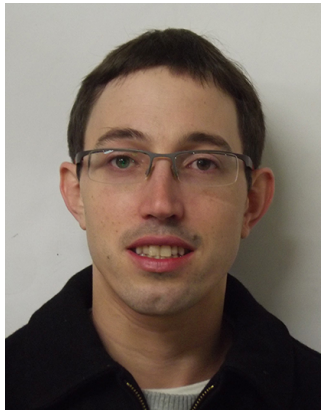
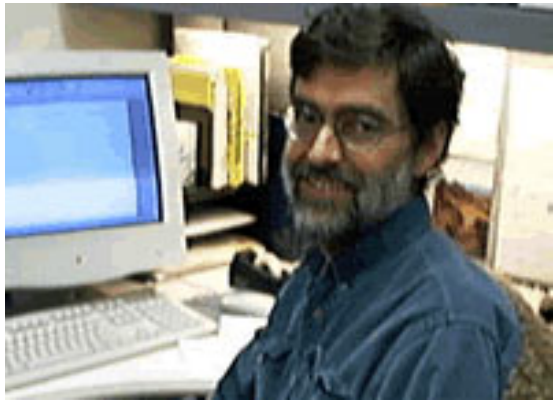
A population genetic model of disease susceptibility and other quantitative traits

Guy Sella
Columbia University



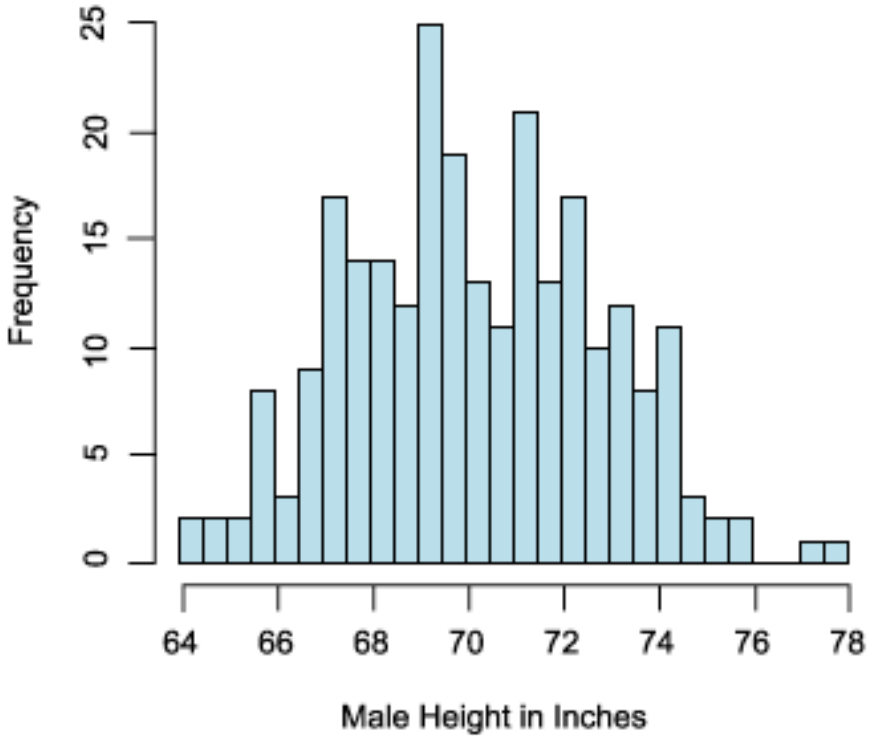
Yuval Simons
(Columbia U.)

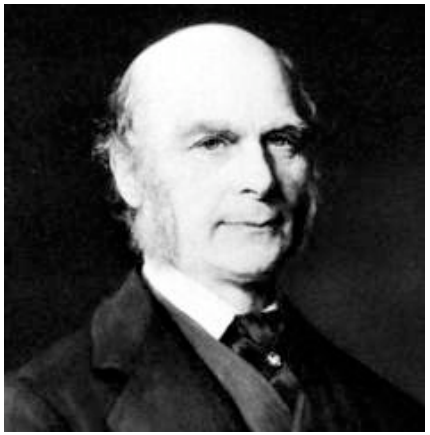
Kevin Bullaughey
(U. Chicago)



Dick Hudson
(U. Chicago)

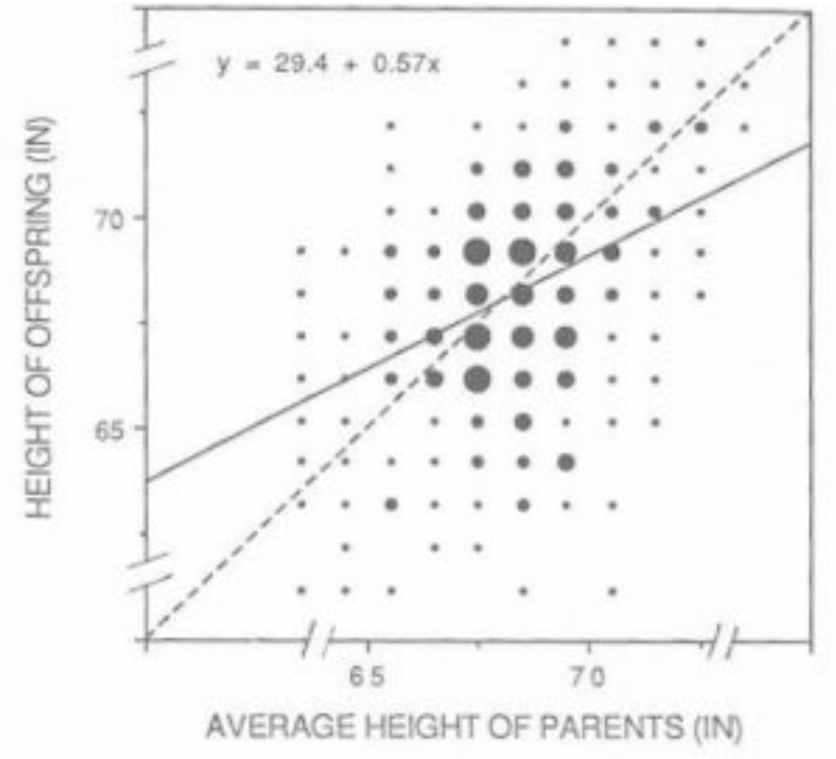
Phenotypic variation in the population is often normally distributed

