Genetic architecture of intelligence from SNP distance measures

Steve Hsu

University of Oregon and BGI
www.cog-genomics.org

Collaborators

ALSPAC / University of Bristol: George McMahon, George Davey Smith

TEDS / King's College London: Ken Hanscombe, Robert Plomin

NIDDK / NIH: James Lee, Shashaank Vattikuti, Carson Chow (Height data from ARIC)

Cognitive Genomics Lab, BGI: Christopher Chang, Laurent Tellier, Rui Yang, Bowen Zhao

Genetic distance measures

Quantitative traits: many alleles, each of small effect. GWAS discovery of individual loci is hard.

But, phenotype differences must be associated with LARGE number of genetic differences.

Investigate pairwise genetic distance as g score (or height) are varied. Extract underlying genetic architecture:

1. Distribution of associated alleles dominated by small MAF (Minor Allele Frequency)
2. More (−) than (+) minor alleles (MAF < 0.5)
3. Rough estimate of 10k causal alleles in total

Data sources and Results

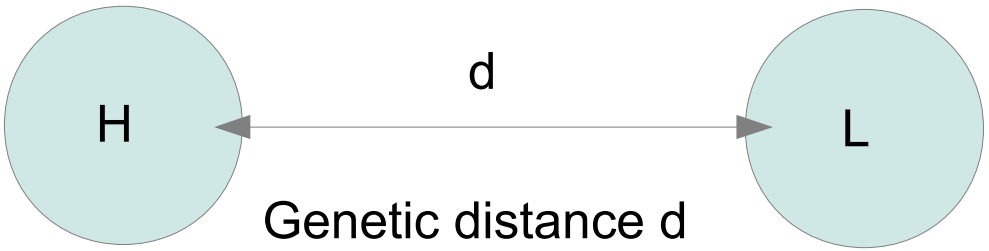
ALSPAC: 4000 individuals, age 15 IQ; 2000 individuals, age 8 IQ
TEDS: 2400 individuals, age 12 IQ
ARIC: 5700 adult heights

ALSPAC: 488k SNPs on chip. Average pairwise distance = 261k ± 1.5k SNPs.

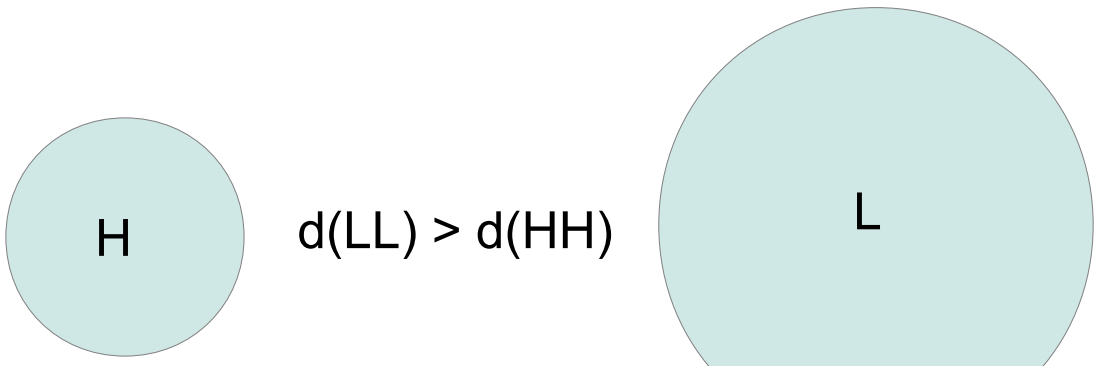
Select outlier groups H and L. Averaging over pairs eliminates fluctuations in distance which are uncorrelated to phenotype.

Average pairwise genetic distance changes with mean IQ and IQ difference: ~ 39 SNPs per population SD

Results



Genetic distance d
varies with avg
phenotype difference



Average intragroup distance
varies with group mean

Genetic distance: architecture from geometry

These two genotypes have a relative Hamming distance of 2:

$\{+ + + + \ominus + + + + + + + + +\}$ vs $\{+ + + + + + + + + + + \ominus + +\}$

These two genotypes have a relative Hamming distance of 6:

$\{+ + \ominus + + + + + + \ominus + + \ominus +\}$ vs $\{+ + + + + \ominus \ominus + + + \ominus + +\}$

More $\ominus$ alleles means greater Hamming distance.

