure by correlation:

$$= \text{Covariance}(y_i, y_j) / \text{Variance}(y)$$

$$= \text{Cov}(y_i, y_j)/V(y)$$

Why do relatives resemble each other?

Why do relatives resemble each other?

Similar:

genes

family environment

country

school

Model phenotype

Phenotype = genetic effect
+ country
+ year of birth
+ family environment

<u>Fixed</u> effects
country, year of birth

<u>Random</u> effects
genetic effect, family environment

we need a model of the covariances between terms

Intermezzo: fixed and random effects

Model phenotype

Phenotype = genetic effect

+ country

+ year of birth

+ family environment

+ individual environment

V(phenotype) = V(genetic effects) + V(family environment)

+ V(individual environment)

Cov(phenotype$_i$, phenotype$_j$) = Cov(genetic effects)

+ Cov(family environment)

Model phenotype

Random effects

genetic effect, common (family) environment (CE)

we need a model of the covariances between terms

Cov(common environment) = 0 if different families

= 1 * V_{CE} if same family

Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive gene action

we need genetic variances and covariances

genotype	BB	Bb	bb
effect	a	0	-a
frequency	p^2	2pq	q^2

(p + q = 1)

$$\text{mean} = a * p^2 + 0*2pq - a*q^2 = (p - q)*a$$

$$V(\text{genetic effect}) = \text{genetic variance} = V_G$$

$$= E(\text{effect}^2) – E(\text{effect})^2$$

$$V_G = a^2 * p^2 + 0*2pq + a^2*q^2 – [(p - q)*a]^2 = 2pqa^2$$

Model: 1 gene, 2 alleles and additive gene action

Covariance between parent and offspring

Parent			Offspring			
Genotype	effect	frequency	BB	Bb	bb	mean
BB	a	p^2	p	q		pa
Bb	0	2pq	0.5p	0.5	0.5q	0.5(p-q)a
bb	-a	q^2		p	q	-qa

Cov(parent genetic value, offspring genetic value)

$= p^2 * a * pa + q^2 * (-a)*(-qa) – [(p - q)a]*[(p - q)a] = pqa^2 = 0.5V_G$

using Cov(y, x) = E(yx) – E(y)E(x)

Model: 1 gene, 2 alleles and additive gene action

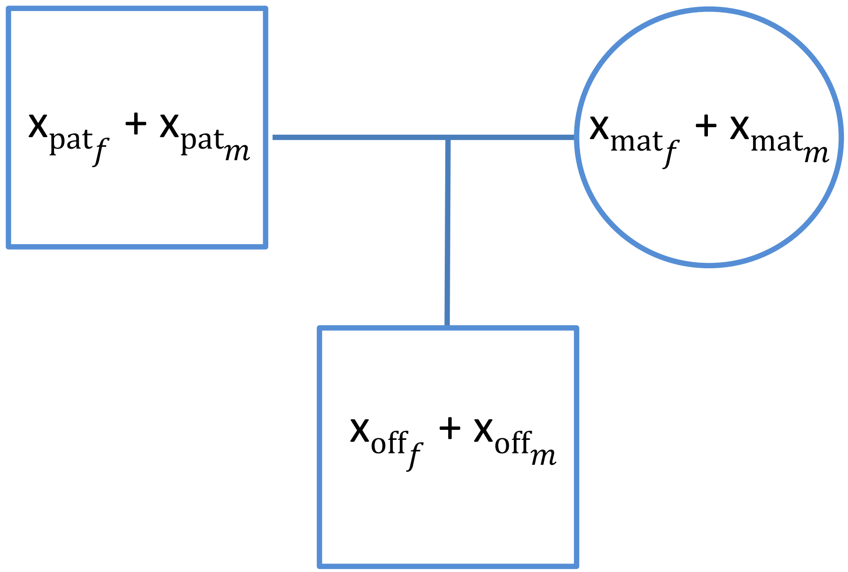
Another way: model genetic value as sum of gametic effects from mother (*m*) and father (*f*)

$$g = x_m + x_f$$

$$V(g) = V(x_m) + V(x_f) = 2V(x)$$

Model: 1 gene, 2 alleles and additive gene action

Another way: model genetic value as sum of <u>gametic effects</u> from mother (*m*) and father (*f*)