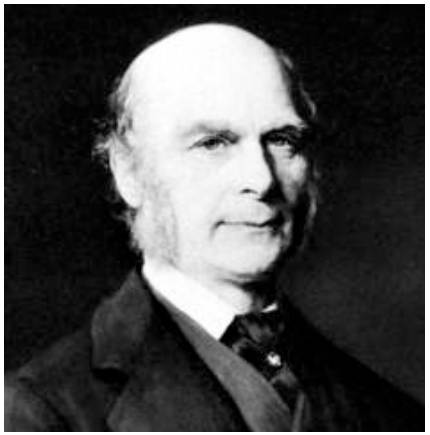




Francis Galton

A substantial portion of this variation is heritable





Francis Galton

An additive model explains these observations

$$P = X_G + X_E = \Sigma_i [X_i^m + X_i^p] + X_E$$

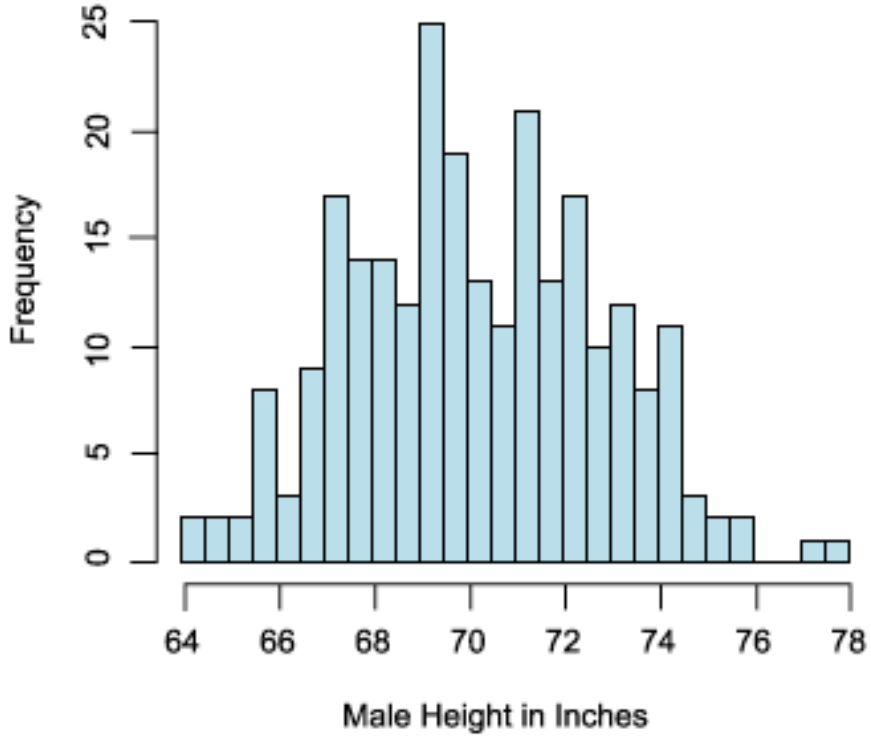
Fisher 1918

X_G & X_E – genetic & environmental contributions

X_i^m & X_i^p – the maternal & paternal additive genetic contribution at locus i

If we assume that $X_E \sim N(0,V_E)$ and the X_is are i.i.d over many loci then

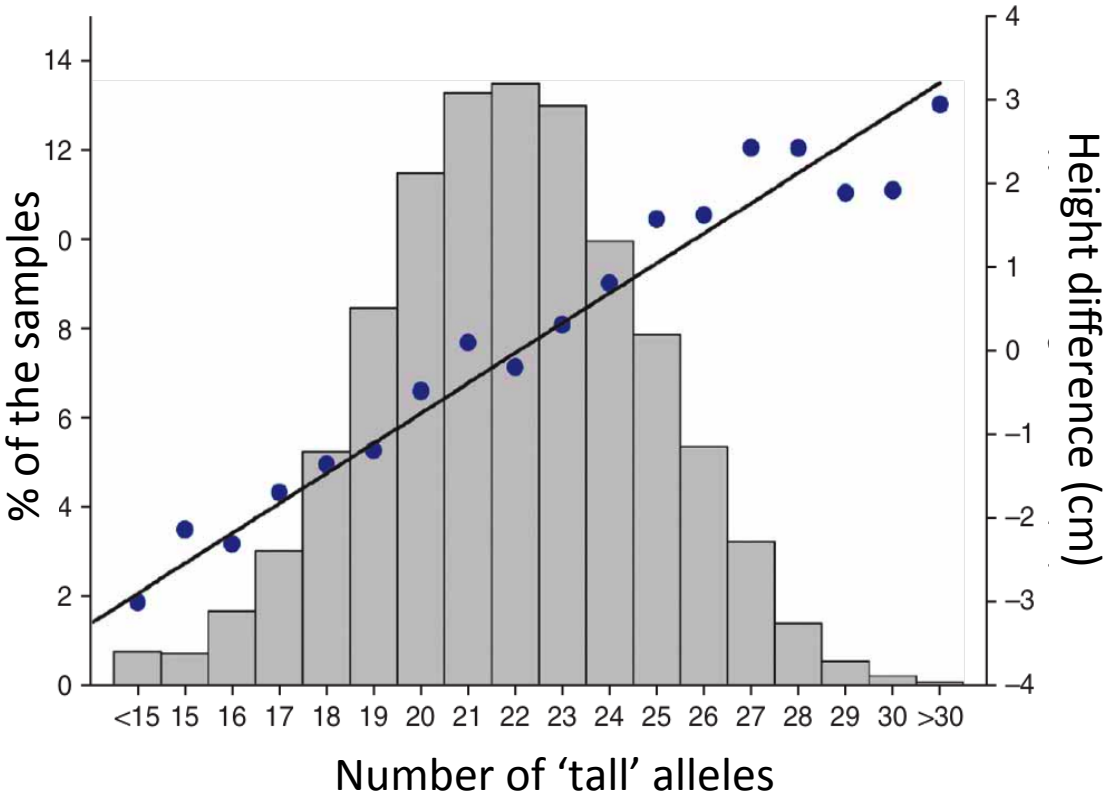
$$P \sim N(\mu_P, V_G+V_E)$$



An example: height

The top 20 loci in the genome underlying Height variation.

Male s.d. difference (95% CI)	Female s.d. difference (95% CI)
0.07 (0.04, 0.09)	0.12 (0.09, 0.14)
0.09 (0.06, 0.12)	0.08 (0.05, 0.11)
0.05 (0.03, 0.08)	0.07 (0.05, 0.10)
–0.08 (–0.11, –0.05)	–0.07 (–0.10, –0.04)
–0.07 (–0.11, –0.03)	–0.07 (–0.11, –0.03)
0.06 (0.03, 0.08)	0.05 (0.03, 0.08)
0.09 (0.05, 0.12)	0.06 (0.03, 0.10)
0.09 (0.05, 0.13)	0.10 (0.06, 0.14)
0.05 (0.02, 0.07)	0.08 (0.05, 0.11)
–0.08 (–0.11, –0.05)	–0.03 (–0.06, 0.00)
–0.04 (–0.07, –0.01)	–0.05 (–0.08, –0.02)
0.06 (0.03, 0.08)	0.04 (0.01, 0.017)
0.04 (0.01, 0.08)	0.07 (0.03, 0.10)
–0.04 (–0.07, 0.00)	–0.09 (–0.12, –0.06)
–0.05 (–0.08, –0.02)	–0.05 (–0.08, –0.02)
0.05 (0.02, 0.07)	0.04 (0.02, 0.07)
–0.06 (–0.10, –0.02)	–0.05 (–0.10, –0.01)
–0.06 (–0.09, –0.03)	–0.05 (–0.08, –0.02)
0.04 (0.01, 0.06)	0.06 (0.03, 0.09)
0.05 (0.02, 0.08)	0.04 (0.01, 0.07)



Weedon et al 2008 Nat. Genetics

An additive model explains these observations

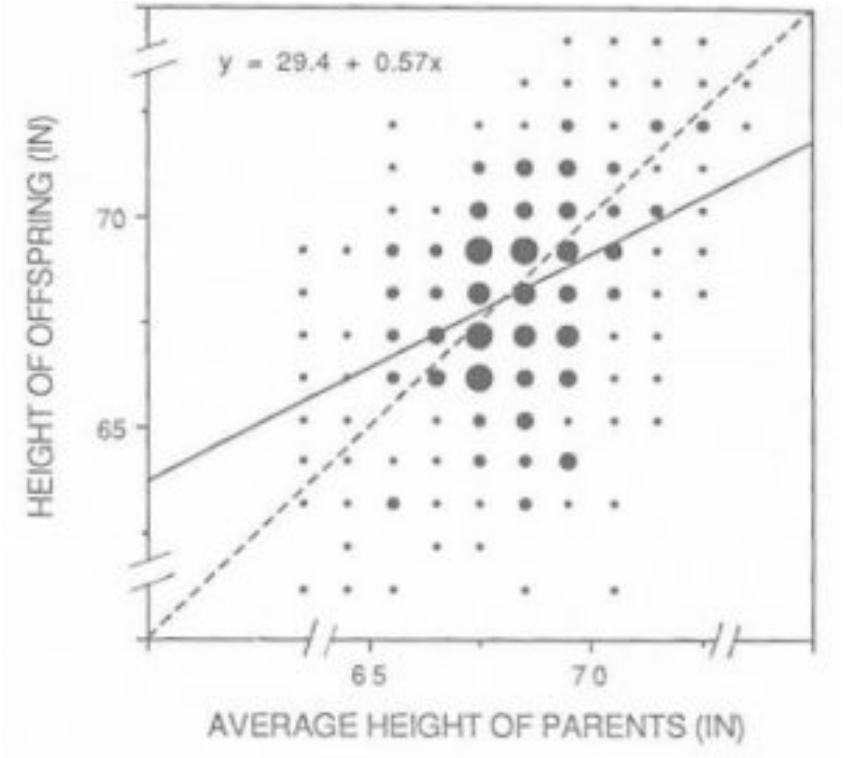
Defining the (narrow sense) heritability as:

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

One can show that:

$$E\left(P_O \middle| P_M = x\right) = h^2 x$$

Fisher 1918



The heritability for most traits is substantial (0.1-0.9)

- Examples from humans (from Byars et al. PNAS 2009):