Note we've made the assumption that (+) is common (MAF > 0.5) and $\ominus$ is uncommon (MAF < 0.5). Otherwise, more (+) alleles would mean greater Hamming distance.

Genetic distance: architecture from geometry

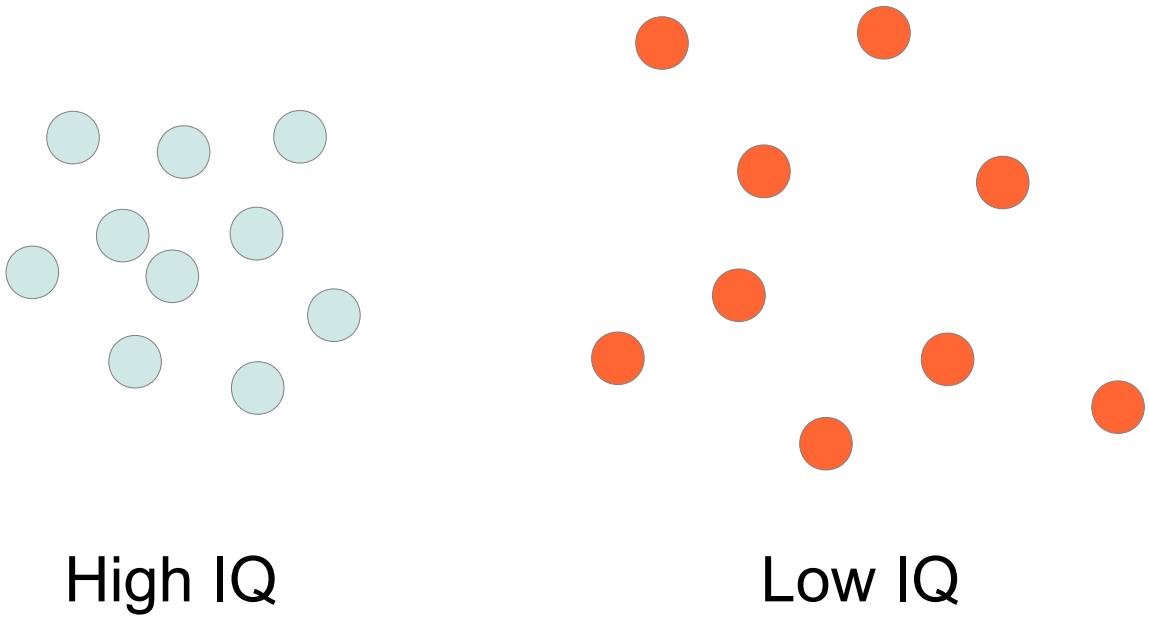
Real genomes are diploid.

Simplest distance measure, analogous to Hamming distance:

AA AA	0
AA Aa	1
AA aa	2

Can also weight by factors of MAF or standardize to obtain different distance measures (e.g., relatedness).

Genetic distance: architecture from geometry



Low IQ = more rare (−) variants. Larger genetic distances between individuals. Similar results for height.

Additive model

n_+ minor alleles with (+) effect on intelligence (MAF < 0.5).

$n_\ominus$ minor alleles with ($-$) effect on intelligence (MAF < 0.5).

