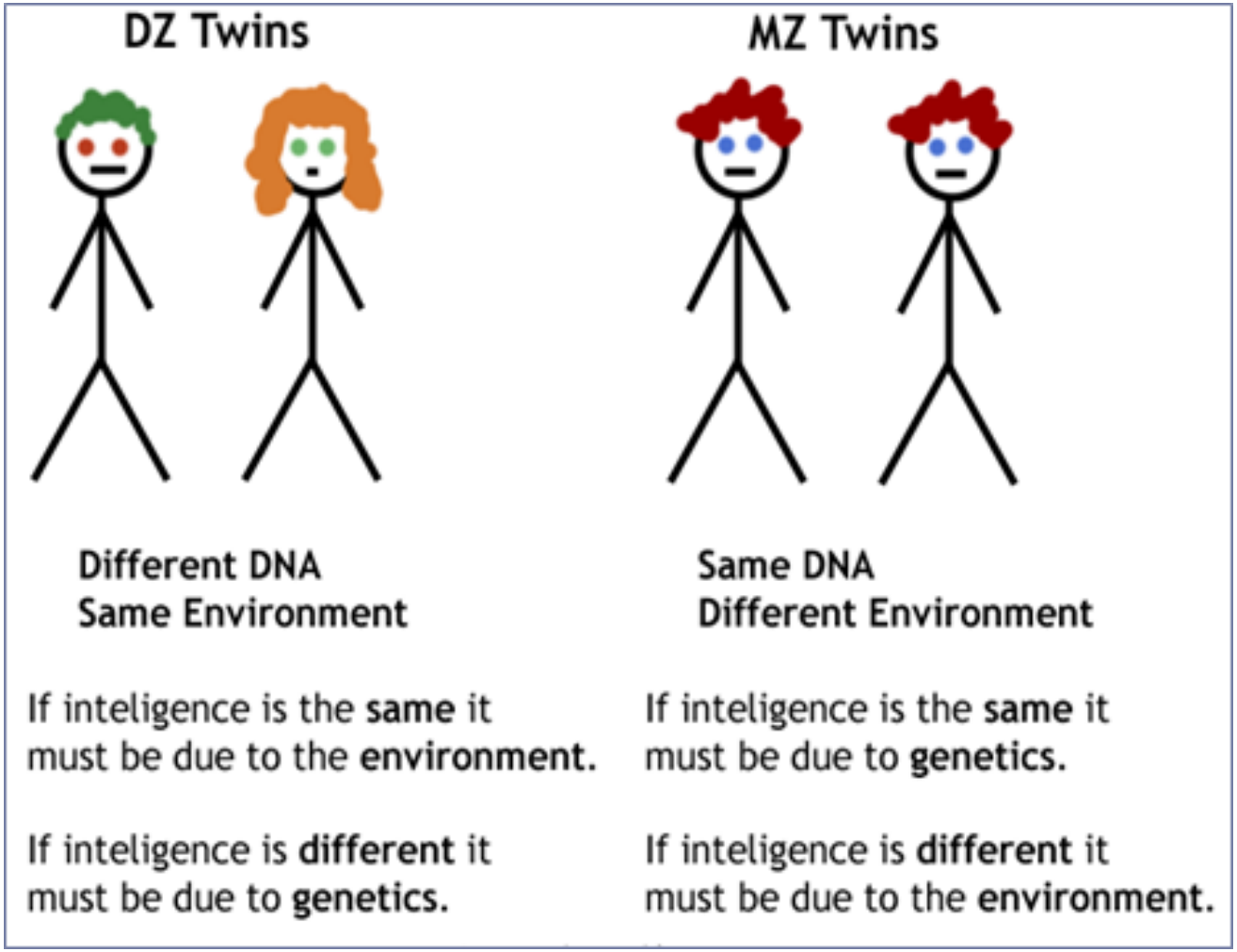
Twin studies



Used to determine the heritability of a trait

Heritability

- The **heritability** (h^2) of a trait is a measure of the degree of similarity between relatives