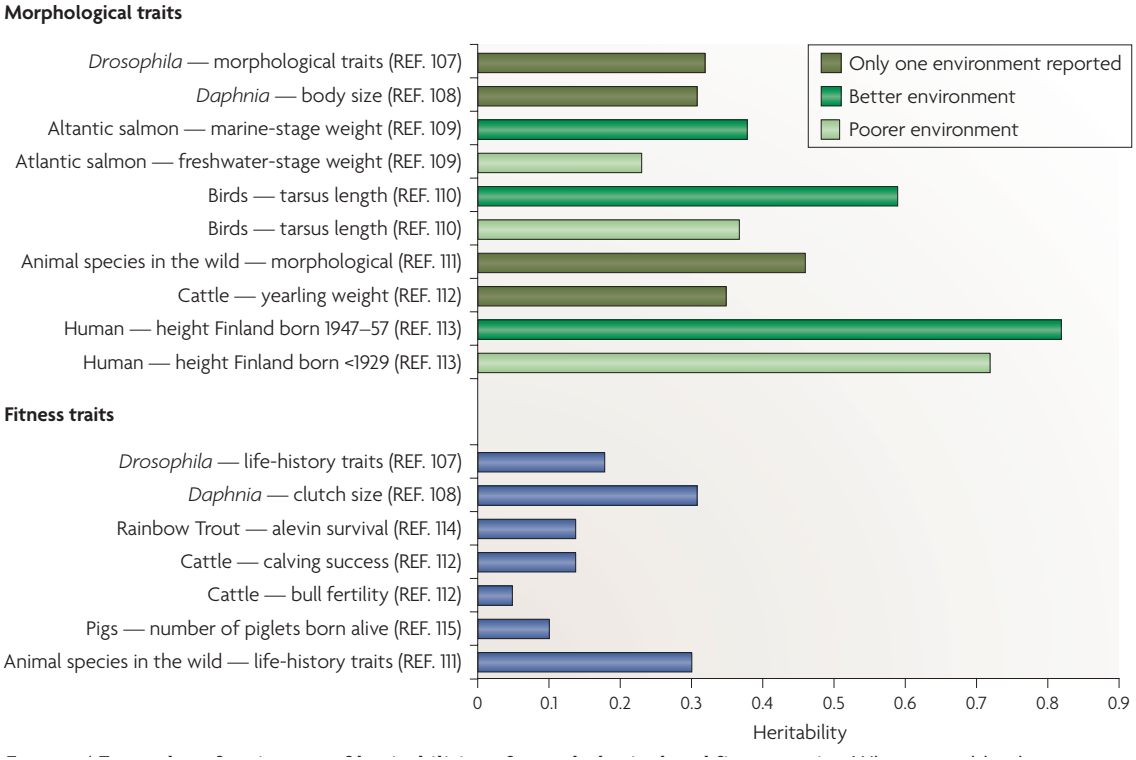
  Height: $h^2 = 0.84$

  Weight: $h^2 = 0.52$

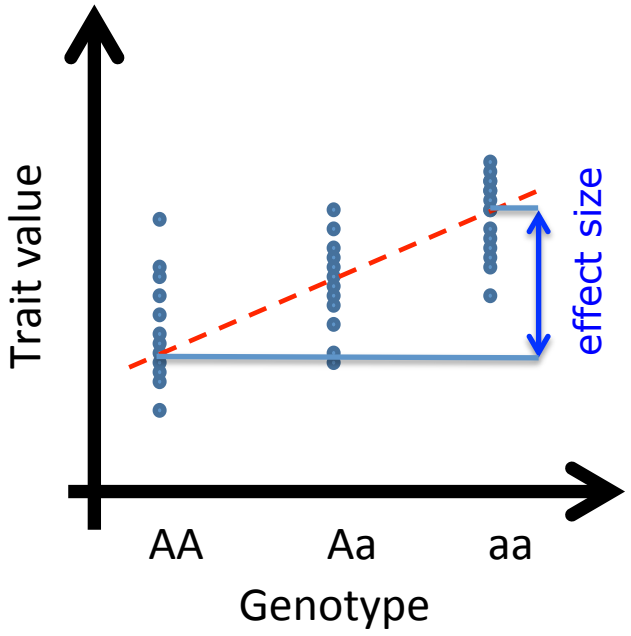
  Total Cholesterol: $h^2 = 0.61$

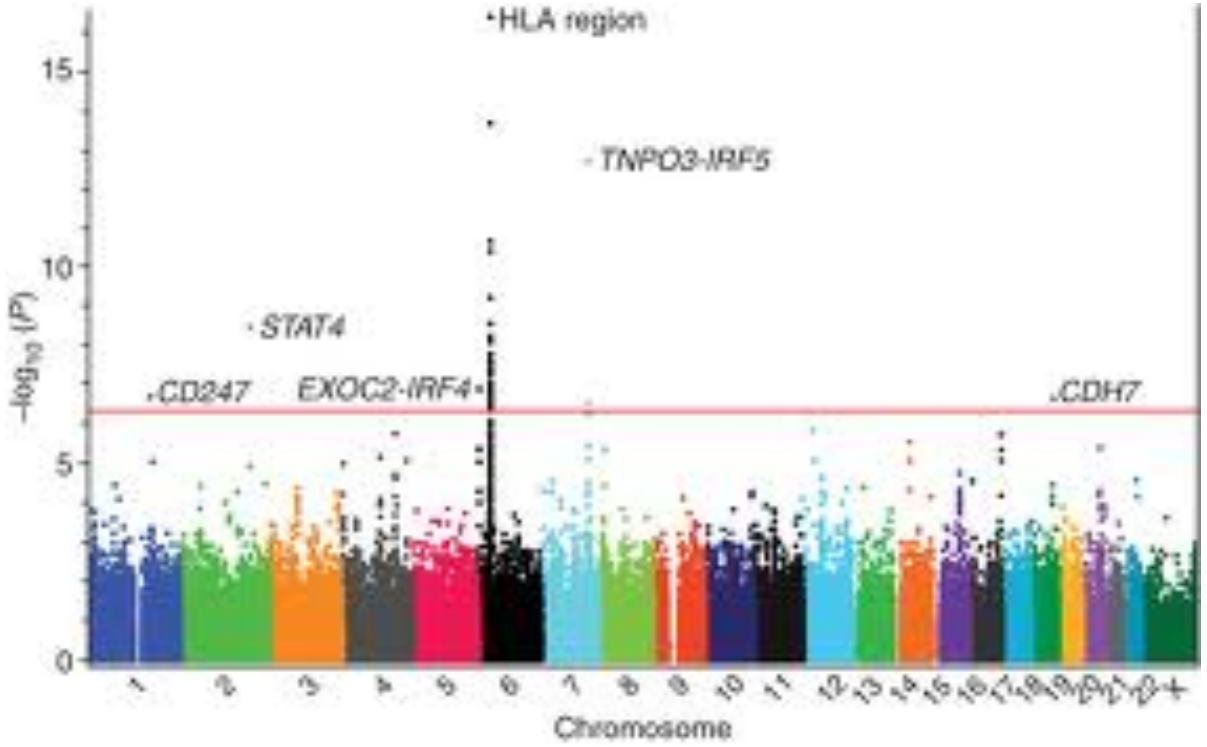
  Age at menopause: $h^2 = 0.47$

  Age at menarche: $h^2 = 0.62$



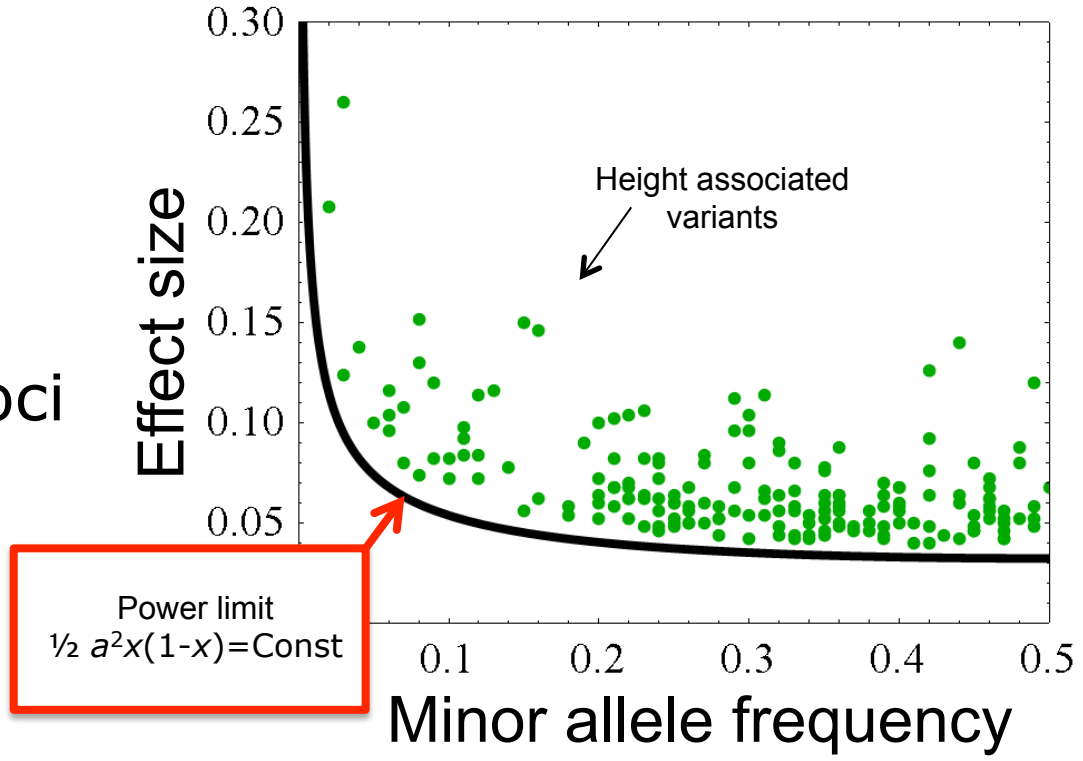
Genome Wide Association Studies





More than 180 SNPs associated with height

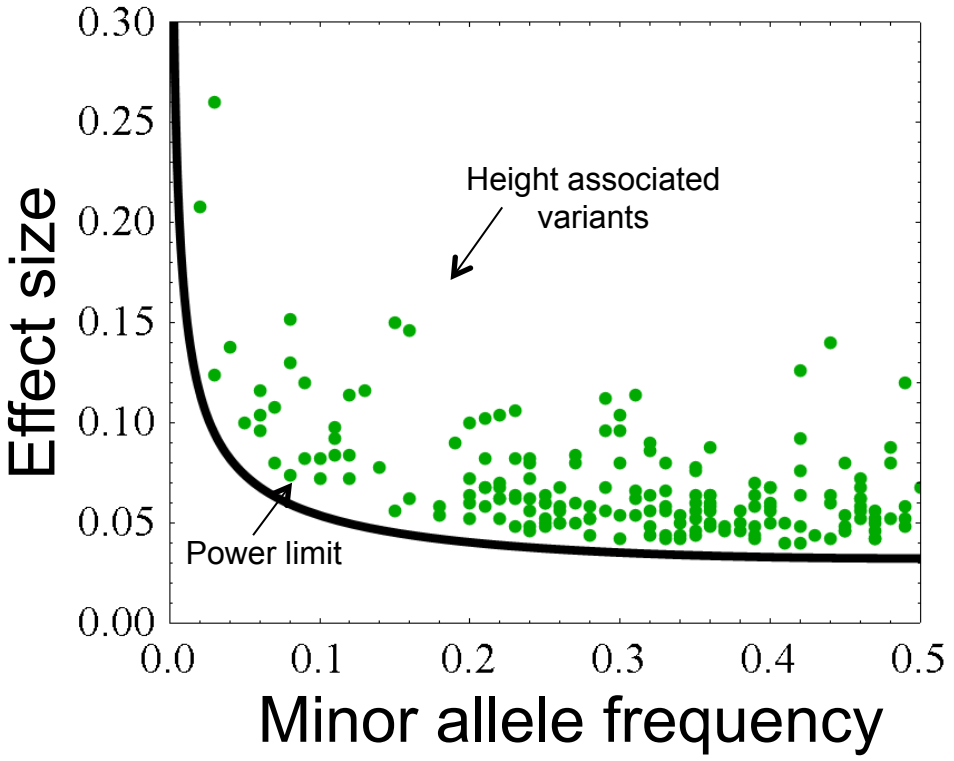
- Genetic Investigation of Anthropometric Traits
- More than 183,000 individuals
- Over 2.5 Million SNPs
- 180 height affecting loci



Missing heritability

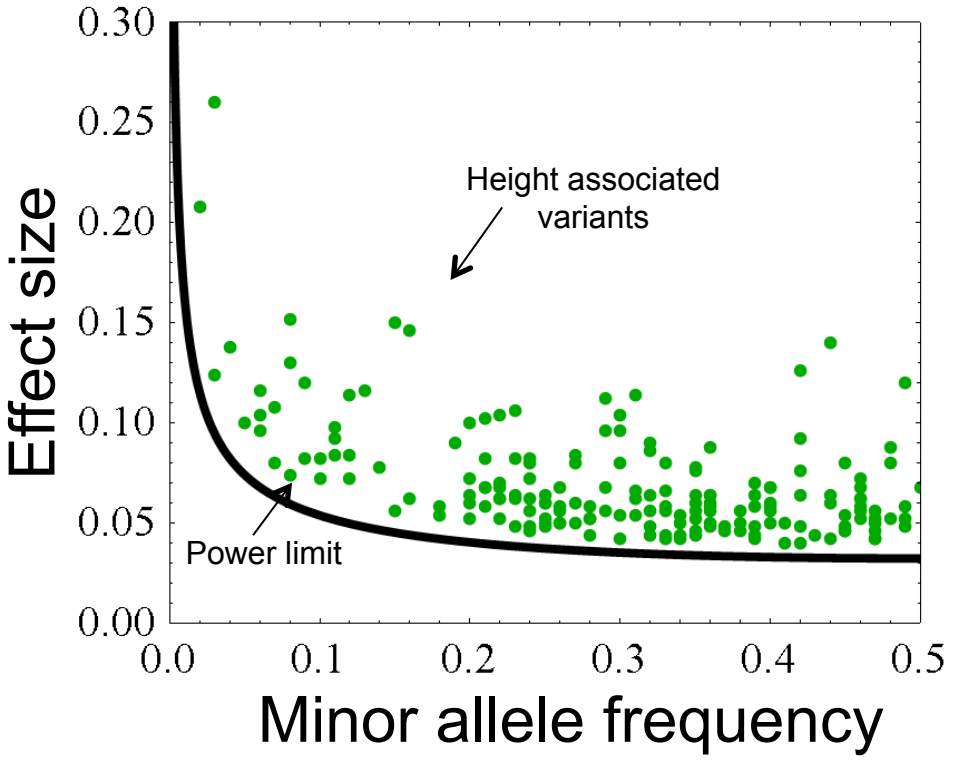
- Genetic Investigation of Anthropometric Traits
- More than 183,000 individuals
- Over 2.5 Million SNPs
- 180 height affecting loci

But, explains only ~10% of the heritable variation in height



Genetic architecture

The number of ALL the variants and their joint distribution of frequencies and effect sizes