Result $d(LL) > d(HH)$ implies that

$$n_\ominus > n_+$$

Plausible that

$$n_\ominus \gg n_+$$

Simplified additive model: spherical cow

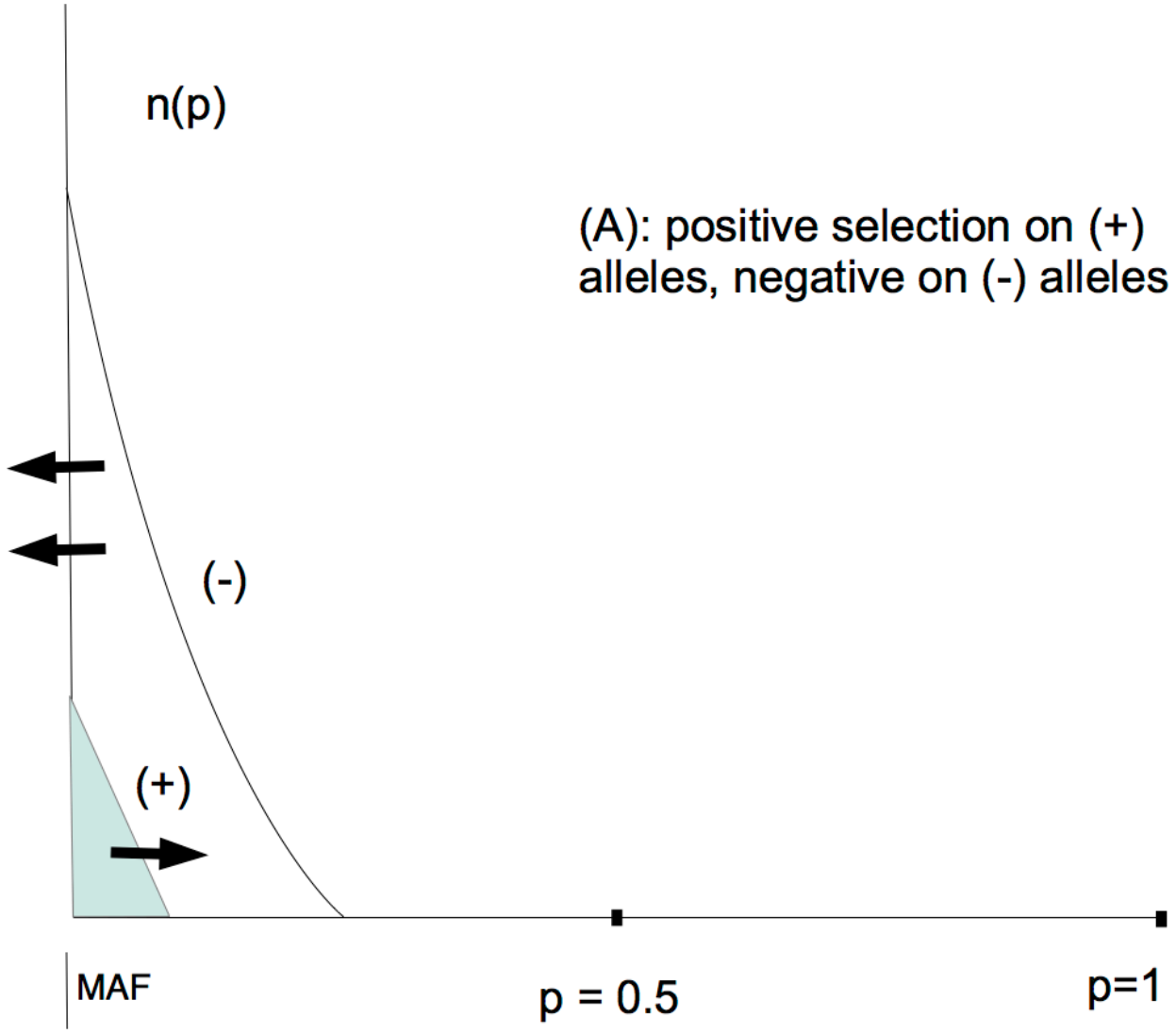
(1) N causal variants, ALL minor alleles have (−) effect on IQ ($n_+ = 0$; $n_\ominus = N$)