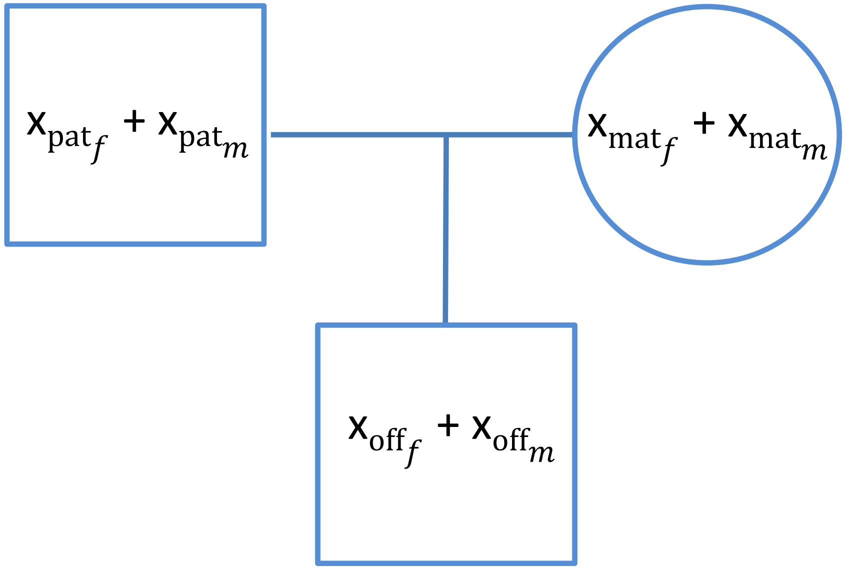


pat = paternal
mat = maternal

$Cov(g_{pat}, g_{off}) = Cov(x_{pat_f} + x_{pat_m}, x_{off_f} + x_{off_m})$

Model: 1 gene, 2 alleles and additive gene action

Another way: model genetic value as sum of gametic effects from mother (*m*) and father (*f*)



$\text{Cov}(g_{pat}, g_{off}) = \text{Cov}(x_{pat_f} + x_{pat_m}, x_{off_f} + x_{off_m})$

$= \text{Cov}(x_{pat_m}, x_{off_m}) + \text{Cov}(x_{pat_m}, x_{off_f}) + \text{Cov}(x_{pat_f}, x_{off_m}) + \text{Cov}(x_{pat_f}, x_{off_f})$

$= \quad 0 \quad + \quad ? \quad + \quad 0 \quad + \quad ?$

? - depends if these gametes are identical by descent (IBD)

Model: 1 gene, 2 alleles and additive gene action

Covariance between parent and offspring:

$\mathrm{Cov}(x_{\mathrm{pat}_m}, x_{\mathrm{off}_f})$ = V(x) if x_{pat_m} is IBD to x_{off_f}

= 0 otherwise

IBD depends on segregation probability:

$\mathrm{Cov}(x_{\mathrm{pat}_m}, x_{\mathrm{off}_f}) = \mathrm{Cov}(x_{\mathrm{pat}_f}, x_{\mathrm{off}_f}) = 0.5V(x)$

$\mathrm{Cov}(g_{\mathrm{pat}}, g_{\mathrm{off}}) = 0 + 0.5V(x) + 0 + 0.5V(x) = V(x) = 0.5V_G$