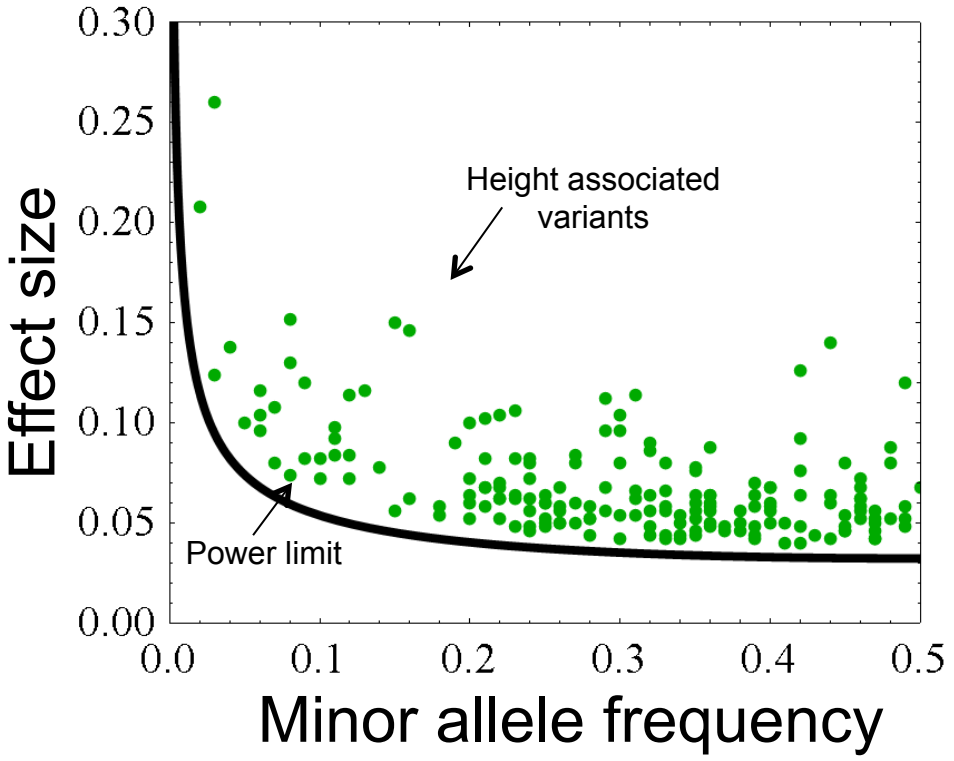
Lango Allen *et al.* , Nature **467**, 832 (2010)

Genetic architecture

The number of ALL the variants and their joint distribution of frequencies and effect sizes

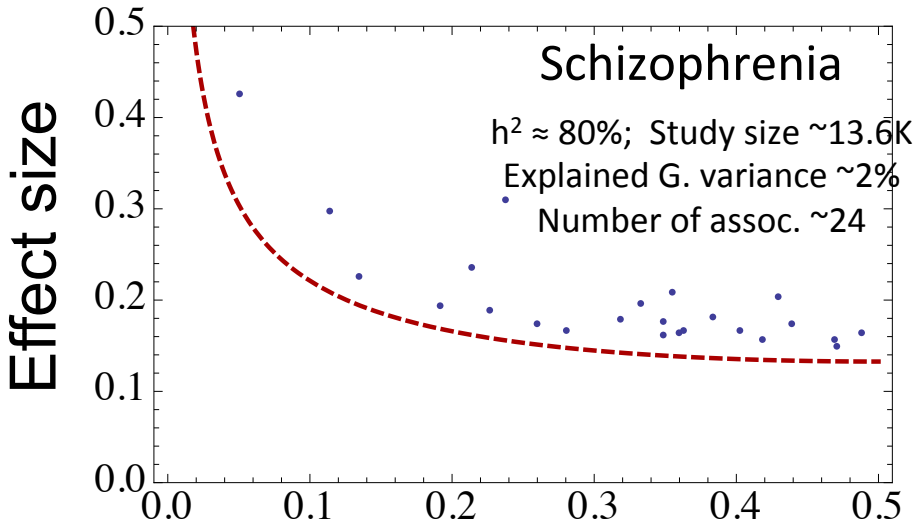
With limited power in GWAS, we can look at:

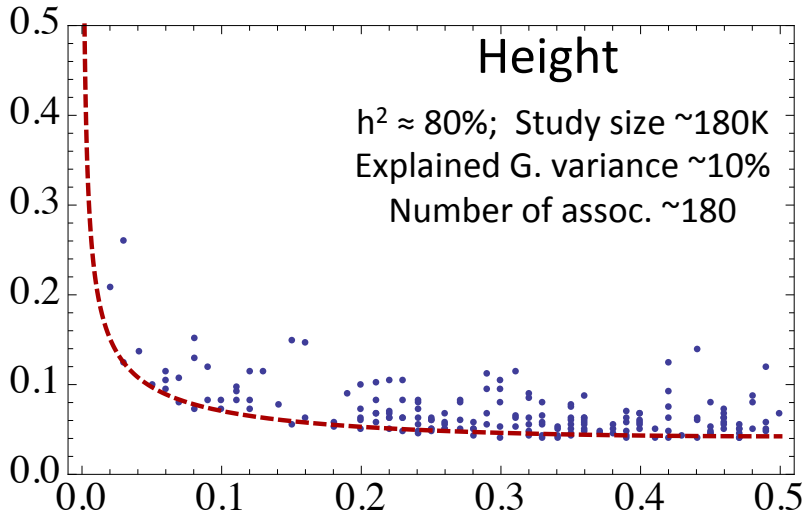
- The discovered number and distribution.
- AND the explained heritability.

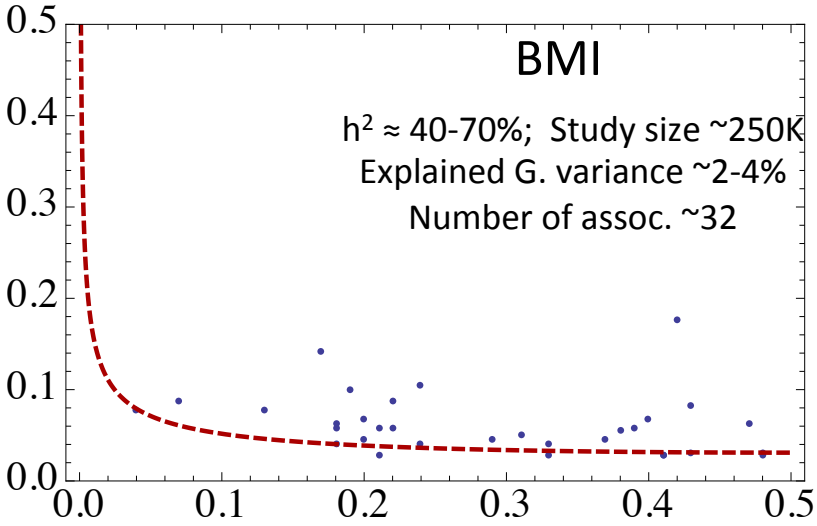


Lango Allen *et al.* , Nature **467**, 832 (2010)

Genetic architecture







While preliminary, we already see differences in:

- Explained variance
- Number discovered

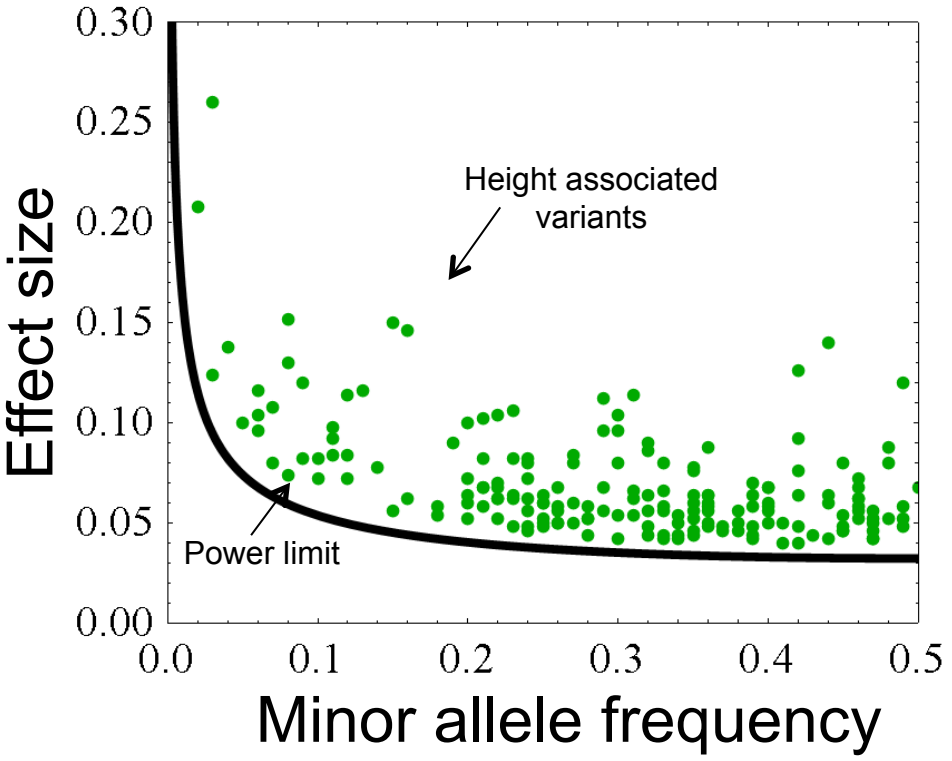
And

- The distribution

Ripke et al, Nature Genetics 45, 1150 (2013)
Lango Allen et al, Nature 467, 832 (2010)
Speliotes et al, Nature 42, 937 (2010)

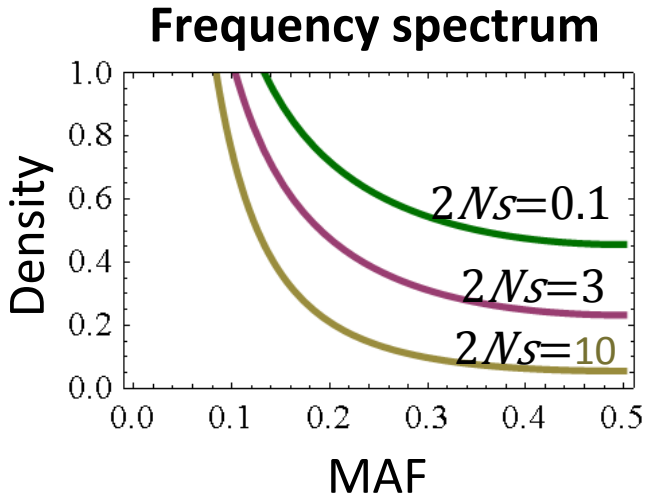
Genetic architecture: some questions

- How is the genetic architecture shaped by population genetic processes? A model.
- Can we use models to inform mapping study design?
- Can we use the results of mapping studies to learn about the forces that shape quantitative genetic variation? Inference.
- What about adaptation?