(2) Typical $MAF < 0.1$

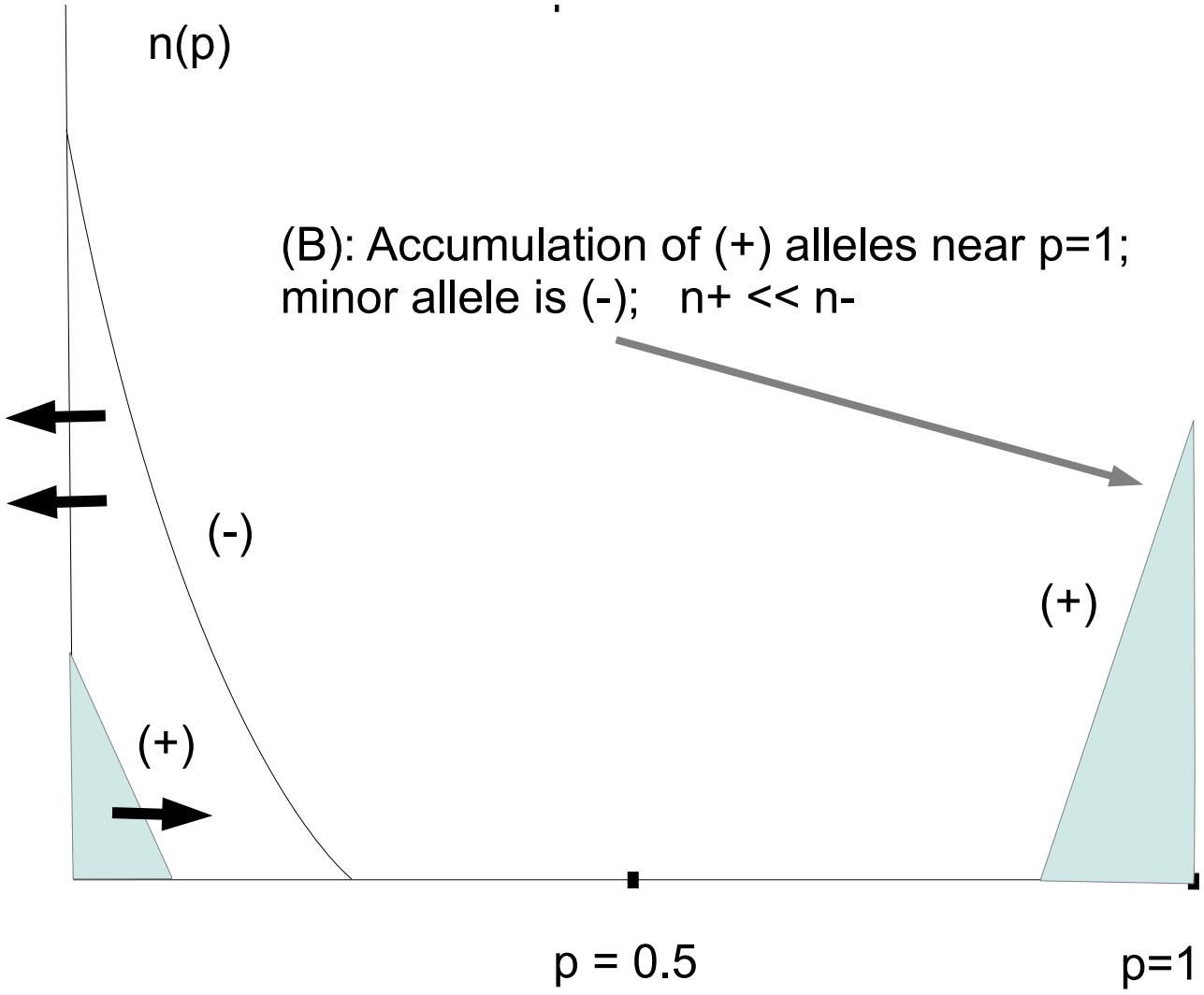
(3) Binomial distribution: 1 SD $\sim (0.1\,N)^{1/2}$

For $N \sim$ 10k, get 1 SD change in intelligence per 30 extra (−) variants.