Probability that relatives share alleles IBD

Covariance between relatives depends on probability that their alleles are IBD

This probability can be calculated from pedigrees

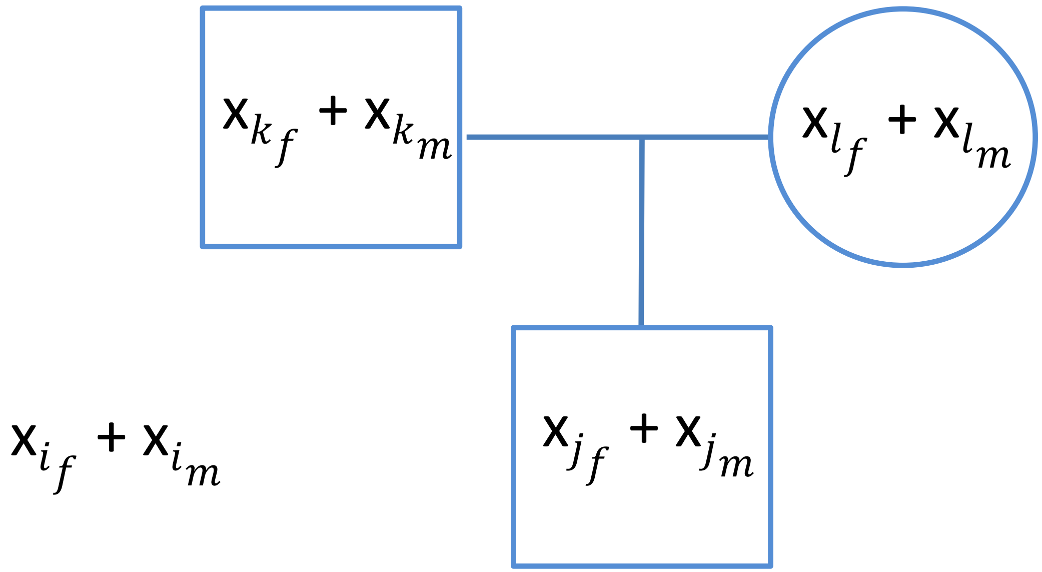
Assume that base individuals at the top of the pedigree (i.e. those without a pedigree) have unrelated alleles i.e. the individuals are unrelated

Recurrence formulae for P(IBD):

$P(x_{i.} \equiv x_{j.})$ = randomly chosen allele in *i* is IBD with randomly chosen allele in *j*

$P(x_{i.} \equiv x_{jf})$ = randomly chosen allele in *i* is IBD with the allele *j* received from their father

Probability that relatives share alleles IBD



Recurrence formulae for P(IBD):

if *i* and *j* are base individuals, $P(x_{i.} \equiv x_{j.}) = 0$

Otherwise,

$$P(x_{i.} \equiv x_{j_f}) = ½[\, P(x_{i.} \equiv x_{k_f}) + P(x_{i.} \equiv x_{k_m})]$$

Relationships between individuals

Usually we analyse measurements on diploid individuals

Cov(g_i, g_j)

= $\mathbf{A}_{ij}$ V_G

= [Cov(x_{i_m},x_{j_m}) + Cov(x_{i_m},x_{j_f}) + Cov(x_{i_f},x_{j_m}) + Cov(x_{i_f},x_{j_f})] V_G

where **A** is the numerator relationship matrix

A = 2 * kinship matrix = 2 * co-ancestry matrix

Estimating genetic variance

Data on phenotypes (y) of related subjects

y = fixed effects + g + e

$V(g) = \mathbf{A} V_G$

$V(e) = \mathbf{I} V_E$

Use ML or REML to estimate variances

Estimating genetic variance

Use ML or REML to estimate variances

- ML finds the value of V_G that maximises the probability of observing the data
- ML estimates all parameters together
  - = estimates variances assuming that fixed effects have been estimated without error
- REML allows for loss of df in estimating fixed effects
- ML $\sigma^2 = \Sigma(y\text{-mean})^2/N$
- REML $\sigma^2 = \Sigma(y\text{-mean})^2/(N\text{-}1)$
- Little difference unless many fixed effects
- Use REML computer programs such as ASREML, GCTA

Intermezzo: In-class demo: (RE)ML

Estimating genetic variance

Example: Data on phenotypes (y) of full sibs

y = fixed effects + g + e

$\mathrm{Cov}(g_i, g_j) = \mathbf{A}_{ij} V_G = 0.5\ V_G$ if *i* and *j* are sibs

Therefore estimate V_G by 2*cov(full-sibs)

h^2 by 2*correlation between full-sibs

What is the covariance between twins?