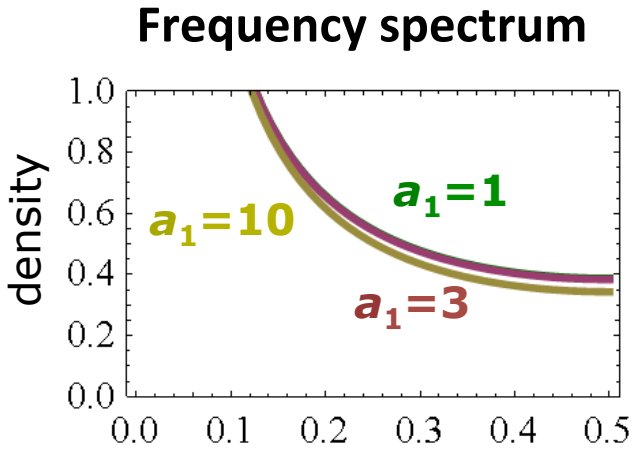
Lango Allen *et al.* , Nature **467**, 832 (2010)

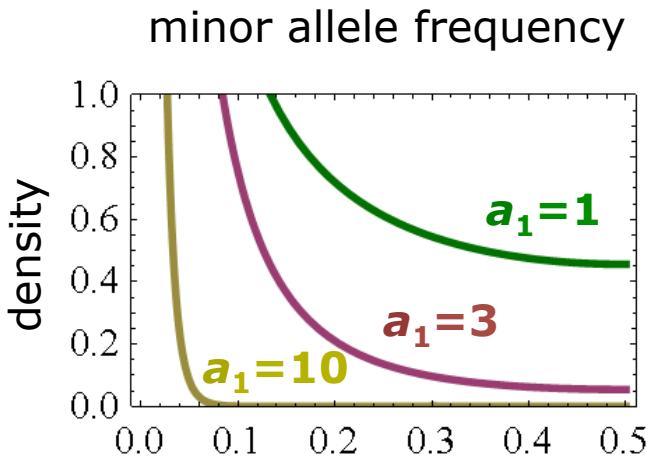
The distribution of selection effects and pleiotropy are key determinants of architecture:



s independent of a^2

s proportional to a^2





Selection & drift determine allele frequencies

Pleiotropy is key to relating selection and effect size

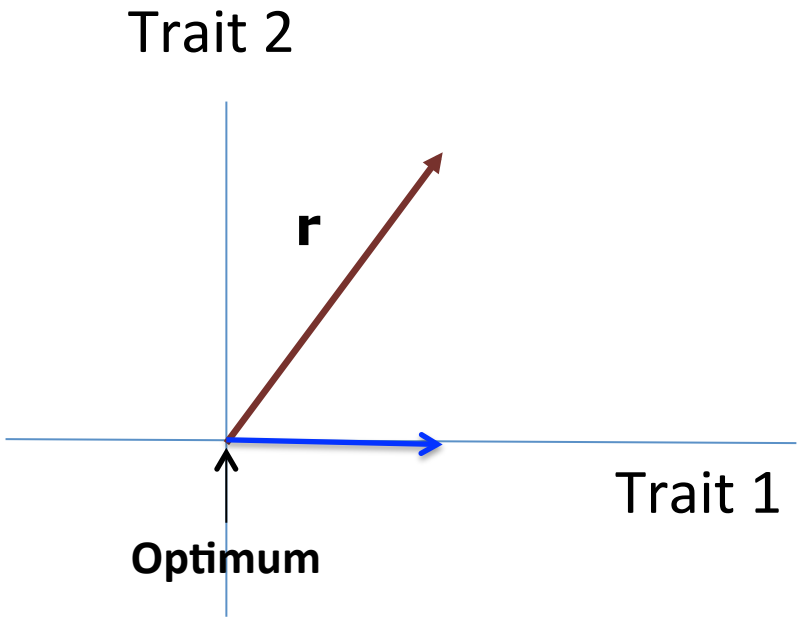
Some requirements from a model for genetic architecture

- Variable degree of pleiotropy.
- Ideally, this would arise from a description of a multi-dimensional phenotype.
- Selection on a variant should derive from the selection on phenotypes.
- We would like to be able to calculate measures of interest, e.g. the density of variance and variants as a function of frequency and effect size…

Fisher's geometric model: phenotype

- The phenotype is represented by a point in an n-dimensional Euclidian space. Each dimension corresponds to a trait.

The dynamic will be affected by all traits, but we will focus on the architecture of a single trait (1) and the dimension will parameterize the degree of pleiotropy.



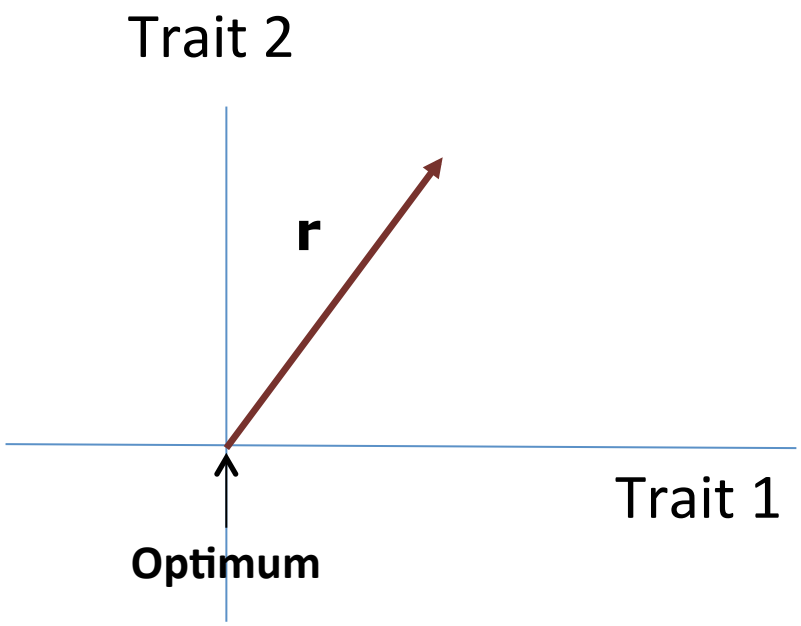
Fisher's geometric model: fitness

- The phenotype is represented by a point in an n-dimensional Euclidian space. Each dimension corresponds to a trait.
- Fitness is a function of the distance from the origin (stabilizing selection).

  We assume that absolute fitness is

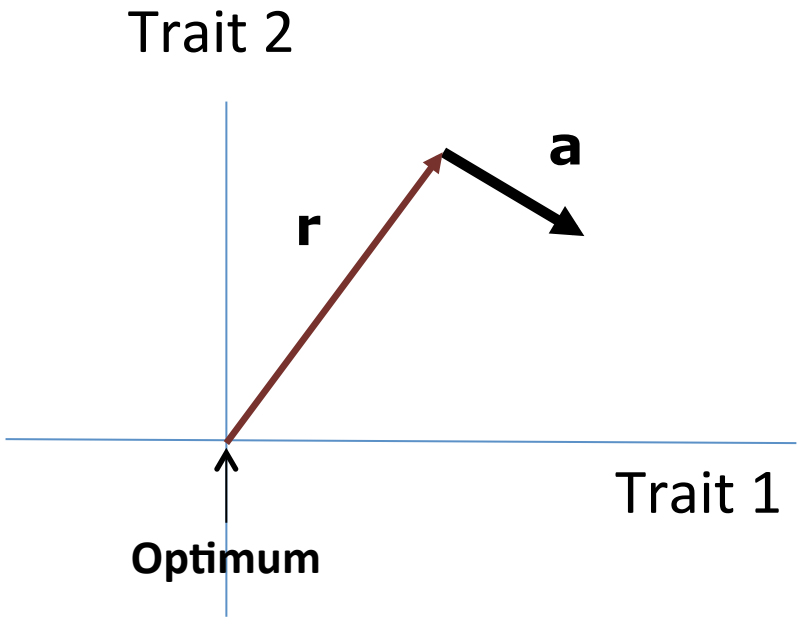
$$W(r) = Exp\left(-\frac{r^2}{2w^2}\right)$$

  where w determines the strength of selection and incorporates the environmental contribution.



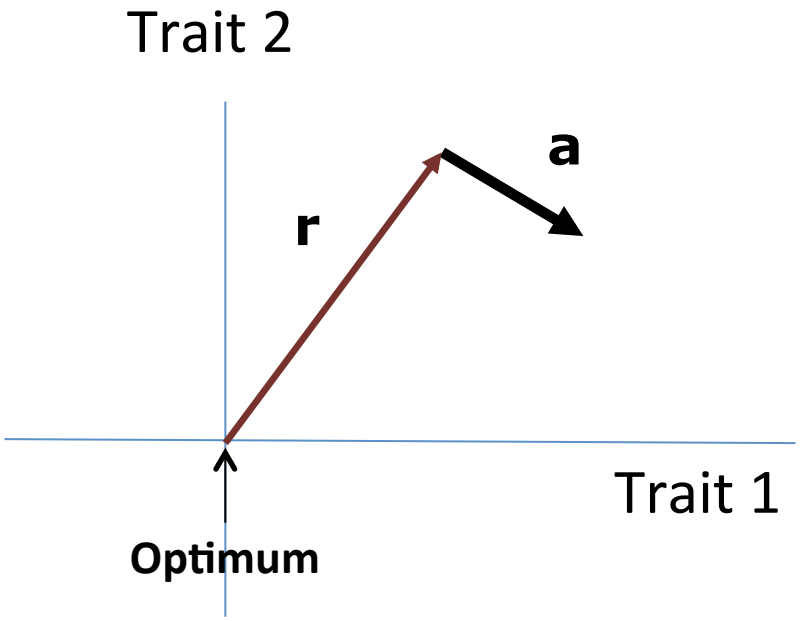
Fisher's geometric model: mutation

- The phenotype is represented by a point in an n-dimensional Euclidian space. Each dimension corresponds to a trait.
- Fitness is a function of the distance from the origin (stabilizing selection). We assume that absolute fitness is $W(r) = Exp\left(-\frac{r^2}{2w^2}\right)$.
- We assume an infinite-sites model, where mutations affecting the trait appear in the population at a rate of $2NU$.
- Mutations sizes a^2 are drawn from a gamma distribution, while their direction distribution is isotropic.



Fisher's geometric model: population dynamic

- The phenotype is represented by a point in an n-dimensional Euclidian space. Each dimension corresponds to a trait.
- Fitness is a function of the distance from the origin (stabilizing selection). We assume that absolute fitness is $W(r) = Exp\left(-\frac{r^2}{2w^2}\right)$.
- We assume an infinite-sites model, where mutations affecting the trait appear in the population at a rate of $2NU$.
- Mutations sizes a^2 are drawn from a gamma distribution, while their direction distribution is isotropic.
- Wright-Fisher sampling with viability selection, mutation, free-recombination, and Mendelian segregation.