Selection and MAF distribution

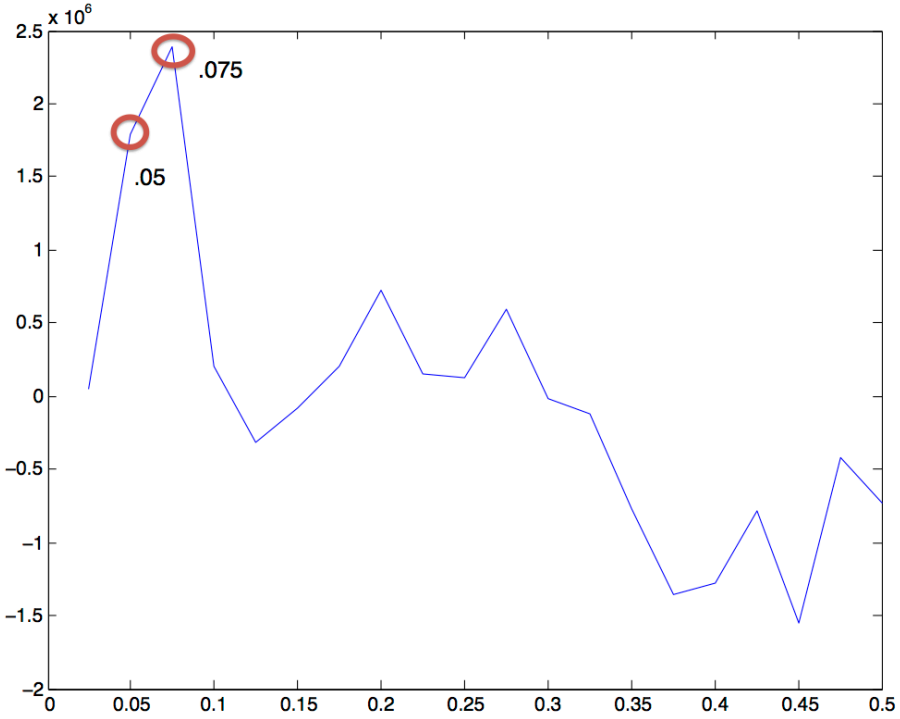


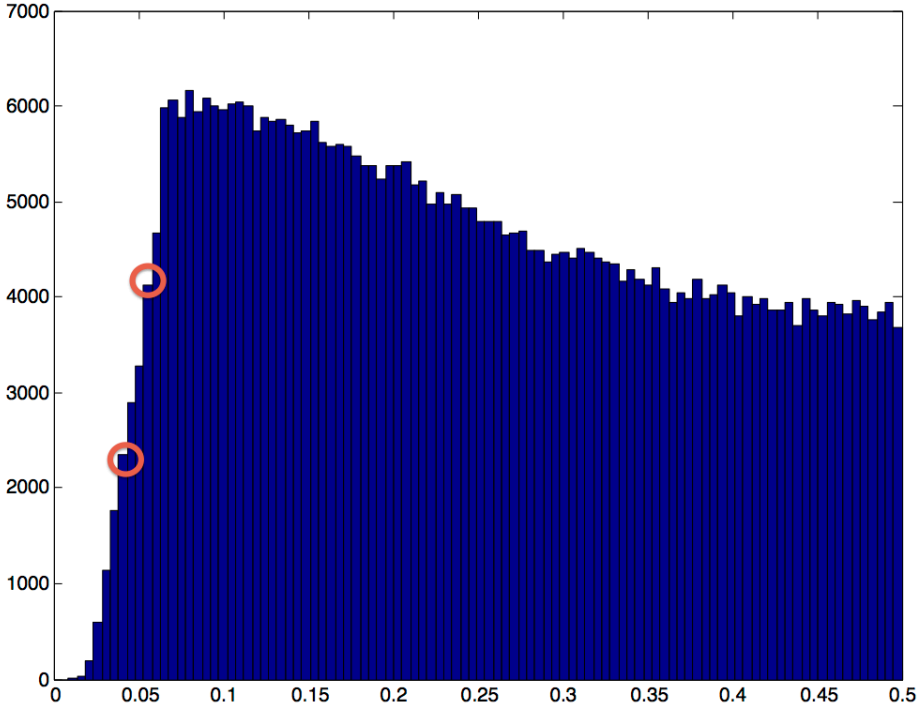
Selection and MAF distribution



MAF distributions

Distribution of associated alleles dominated by MAF < 0.1.