$$h^2 = \frac{V_A}{V_P} = \frac{V_A}{V_G + V_E}$$

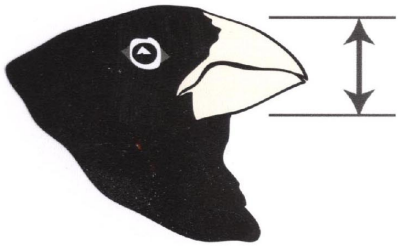
- **V_P (Total phenotypic variance)**
- **V_A (Additive genetic variance)** - variation due to the additive effects of alleles
- **V_G (Genetic variance)** - the variance among the mean phenotypes of different genotypes
- **V_E (Environmental variance)** - the variance among phenotypes expressed by replicate members of the same genotype

Heritability

$$h^2 = \frac{V_A}{V_P} = \frac{V_A}{V_G + V_E} = \frac{V_A}{(V_A + V_D + V_I) + V_E}$$

- Since heritability is a function of the environment (V_E), it is a environment-dependent measure
- Heritability ranges from 0 to 1 (e.g. traits with no genetic variation have a heritability of 0)
- Heritability is usually estimated by family-based study

Estimating heritability using regression

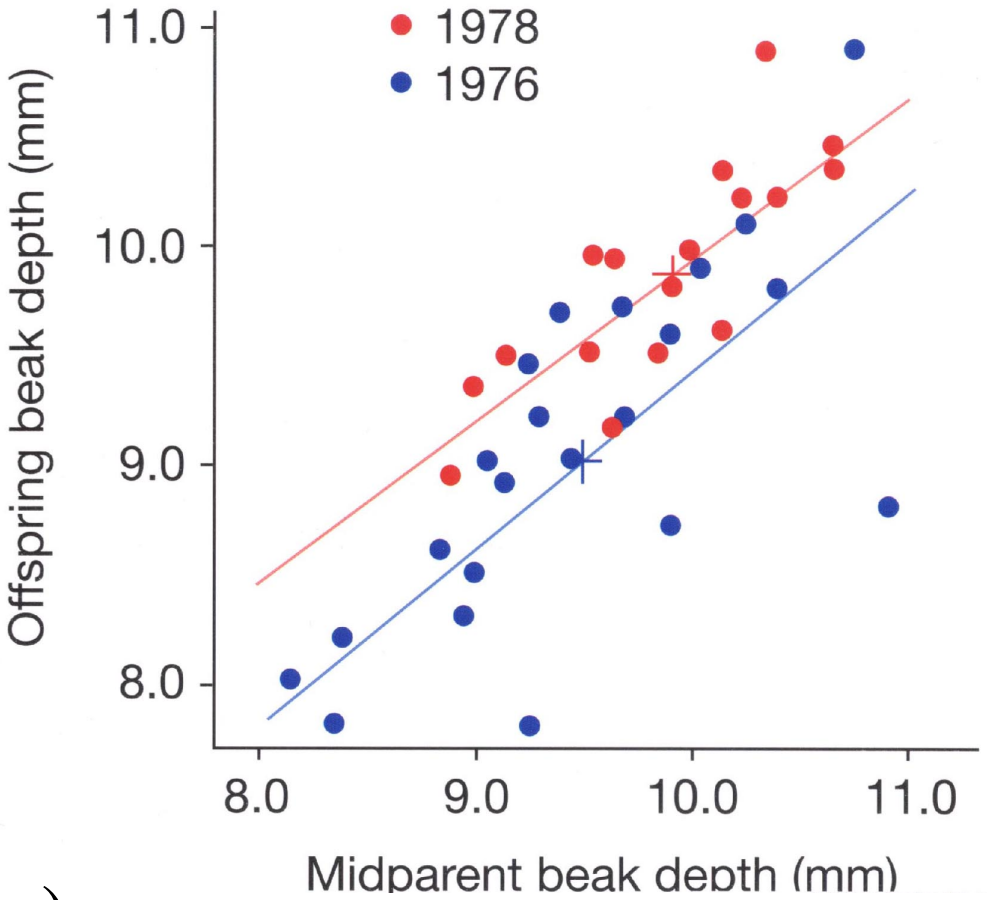


Beak depth of Darwins' finches

$$Cov(x,y) = \frac{1}{N}\sum (x_i - \bar{x})(y_i - \bar{y})$$