Fisher's geometric model: genotype to phenotype

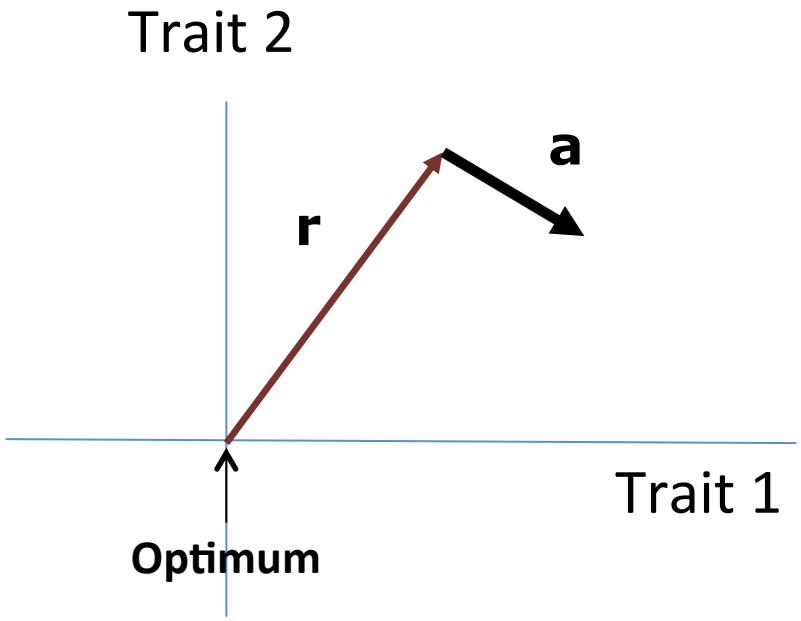
The genetic component of the trait for can be separated into the segregating and fixed and component $r = r_f + r_s$.

Segregating: If there are *L* segregating sites then the phenotypic contribution for individual *i* is

$$\boldsymbol{r}_s^i = \frac{1}{2}\sum_{l=1}^{L} g_{i,l}\boldsymbol{a}_l,$$

where $g_{i,l} = 0{,}1\ or\ 2$ is the number of mutant alleles carried by individual *i* at site *l* and $\boldsymbol{a}_l$ is the phenotypic effect of the mutation at locus *l* (specifically, a homozygote).

Fixed: We do bookkeeping.



Some requirements from a model for genetic architecture

- Variable degree of pleiotropy.
- Ideally, this would arise from a description of a multi-dimensional phenotype.
- Selection on a variant should derive from the selection on phenotypes.
- We would like to be able to calculate measures of interest, e.g. the density of variance and variants as a function of frequency and effect size…

The steady state distribution of phenotypes in the population is simple

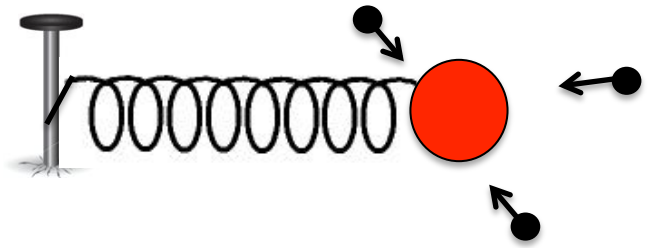
Under sensible conditions ($1 \gg U \gg 1/2N$), it is normal around the origin:

$$f(r) = \frac{1}{(2\pi\sigma^2)^{\frac{n}{2}}} exp\left(-\frac{r^2}{2\sigma^2}\right),$$

where $\sigma^2 \ll w^2$.

Intuition:

- Normal: ***r*** is a sum over i.i.d. contributions over many loci
- Mean at optimum ($V(\|\bar{r}\|) \approx \frac{nw^2}{2N}$):



(An Ornstein–Uhlenbeck process)

Lande (1976)

The dynamic of a segregating variant follows (the first two moments)

Consider a mutations with effect size ***a*** and frequency *q*

The distribution of its phenotypic background is

$$f(\vec{X}|\vec{\alpha},q)=\frac{1}{(2\pi\sigma^2)^{n/2}}Exp\left(-\frac{\left(\vec{X}+q\vec{\alpha}\right)^2}{2\sigma^2}\right).$$

The expected fitness of the three genotypes is therefore

$$W_{00}=\int_{\vec{X}} f(\vec{X}|\vec{\alpha},q)W(\vec{X})\cdots$$

It follows that

$$E(\Delta q)=-pq\frac{p(W_{00}-W_{01})+q(W_{01}-W_{11})}{\bar{W}}\approx-\frac{a^2}{4w^2}pq\left(q-\frac{1}{2}\right)$$

and

$$V(\Delta q)\approx\frac{pq}{2N}$$

Now we can use the diffusion to calculate summaries of interest

Two notes on the first moment

$$E(\Delta q) = -\frac{a^2}{4w^2} pq\left(q - \frac{1}{2}\right)$$

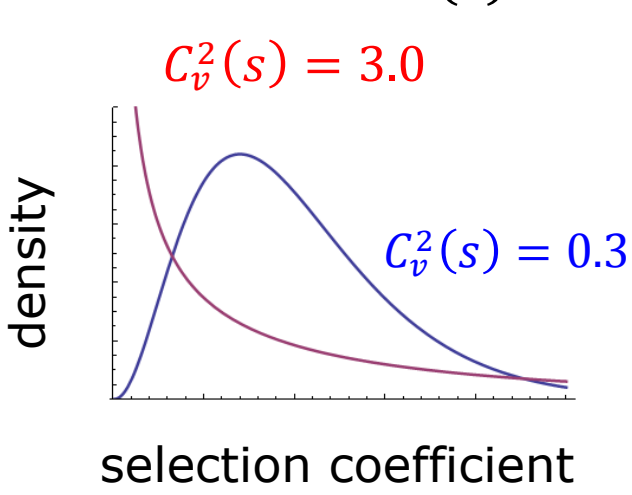
This is the standard under-dominant form (Robertson 1961).

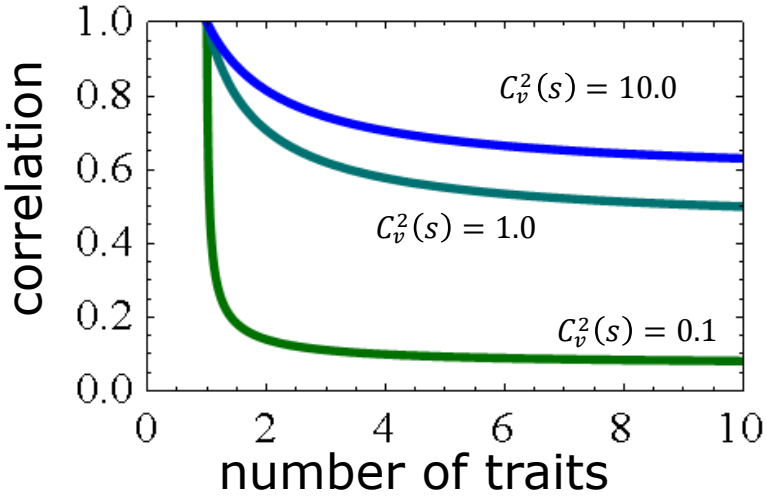
- The reason isn't heterozygote disadvantage but rather selection to reduce the phenotypic variance ½ a^2pq.
- The selection coefficient s is $a^2/2w^2$. This is why Na^2/w^2 is a natural unit for mutation size.

The relationship between selection coefficients and effect sizes

$$Corr(s, a_1^2) = \sqrt{\frac{1}{1 + 2\frac{n-1}{n+2}(1 + 1/C_v^2(s))}}$$

$$C_v^2(s) \equiv \frac{V(s)}{E^2(s)}$$





The effects of pleiotropy and selection on genetic architecture (2Ns=3)

peaked

Coefficient of variation

wide

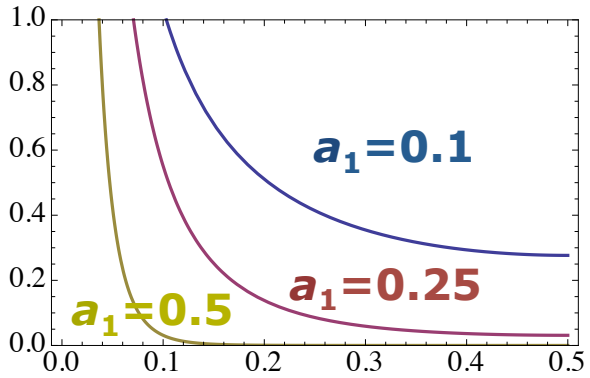
$C_v^2(s)=1/3$

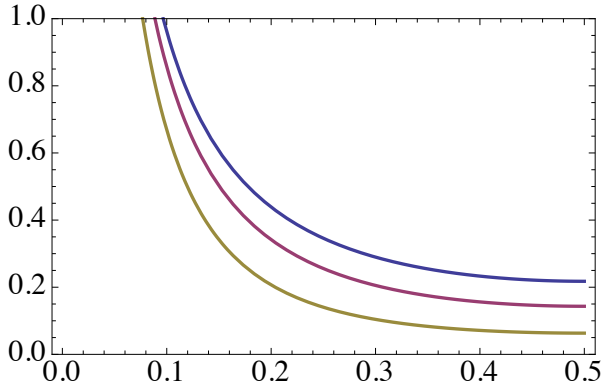
$C_v^2(s)=3$

density

n=1
(corr(s,a_1^2)=1)

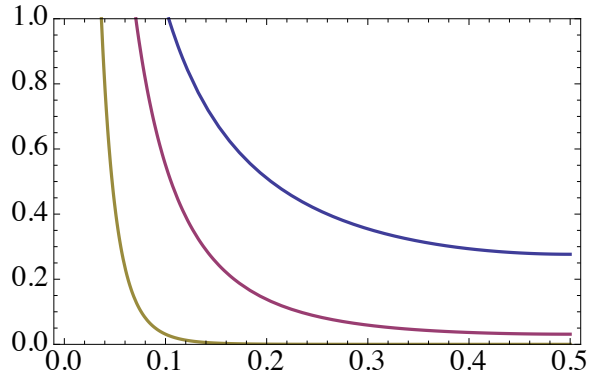
n=100
(corr(s,a_1^2)=0.34)