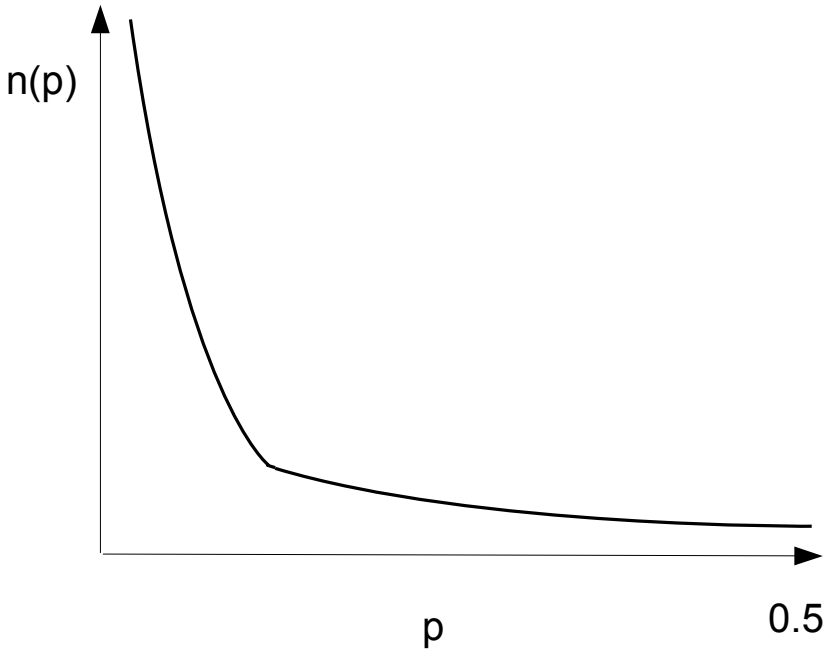
Left: contributions to H–L genetic distance by MAF. Right: density of SNPs on chip.

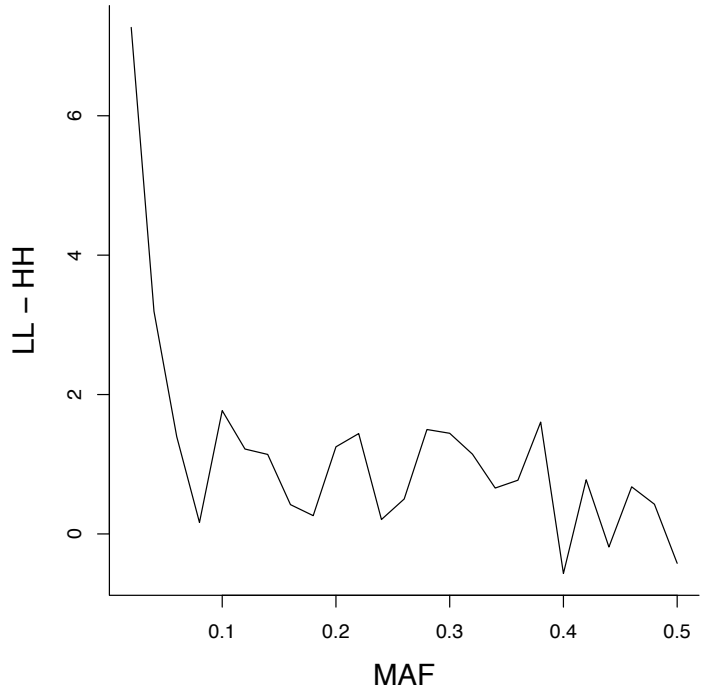
MAF distributions

Modulo statistical errors, can extract

$$n(p) = \text{density of associated SNPs}$$

Result consistent with "L shape" suggested by population genetics models.





Implications of low MAF: missing heritability

GCTA: heritability on chip is roughly $h^2 \sim 0.5$. (Specifically, 0.56 for ALSPAC.)

But, expect larger total additive heritability, perhaps even $h^2 \approx 0.8$!

Yang et al. 2010: causal variants at low MAF are poorly tagged by chip; if MAF of many causal variants < 0.1, can recover "missing heritability".

Implications of low MAF: epistasis, additivity and all that

Why is most of the variance additive? Where is the epistasis that our wet lab colleagues see every day?

If most causal variants are rare (e.g., MAF < 0.1), then when two individuals differ at a locus we likely find AA vs Aa. Very few individuals are aa.

Therefore, even if the effect of aa is not twice that of Aa (non-additivity or non-linearity), the relative size of population level non-additive effects is still small – suppressed relative to additive effects by of order MAF.

(Similar argument for gene-gene interactions, etc.)

Geniuses and Giants: Fewer deleterious alleles.

(A) 39 SNPs per SD of IQ suggests roughly 10k causal variants.

(B) Exceptional cognitive ability = of order 100's fewer rare (−) variants than an average person.

Many caveats to estimate (A); uncertainty in (B) is smaller due to SD $\sim \sqrt{N}$.

Toy model: 10k causal variants, typical MAF = 0.1 : average person has ~ 1000 randomly distributed (−) variants; little overlap between individuals in locations of (−)'s. A genius or giant has ~ 100 fewer (−) alleles: ~ 900 (−) variants in total.