$$Var(x) = \frac{1}{N}\sum (x_i - \bar{x})^2$$

$$h^2 = slope = \frac{Cov(x,y)}{Var(x)}$$



Using unrelated individuals

- However, estimates of h^2 become less precise as number of close relatives in the sample decreases
- It is difficult to estimate **SNP-based heritability** using genetic data from unrelated individuals

(2010)

nature genetics

ANALYSIS

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

SNP-based heritability

Using unrelated individuals has some key advantages:

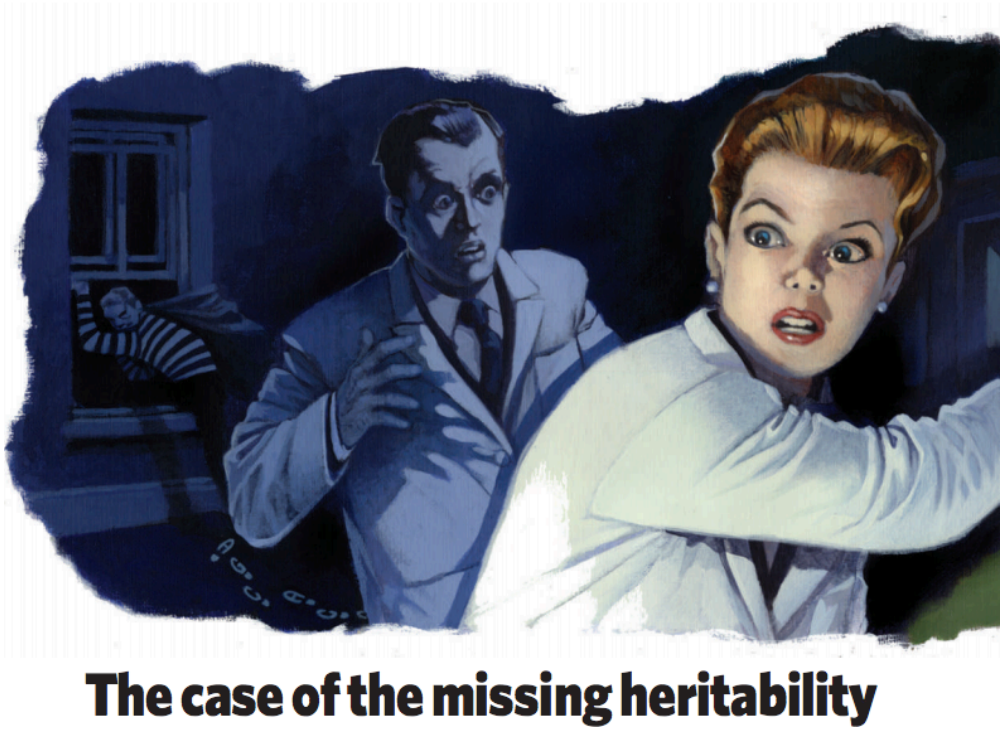
- The resulting estimate, referred to as "SNP-based heritability" (h^2_{SNP}), is an estimate of the total variance explained by all SNPs
- Can use GWAS data, so sample sizes are much larger than using family data
- Can discover a large number of common SNPs with effect sizes too small to pass the stringent GWAS threshold ($p<5\times10^{-8}$), especially for complex traits