Model with dominance

Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

genotype	BB	Bb	bb
effect	a	d	-a
frequency	p^2	$2pq$	q^2

$(p + q = 1)$

$$\text{mean} = a * p^2 + d*2pq - a*q^2 = (p - q)*a + 2pqd$$

$$V(\text{genetic effect}) = V_G = E(\text{effect}^2) - E(\text{effect})^2$$

$$V_G = a^2 * p^2 + d^2*2pq + a^2*q^2 - [(p - q)*a + 2pqd]^2$$

$$= 2pq\alpha^2 + (2pqd)^2, \qquad \text{where } \alpha = a + (q - p)d$$

Covariance between genetic effects of relatives

For the covariance between relatives we need to decompose V_G into an additive and dominance variance

Parameterise the genetic value as:

g = mean + additive effect + dominance deviation

g = mean + paternal allele effect + maternal allele effect + interaction of alleles

Genotype	BB	Bb	bb	
effect	a	d	-a	
frequency	p^2	2pq	q^2	(p + q = 1)
mean	(p-q)a + 2pqd	(p-q)a + 2pqd	(p-q)a + 2pqd	
additive	$2q\alpha$	$(q-p)\alpha$	$-2p\alpha$	$\alpha = a + (q - p)d$
dominance dev.	$-q^2d$	2pqd	$-p^2d$	

mean(additive effect) = 0, mean(dominance deviation) = 0

Cov(additive effect, dominance deviation) =0

Genetic variance $= V_G = 2pq\alpha^2 + (2pqd)^2 = V_A + V_D$

Intermezzo: Genetic variance for the 1-locus Fisher/Falconer model

http://cnsgenomics.com/shiny/Falconer/

Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

$$\text{Cov}(g_i, g_j) = \text{Cov}(a_i + d_i, a_j + d_j) = \text{Cov}(a_i, a_j) + \text{Cov}(d_i, d_j)$$
$$= \mathbf{A}_{ij} V_A + \mathbf{D}_{ij} V_D$$

$\mathbf{D}_{ij}$ = P(*i* and *j* inherit the same genotype IBD)

$\mathbf{D}_{ij}$ = 1 for MZ twins, ¼ for full-sibs, 0 for parent-offspring

Covariance between genetic effects of relatives

Model with 1 gene, 2 alleles and additive and dominant gene action

Relationships	MZ twins	full-sibs	halfsibs	P-O
A	1	0.50	0.25	0.5
D	1	0.25	0	0

Therefore can estimate both V_A and V_D by using <u>multiple</u> relationships

Covariance between environmental effects of relatives

y = mean + genetic effect + common environment effect

+ individual environment effect

$y = \text{mean} + g + e_c + e$

$\text{Cov}(e_{c_i}, e_{c_j}) = V_C$ if *i* and *j* in same family, zero otherwise

Relationships	MZ twins	full-sibs.	halfsibs	P-O
A	1	0.50	0.25	0.5
D	1	0.25	0	0
E_c	1	1	?	?

Covariance between relatives

Estimating V_A, V_D and V_C

Difficult (in humans)!

Assume $V_D = 0$

$V_A = 2[\mathrm{Cov}(\text{MZ twins}) - \mathrm{Cov}(\text{full-sibs})]$

Relationships	MZ twins	full-sibs.	halfsibs	P-O
A	1	0.50	0.25	0.5
D	1	0.25	0	0
E_c	1	1	?	?

Covariance between relatives

Can add epistatic interactions to model

g = mean + additive + dominance + epistasis

g = mean + a + d + aa

$$\mathrm{Cov}(g_i, g_j) = \mathbf{A}_{ij} V_A + \mathbf{D}_{ij} V_D + \mathbf{A}_{ij}^2 V_{AA}$$

Relationships	MZ twins	full-sibs	halfsibs	P-O
A	1	0.50	0.25	0.5
D	1	0.25	0	0
AxA	1	0.25	0.0625	0.25

Summary
Resemblance between relatives

Model phenotypes by fixed effects and random effects including genetic value (additive, dominance, epistatic)

Model covariance of genetic effects by relationship estimated from pedigree (or SNP genotypes)

Estimate genetic variance by REML

Empirical results on the resemblance between relatives in humans

1. Pearson & Lee 1903

2. From 1903 to 2010: correlations are stable over time

3. Height and BMI in Sweden: simple models don't explain empirical results

VOLUME II NOVEMBER, 1903 No. 4

ON THE LAWS OF INHERITANCE IN MAN*.

