Common variants and their contribution to heritability ("GWAS and heritability")

peter.visscher@uq.edu.au

Program in Complex Trait Genomics

The University Of Queensland AUSTRALIA | IMB
Institute for Molecular Bioscience

The University Of Queensland AUSTRALIA | QBI
Queensland Brain Institute

The original definition of 'missing heritability'

$$Missing\ h^2 = \hat{h}^2_{pedigree} - \hat{h}^2_{GWS\ loci}$$

NB both are estimates that can be biased (up or down)

My 2009 presentation

- Theory and applications of quantitative genetics: heritability, estimation and prediction
- Estimation of heritability using DNA markers:
  - Using segregation within families
  - Using GWAS data on "unrelated" individuals (unpublished data that became **Yang et al. 2010 NG**)

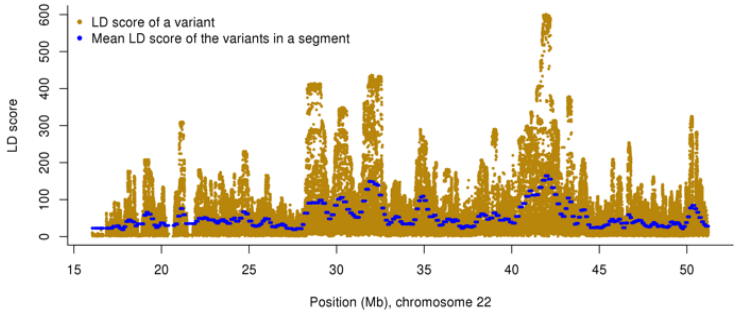
Yang et al. 2010 NG: SNP-heritability

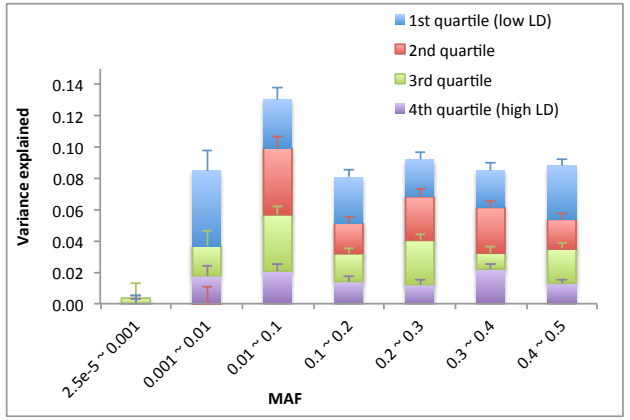
- Estimation, not hypothesis testing
- Variance explained by all genotyped SNPs ~ 45% for height
- Contrast 45% with 5% from GWS SNPs (Manolio 2009)
- Larger GWAS sample size → discovery of more GWS loci
- 'Infinite' sample size → 45% of variance explained by GWS SNPs; prediction R^2 → 45%

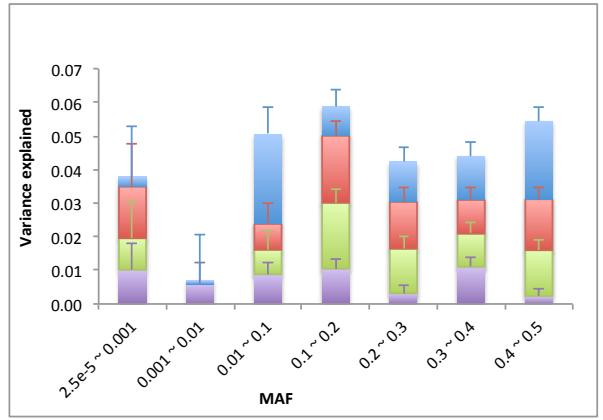
Common SNPs explain a large proportion of the heritability for human height