- **SNP-based heritability** (h^2_{SNP}): the degree to which individual genetic variation (i.e. SNPs) accounts for phenotypic variation seen in a population

The missing heritability problem --- GWAS

- Although GWAS has found a number of associations for a wide range of phenotypes, the proportion of variance explained by the associations for any particular phenotype (h^2_{GWAS}) was typically slight compared to the phenotype's heritability

- Before 2012, human geneticists were referring to the “missing heritability problem”



Maher, Nature news feature (2008)

The missing heritability problem --- Example

- The classic example was height
- The heritability is about 80%. In 2008, 20 associations had been found, but these explained only a few percent of variation (Genome-wide association analysis identifies 20 loci, *Nature Genetics*)

Genome-wide association analysis identifies 20 loci that influence adult height

Michael N Weedon[1,2,23], Hana Lango[1,2,23], Cecilia M Lindgren[3,4], Chris Wallace[5], David M Evans[6], Massimo Mangino[7], Rachel M Freathy[1,2], John R B Perry[1,2], Suzanne Stevens[7], Alistair S Hall[8], Nilesh J Samani[7], Beverly Shields[2], Inga Prokopenko[3,4], Martin Farrall[9], Anna Dominiczak[10], Diabetes Genetics Initiative[21], The Wellcome Trust Case Control Consortium[21], Toby Johnson[11–13], Sven Bergmann[11,12], Jacques S Beckmann[11,14], Peter Vollenweider[15], Dawn M Waterworth[16], Vincent Mooser[16], Colin N A Palmer[17], Andrew D Morris[18], Willem H Ouwehand[19,20], Cambridge GEM Consortium[22], Mark Caulfield[5], Patricia B Munroe[5], Andrew T Hattersley[1,2], Mark I McCarthy[3,4] & Timothy M Frayling[1,2]

Adult height is a model polygenic trait, but there has been limited success in identifying the genes underlying its normal variation. To identify genetic variants influencing adult human height, we used genome-wide association data from 13,665 individuals and genotyped 39 variants in an additional 16,482 samples. We identified 20 variants associated with adult height ($P < 5 \times 10^{-7}$, with 10 reaching $P < 1 \times 10^{-10}$). Combined, the 20 SNPs explain ~3% of height variation, with a ~5 cm difference between the 6.2% of people with 17 or fewer 'tall' alleles compared to the 5.5% with 27 or more 'tall' alleles. The loci we identified implicate genes in Hedgehog signaling (*IHH, HHIP, PTCH1*), extracellular matrix (*EFEMP1, ADAMTSL3, ACAN*) and cancer (*CDK6, HMGA2, DLEU7*) pathways, and provide new insights into human growth and developmental processes. Finally, our results provide insights into the genetic architecture of a classic quantitative trait.

The missing heritability problem SOLVED

ANALYSIS

nature genetics

Common SNPs explain a large proportion of the heritability for human height

Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

SNPs discovered by ge
account for only a sma
complex traits in huma
heritability? We estima
human height explaine
3,925 unrelated indivi
validated the estimatio
the observed genotype
can be explained by co
most of the heritability
been detected because
to pass stringent signifi
that the remaining heri
disequilibrium betwee
exacerbated by causal
frequency than the SN

ANALYSIS

nature genetics

Genome partitioning of genetic variation for complex traits using common SNPs

Jian Yang[1]*, Teri A Manolio[2], Louis R Pasquale[3], Eric Boerwinkle[4], Neil Caporaso[5], Julie M Cunningham[6], Mariza de Andrade[7], Bjarke Feenstra[8], Eleanor Feingold[9], M Geoffrey Hayes[10], William G Hill[11], Maria Teresa Landi[12], Alvaro Alonso[13], Guillaume Lettre[14], Peng Lin[15], Hua Ling[16], William Lowe[17], Rasika A Mathias[18], Mads Melbye[8], Elizabeth Pugh[16], Marilyn C Cornelis[19], Bruce S Weir[20], Michael E Goddard[21,22] & Peter M Visscher[1]