I. INHERITANCE OF PHYSICAL CHARACTERS.

BY KARL PEARSON, F.R.S., assisted by ALICE LEE, D.Sc.
University College, London.

DIAGRAM IV. *Distribution of Stature.*

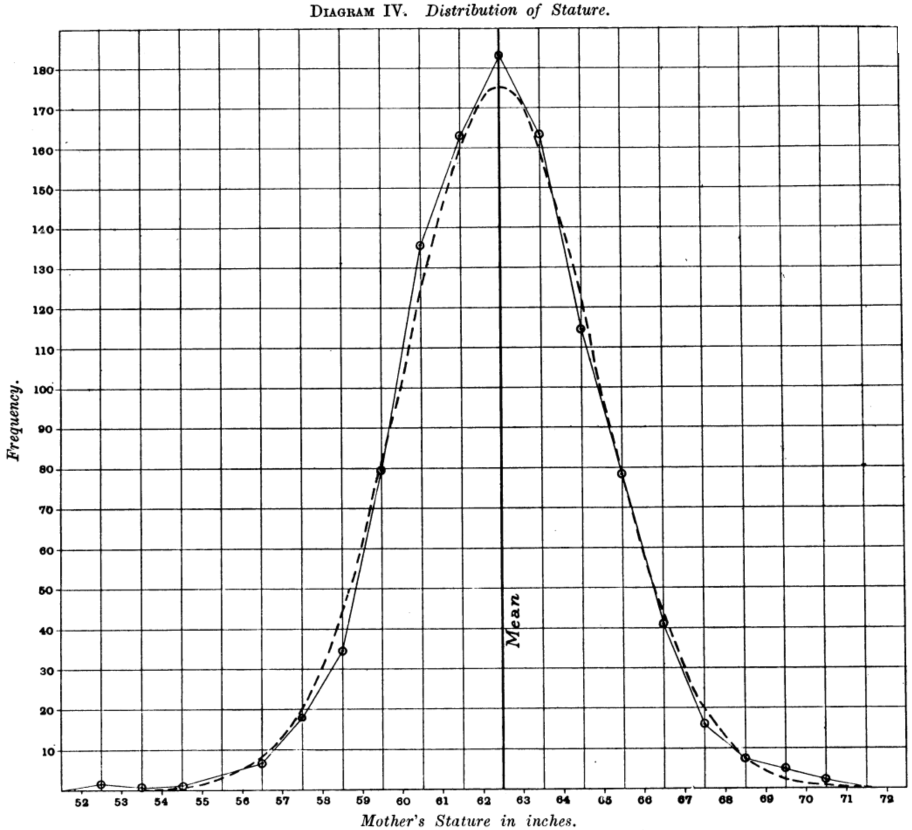
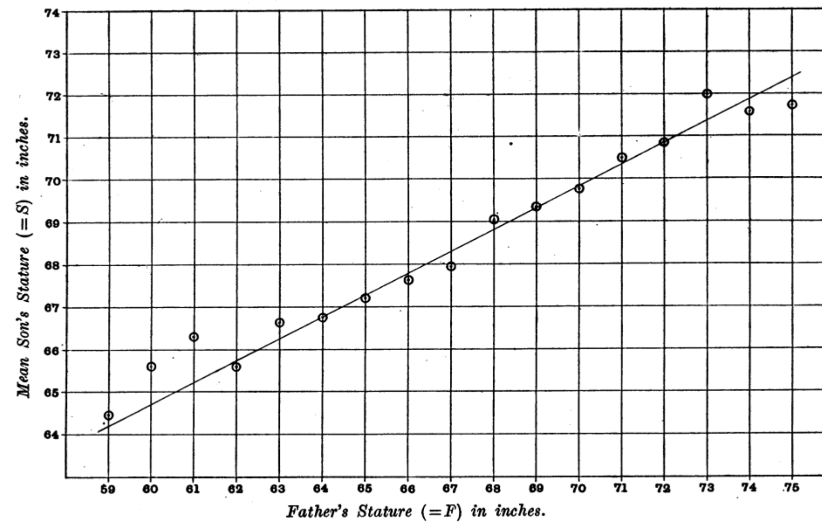