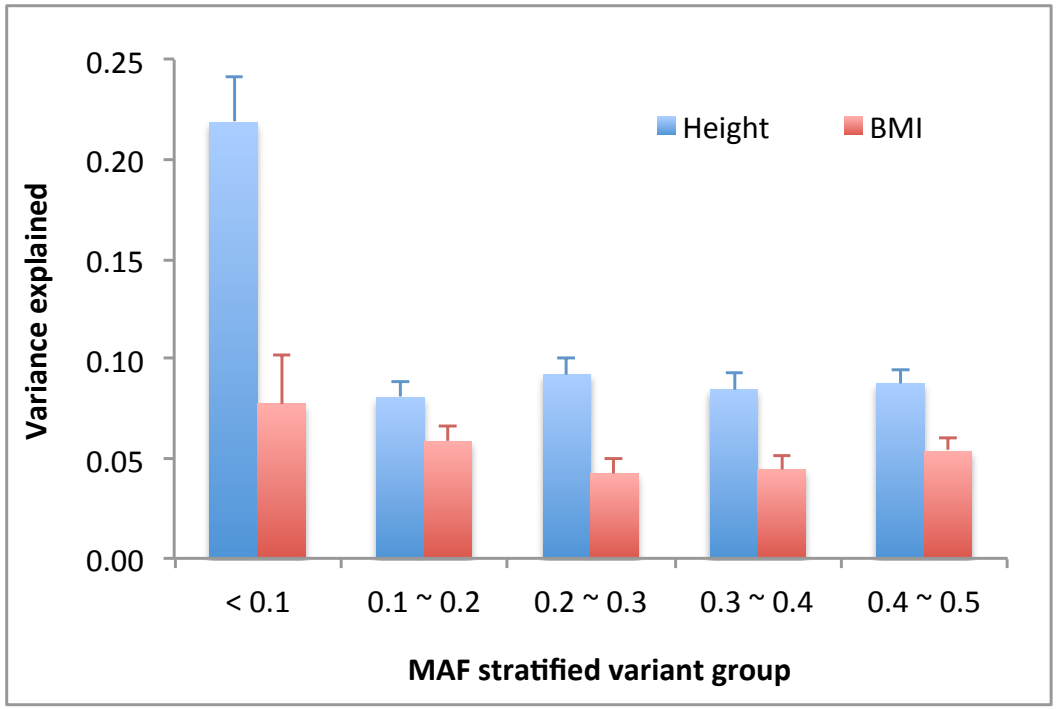
Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

Robust estimation from imputed variants by accounting for LD and MAF









Totals
57% for height
27% for BMI

Yang et al. 2015 (Nature Genetics)

Re-reading Manolio et al. 2009

"Many explanations for this missing heritability have been suggested, including

- much larger numbers of variants of smaller effect yet to be found;
- rarer variants (possibly with larger effects) that are poorly detected by available genotyping arrays that focus on variants present in 5% or more of the population;
- structural variants poorly captured by existing arrays;
- low power to detect gene–gene interactions;
- and inadequate accounting for shared environment among relatives."

“much larger numbers of variants of smaller effect yet to be found”

- Cumulatively, common variants explain ~1/3 to ~2/3 of heritability (GREML and LD Score regression methods)
- Much larger numbers of variants have indeed been found e.g.
  - from 40 to 3000+ for height
  - 8 to 700+ for BMI
  - 0 to 1000+ for educational attainment / IQ
  - 1 to 250 for schizophrenia
  - 32 to 200 for inflammatory bowel disease
  - 18 to 150 for Type 2 diabetes

“rarer variants (possibly with larger effects)“

- Evidence for natural selection: rare(r) variants associated with complex traits have larger effects
  - height
  - BMI
  - disease

- But cumulatively, rare variants contribute a small amount of heritability
  - T2D
  - Height, BMI

New definitions of ‘heritability’ since 2009…

- Missing
- Phantom
- Pedigree
- SNP
- Hiding
- Genomic
- etc.

New data since 2009

- GWAS summary statistics
- More and ever-larger GWAS
- Transcriptional and epigenetic resources
- Fully sequenced reference panels
  - imputation accuracy down to MAF = 0.5%