ited. That is, the proportion
common SNPs depends on
se SNPs. For these reasons,
ted for by SNPs that reach
rtain to be smaller than the

o focus on the estimation of
r. Recently, we showed how
% of phenotypic variation
mmon SNPs from a sample
he British Isles[4]. In a sepa-
ce for height onto chromo-
which captures the effects
the variance was explained
Here we take these studies
1,586 unrelated European
its. We partitioned additive

Table 1 Estimates of the variance explained by all autosomal SNPs for height, BMI, vWF and QTi

		No PC[a]		10 PCs[b]			
Trait	*n*	h^2_G (s.e.)[c]	*P*	h^2_G (s.e.)	*P*	Heritability[d]	GWAS[e]
Height	11,576	0.448 (0.029)	4.5×10^{-69}	0.419 (0.030)	7.9×10^{-48}	80–90%[32]	~10%[23]
BMI	11,558	0.165 (0.029)	3.0×10^{-10}	0.159 (0.029)	5.3×10^{-9}	42–80%[25,26]	~1.5%[14]
vWF	6,641	0.252 (0.051)	1.6×10^{-7}	0.254 (0.051)	2.0×10^{-7}	66–75%[33,34]	~13%[15]
QTi	6,567	0.209 (0.050)	3.1×10^{-6}	0.168 (0.052)	5.0×10^{-4}	37–60%[35,36]	~7%[16]

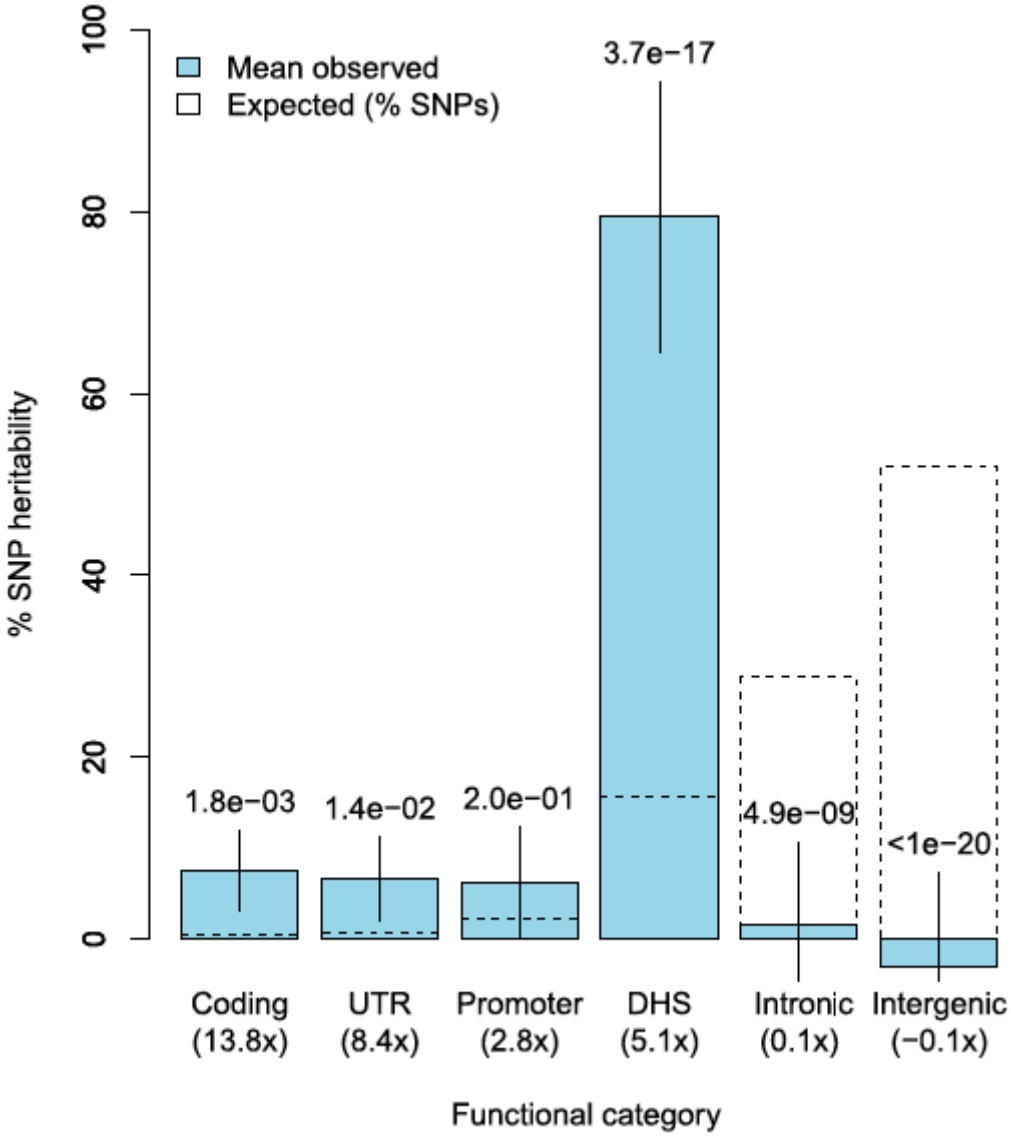
h^2_{SNP} estimated with unrelated individuals explain more total phenotypic variance than h^2_{GWAS} ($h^2_{SNP} >> h^2_{GWAS}$)