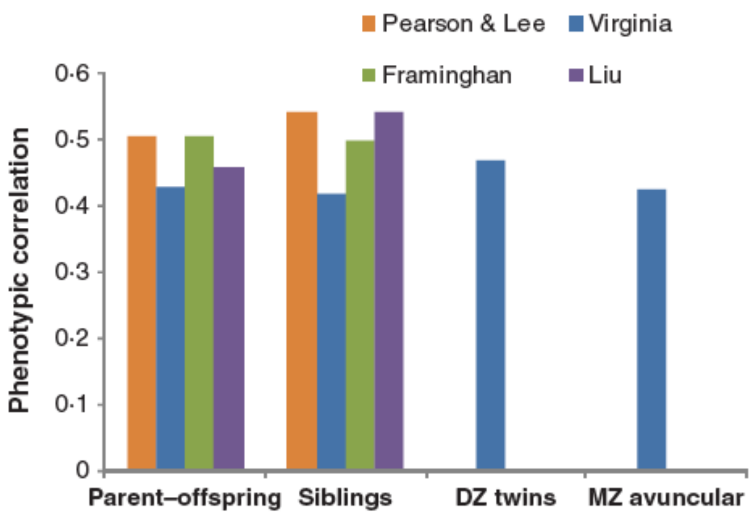
DIAGRAM I. *Probable Stature of Son for given Father's Stature.*

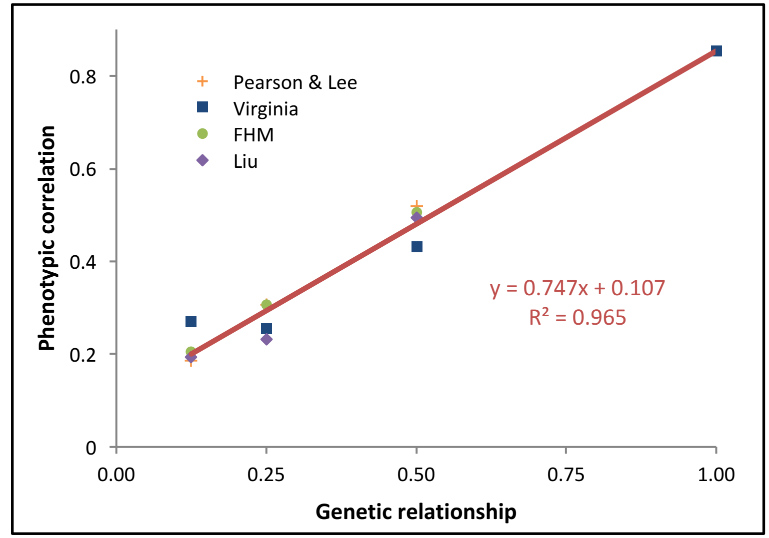
Regression Line: $S = 33{\cdot}73 + {\cdot}516\,F$. 1078 Cases.