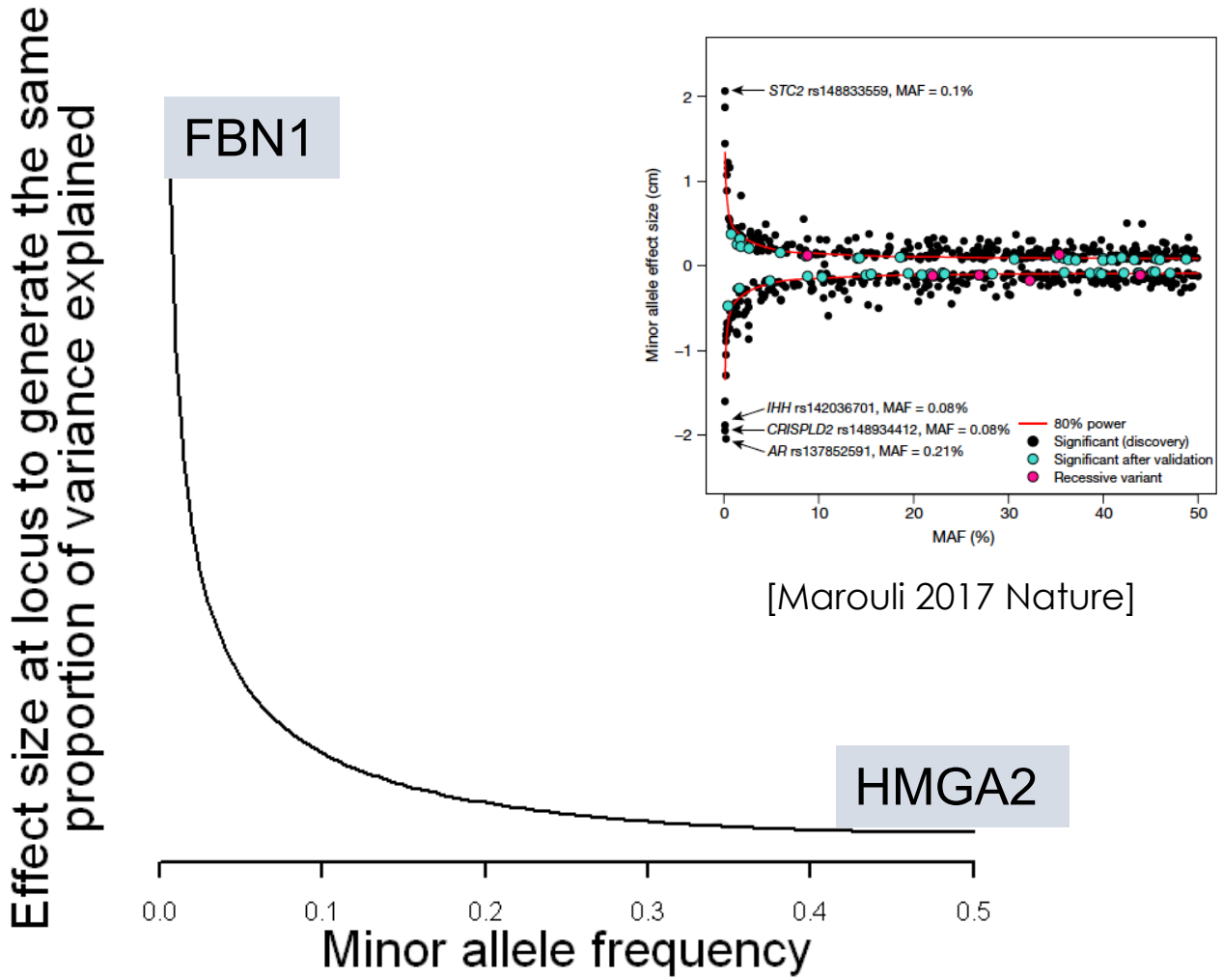
- Large single cohort studies, e.g. UK Biobank
- Contributions from commercial companies e.g. 23andMe

New methods since 2009

- GREML (Yang 2010, 2015 NG; 2011 AJHG)
- LD score regression (Bulik-Sullivan 2015 NG 2x)
- Prediction methods (Purcell 2009 Nature; Zhou-Stephens 2012 PLOS Genetics, 2013 NG; Moser 2015 PLOS Genetics; Turley 2017 NG; Maier 2018 Nat Comms)
- Causal inference (MR, SMR, GSMR, PrediXcan, MetaXcan)

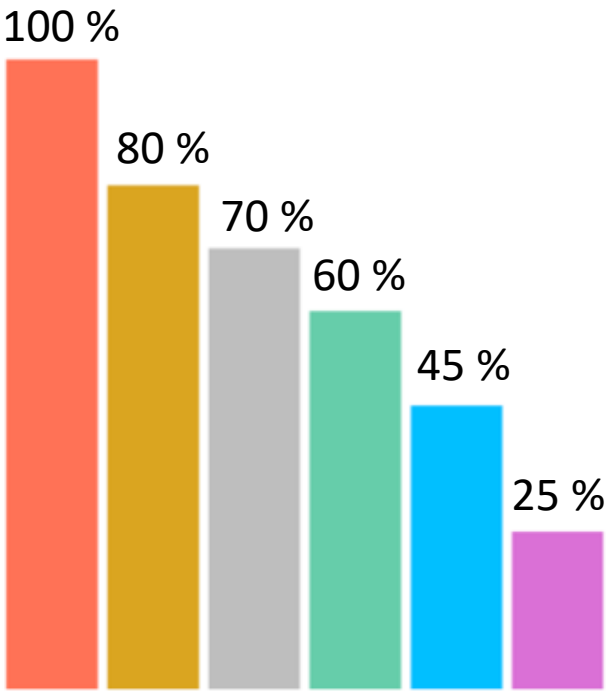
Mendelian forms of “tallness” and “shortness” exist, but most variation is polygenic