h^2_{SNP} explained by DNase I hypersensitivity sites

11 traits in this analysis

6 autoimmune traits
- Ankylosing spondylitis
- Crohn disease
- Multiple sclerosis
- Rheumatoid arthritis
- Type I diabetes
- Ulcerative colitis

5 non-autoimmune traits
- Schizophrenia
- Bipolar disorder
- Coronary artery disease
- Hypertension
- Type II diabetes



Gusev et al (2014) AJHG

How to calculate h^2_{SNP}

- Collect GWAS data for a particular trait (e.g. > 5000 individuals with genome-wide genotyping)
- Compute allelic correlations *K*
- Remove individuals so that no pair remains with $K_{i,j} > 0.05$ (closely related individuals)
- Perform **GREML model** to estimate h^2_{SNP}

How to calculate h^2_{SNP}

- Collect GWAS data for a particular trait (e.g. > 5000 individuals with genome-wide genotyping)
- Compute allelic correlations *K*
- Remove individuals so that no pair remains with $K_{i,j} > 0.05$ (closely related individuals)
- Perform **GREML/GCTA model** to estimate h^2_{SNP}

$$y = g + e$$

Phenotypes

Additive genetic variance

$$g \sim N(0, A\sigma^2_{g(SNP)})$$

Genetic relationship matrix (GRM)

Residuals

$$e \sim N(0, I\sigma^2_e)$$

Identity matrix

How to calculate h^2_{SNP}

- Collect GWAS data for a particular trait (e.g. > 5000 individuals with genome-wide genotyping)
- Compute allelic correlations *K*
- Remove individuals so that no pair remains with $K_{i,j} > 0.05$ (closely related individuals)
- Perform **GREML/GCTA model** to estimate h^2_{SNP}

$$y = g + e$$

Phenotypes

Additive genetic variance

$$g \sim N(0, A\sigma^2_{g(SNP)})$$

Genetic relationship matrix (GRM)

Residuals

$$e \sim N(0, I\sigma^2_e)$$

Identity matrix

$$h^2_{SNP} = \frac{\sigma^2_{g(SNP)}}{\sigma^2_{g(SNP)} + \sigma^2_e}$$