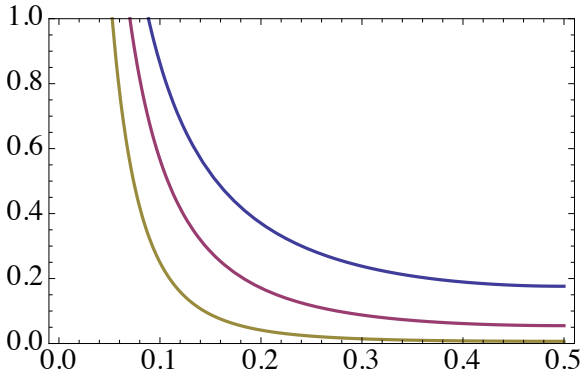




minor allele frequency

minor allele frequency





(corr(s,a_1^2)=1)

(corr(s,a_1^2)=0.53)

pleiotropy

Simulated GWAS:
Study size 10K & explained variance is 20%

peaked

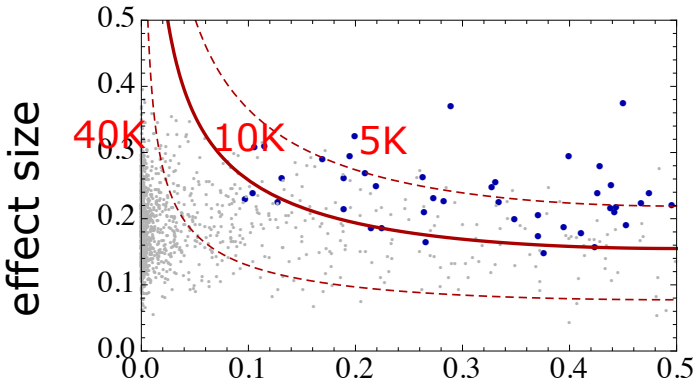
Coefficient of variation

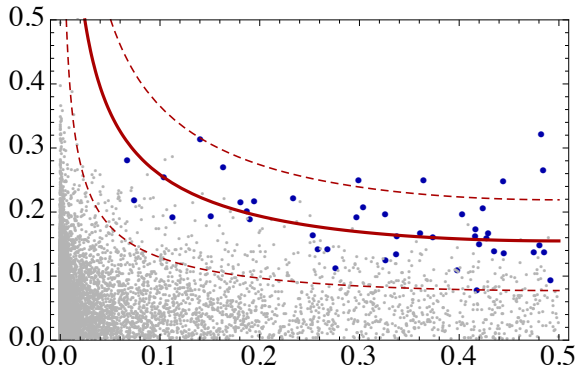
wide

n=1
(corr(s,a_1^2)=1)

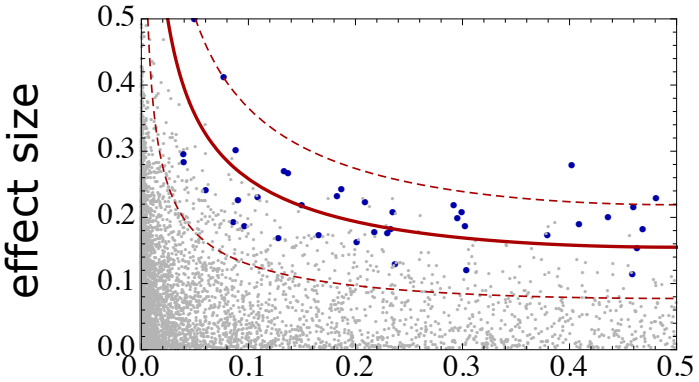
n=100
(corr(s,a_1^2)=0.34)

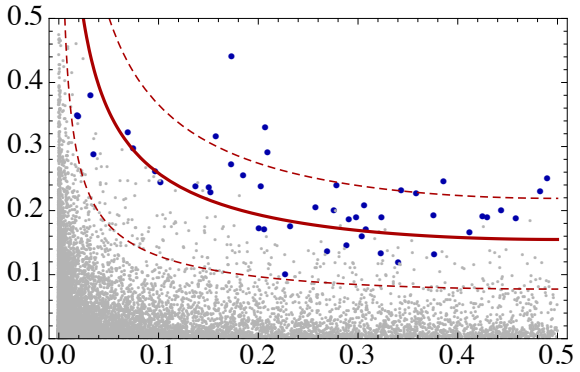
$C_v^2(s)$=1/3





$C_v^2(s)$=3



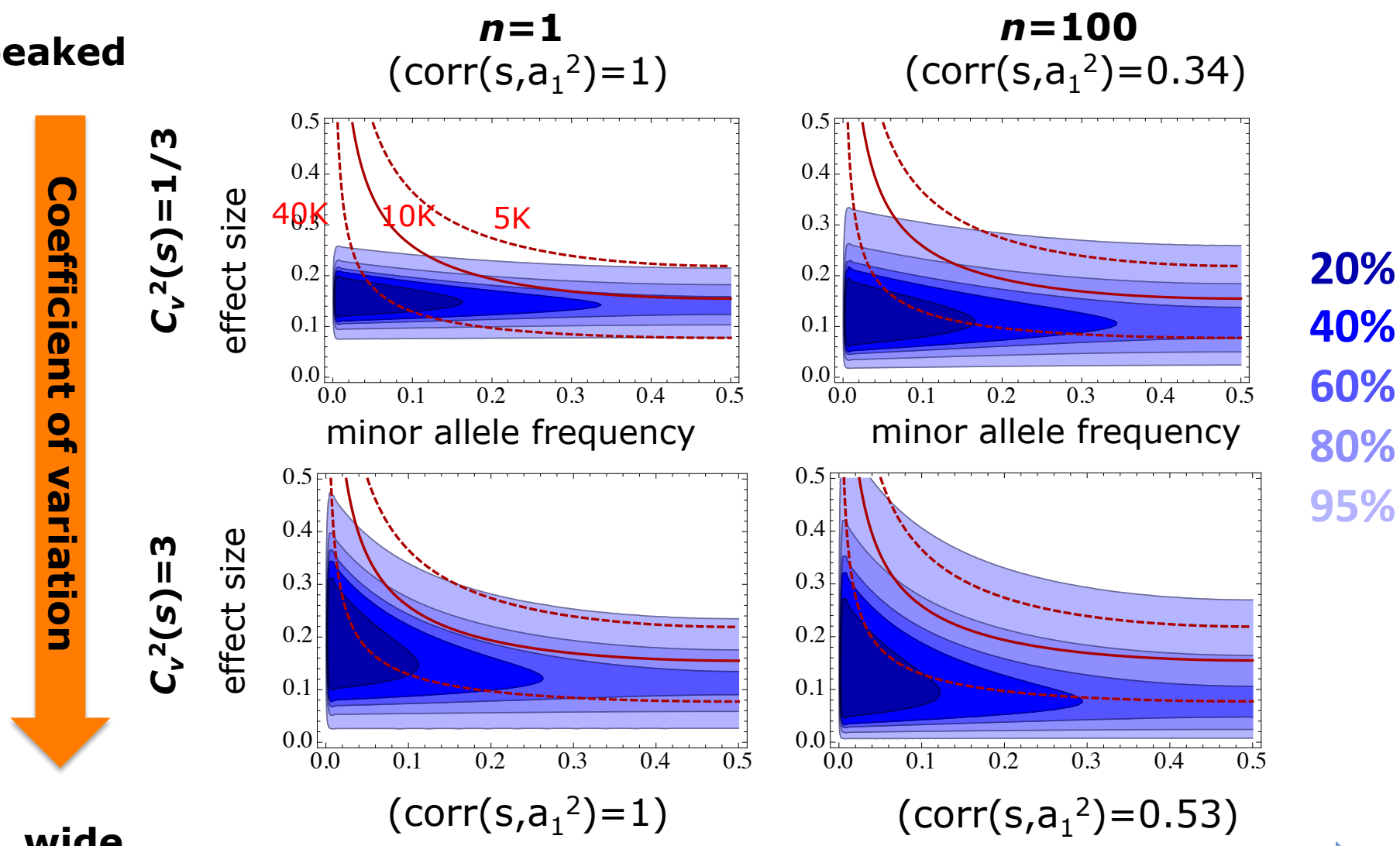


(corr(s,a_1^2)=1)

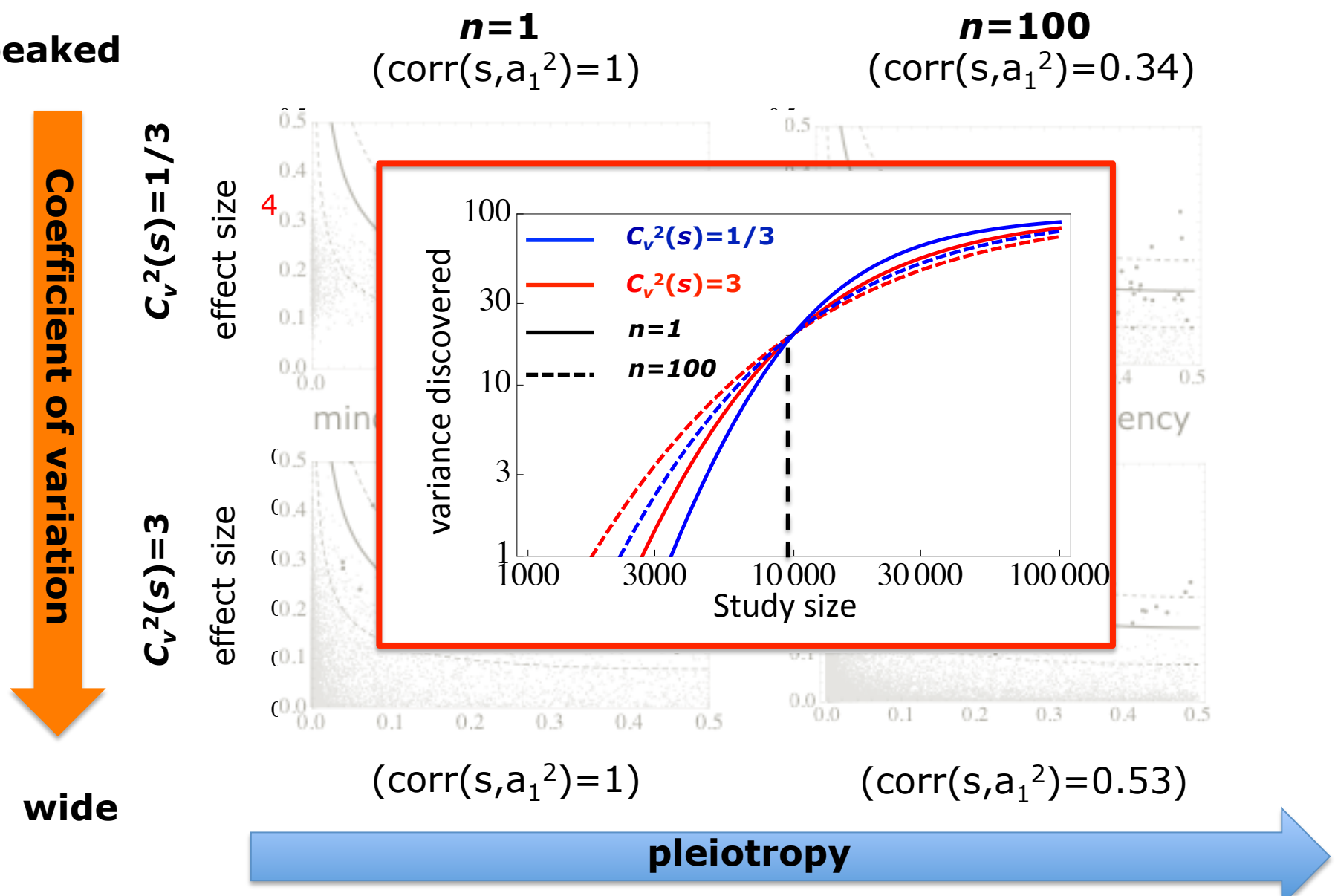
(corr(s,a_1^2)=0.53)