The combination of allele frequency and effect size determines the contribution to heritability



[Marouli 2017 Nature]

Partitioning variance of height 2018



Total variance
Heritability (based on Twin or family studies)
Within-family estimates
SNP heritability from imputation to sequenced reference
SNP-heritability (variance explained by all genotyped SNPs on the Chip)
Variance explained by genome wide significant SNPs

Prediction R^2 is approaching 40%

Variance explained by WGS unknown

Slide by Loic Yengo

Variance explained for BMI

Twin studies	70-80%
Non-twin family studies	40-50%
Within-family segregation	40%
Whole-genome imputation	27%
HapMap3 SNPs	22%
GWS loci	5%

Difference between within-family and population estimates of SNP effects

- Population stratification
- G-E correlation (Nature of Nurture)
- Assortative mating

- Ratio within to population estimates
  - Height ~0.9
  - Educational attainment ~0.5

[Lee Nature Genetics 2018 in press; Kong Science 2018]

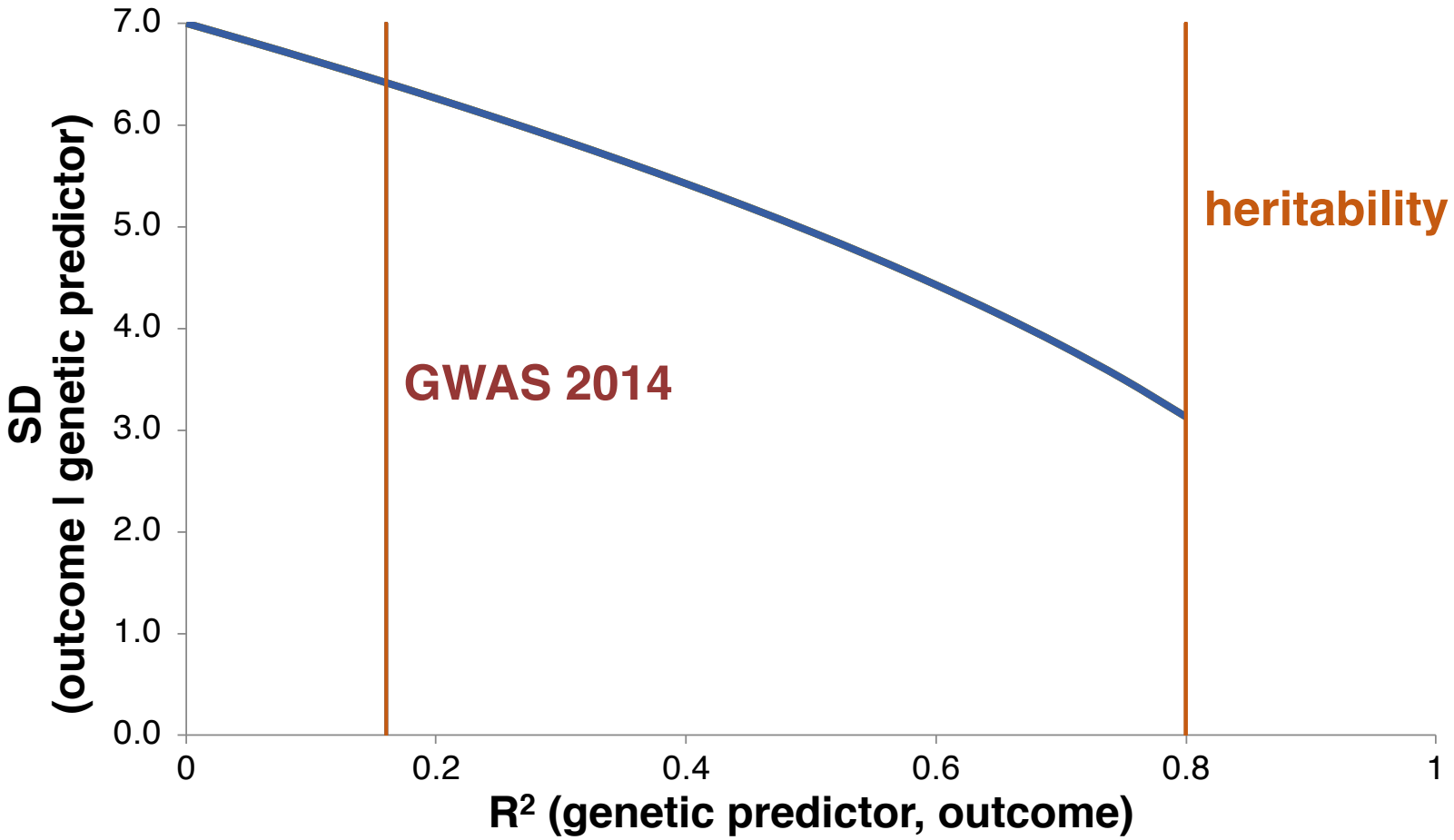
Non-additive genetic variance from GWAS data