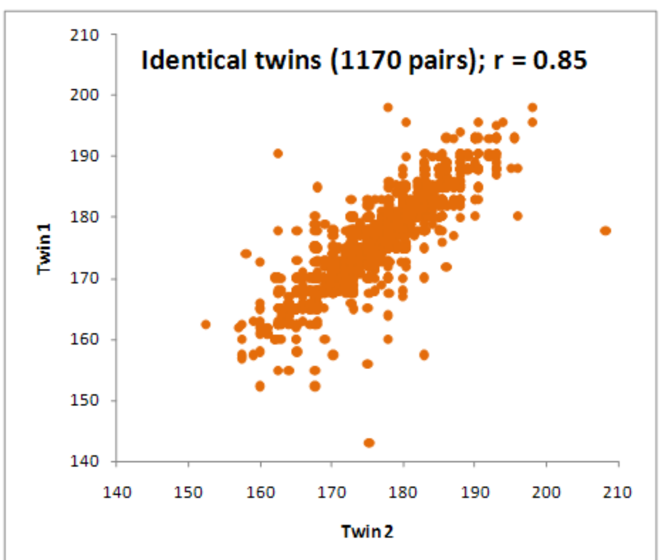


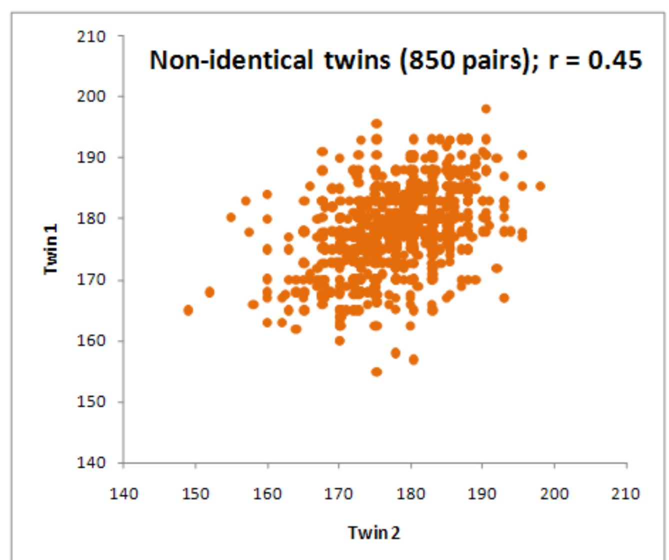
PAIR	CORRELATION	SE
Spouse	0.28	0.02
Son-Father	0.51	0.02
Daughter-Father	0.51	0.01
Son-Mother	0.49	0.02
Daughter-Mother	0.51	0.01
Brother-brother	0.51	0.03
Sister-sister	0.54	0.02
Brother-sister	0.55	0.01

Resemblance between relatives (height)



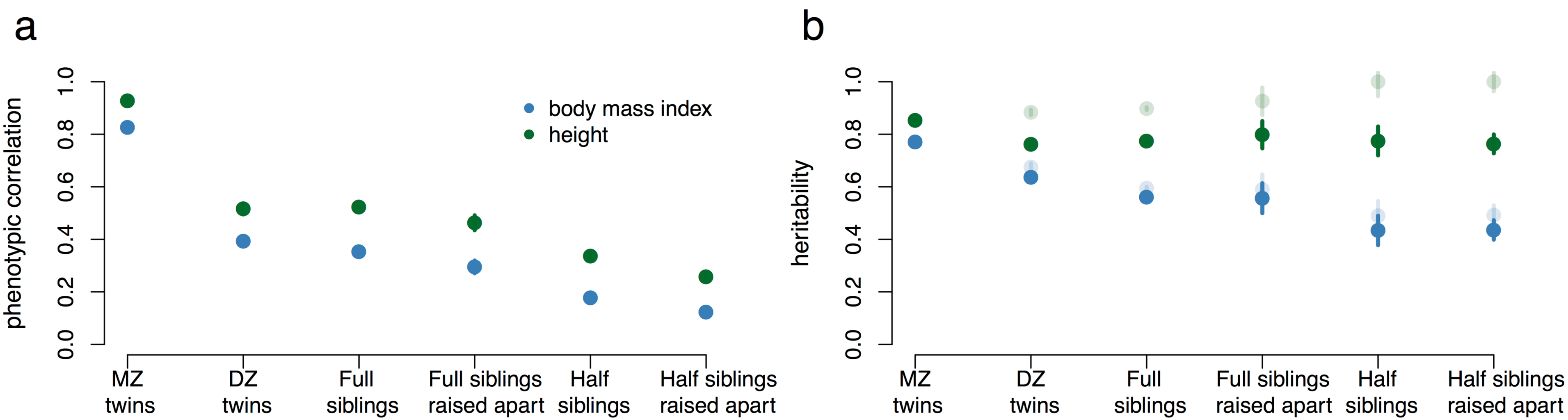






More data on height (and BMI)

Data from ~172,000 18-year old brother pairs



[Robinson et al. Nature Genetics]