pleiotropy

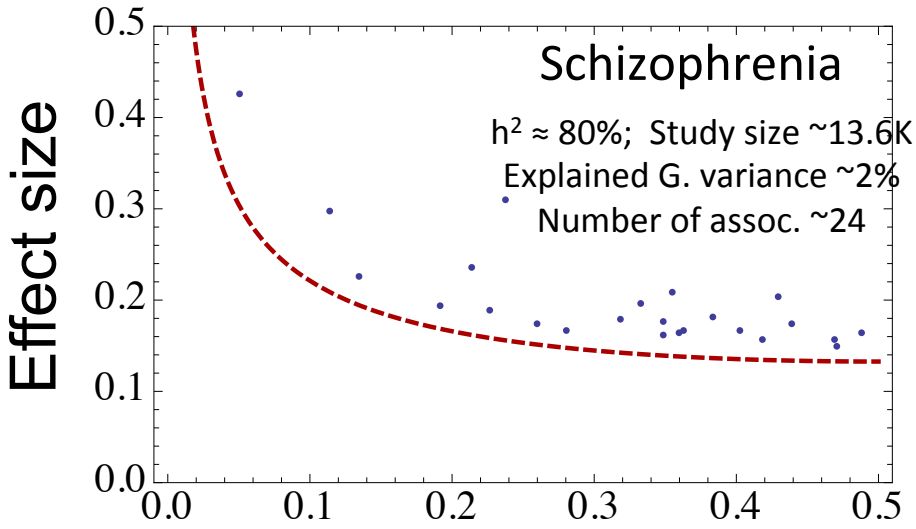
The effects of pleiotropy and selection on genetic architecture (2Ns=3): density of variance

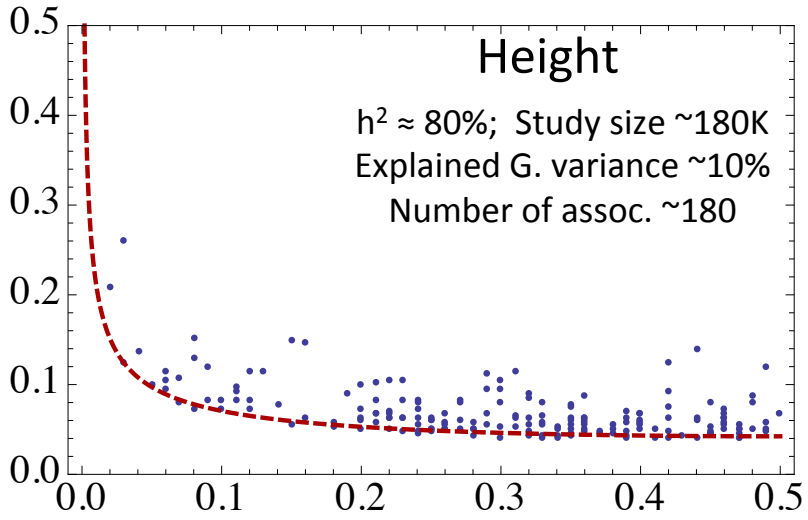


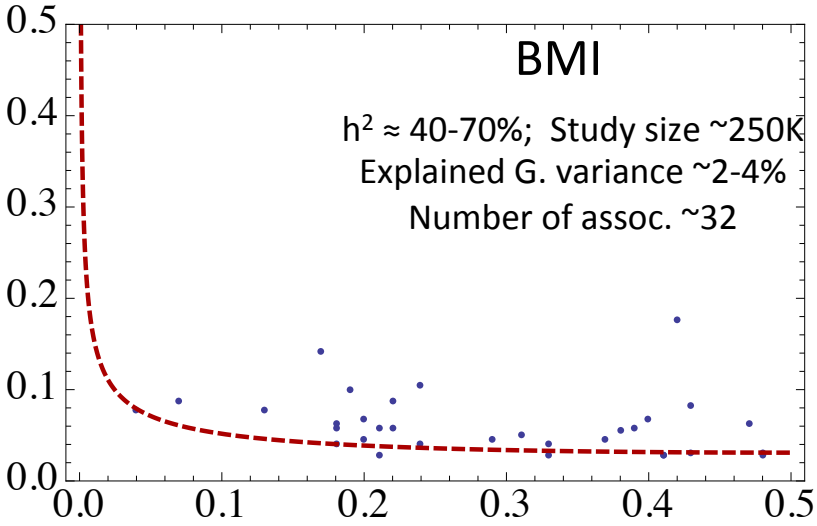
The effects of pleiotropy and selection on genetic architecture (2Ns=3): density of variance



Genetic architecture







While preliminary, we already see differences in:

- Explained variance
- Number discovered

And

- The distribution

Ripke et al, Nature Genetics 45, 1150 (2013)
Lango Allen et al, Nature 467, 832 (2010)
Speliotes et al, Nature 42, 937 (2010)

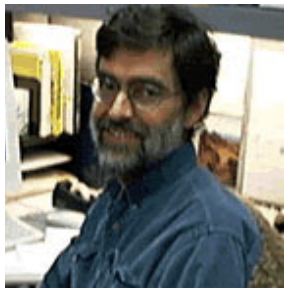
Summary and future directions

- We solved an intuitive model for how pleiotropy and the distribution of selection coefficients affect architecture.
- Other models for architecture (solution, robustness, apparent and balancing selection).
- Inference (implication to mapping and to understanding genetic variation).
- Polygenic adaptation.

Acknowledgements

Yuval Simons
(Columbia U.)

Kevin Bullaughey
(U. Chicago)

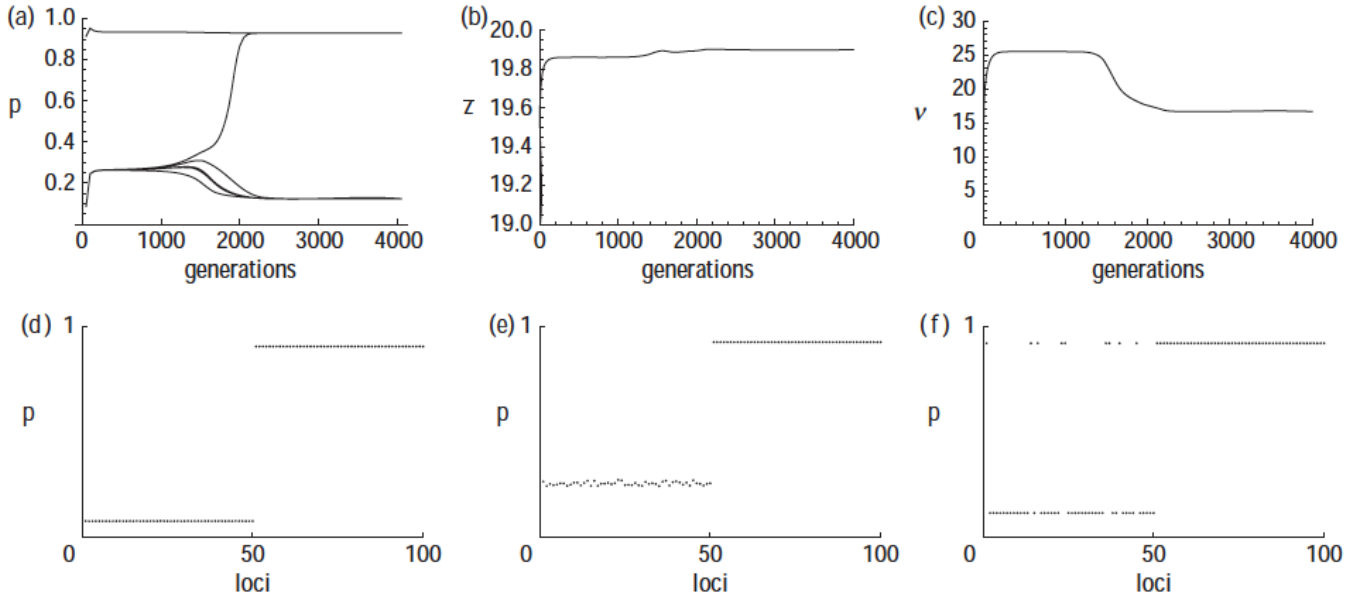


Dick Hudson
(Chicago U.)

Interested in joining the lab?

Biological Sciences, Columbia

Polygenic adaptation



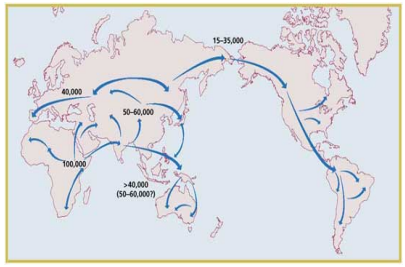


Figure 1. (*a*) Allele frequencies, (*b*) trait mean and (*c*) genetic variance plotted against time. A population is initially at equilibrium with stabilizing selection $s = 0.05$ towards $z_{opt} = 0$ acting on an additive trait, with $n = 100$ loci of effect $\gamma = 1$; the mutation rate is $\mu = 0.002$ per locus, which maintains a genetic variance of $v = 4n\mu\gamma s = 16$. The optimum then shifts abruptly to $z_{opt} = 20$, and the mean responds almost immediately (*b*). The variance increases abruptly (*c*) as, the allele frequencies at all the '−' loci increase substantially (*d*). However, this new state is unstable, and slight variations in the initial conditions cause some loci to shift down, and some to shift up. As a result, the genetic variance returns to its original value. The lower row shows snapshots of allele frequencies at times (*d*) 0, (*e*) 800 and (*f*) 3000 generations.

From de Vladar & Barton (2010)