- Few examples from GWS loci
  - but loci detected from additive models
- Greater loss of information due to imperfect LD
  - r^4 vs r^2
- Estimation of dominance variance
  - 3% from 79 traits on N = 6700 (Zhu 2015 AJHG)
  - <1% from 20 traits on N = 350,000 (Rohart 2018 unpublished)
- Lack of power to detect AxA variance
- Confounding with non-genetic effects from family data

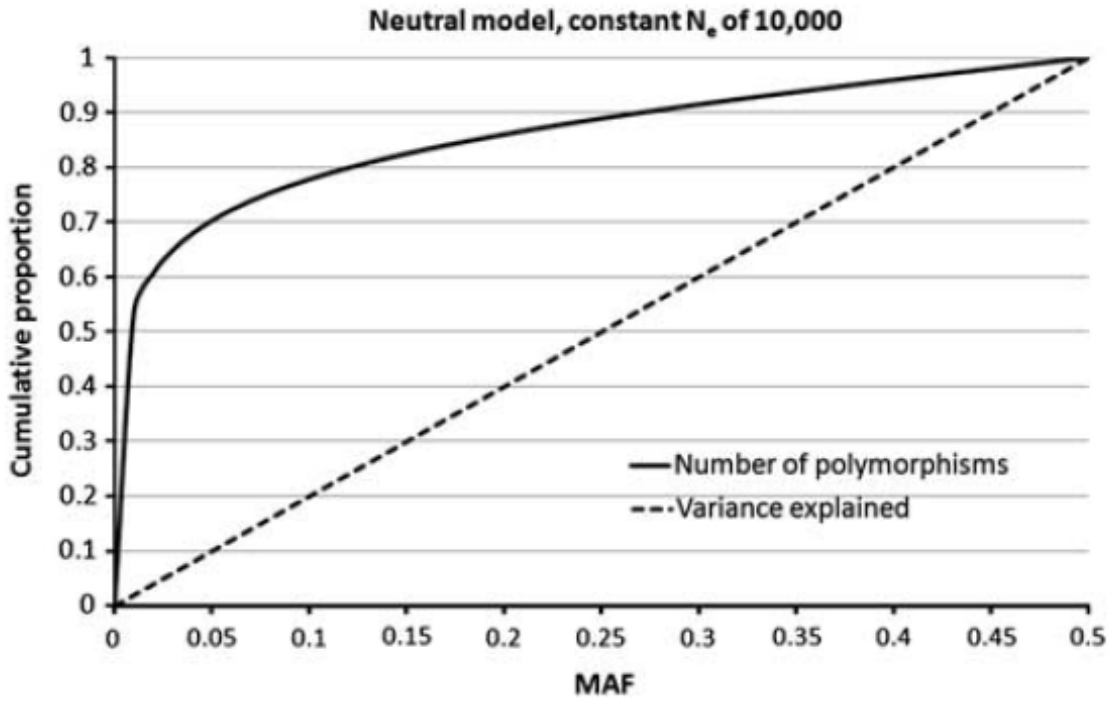
Prediction

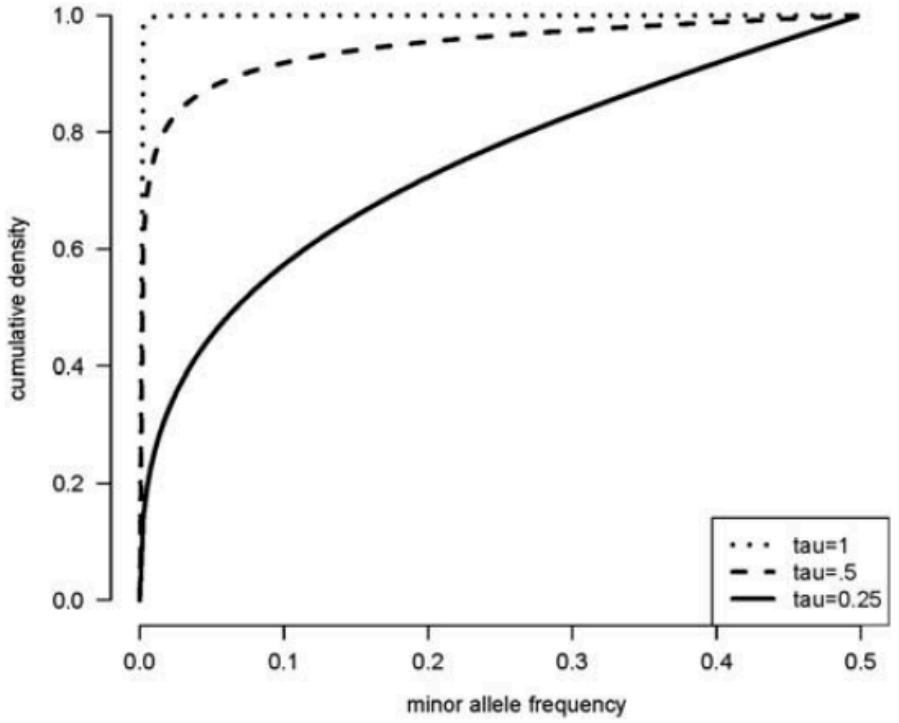
- Prediction from DNA sequence (or imputed SNP array) is limited by
  - how much phenotypic variance is captured by all variants
  - how well the effects of all variants are estimated

Imprecision Medicine



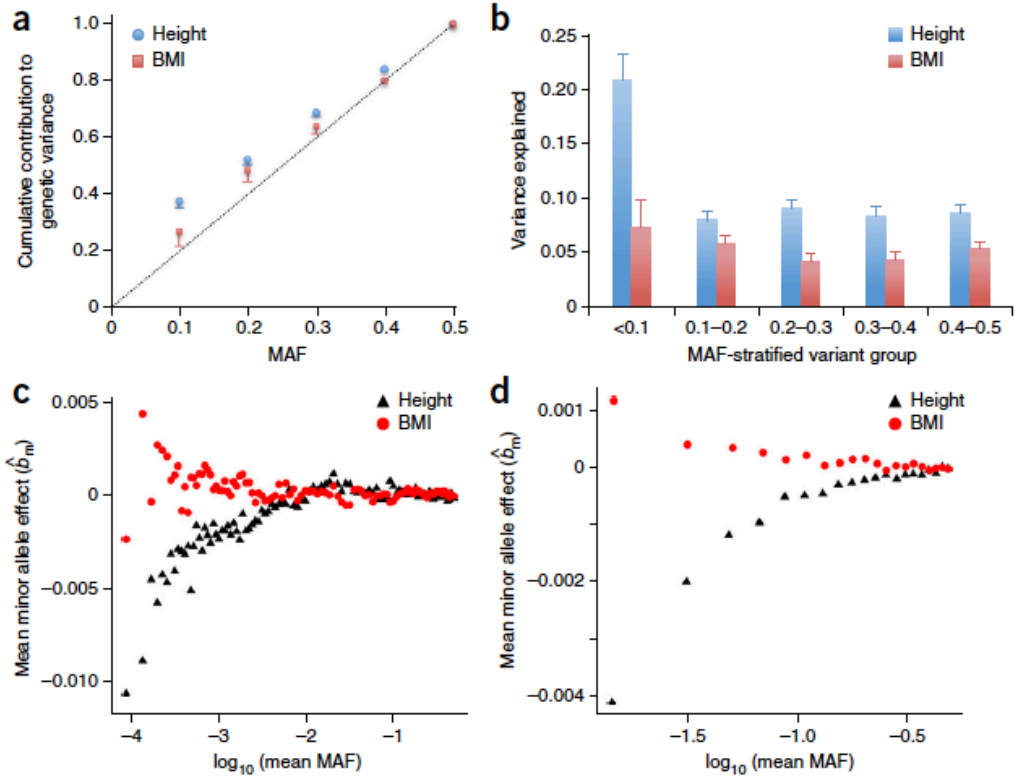
Past natural selection determines genetic architecture today





[Eyre-Walker 2010 PNAS; Visscher 2013 Mol Psych]

Evidence for association effect size and allele frequency among common variants



[Yang 2015 Nature Genetics]

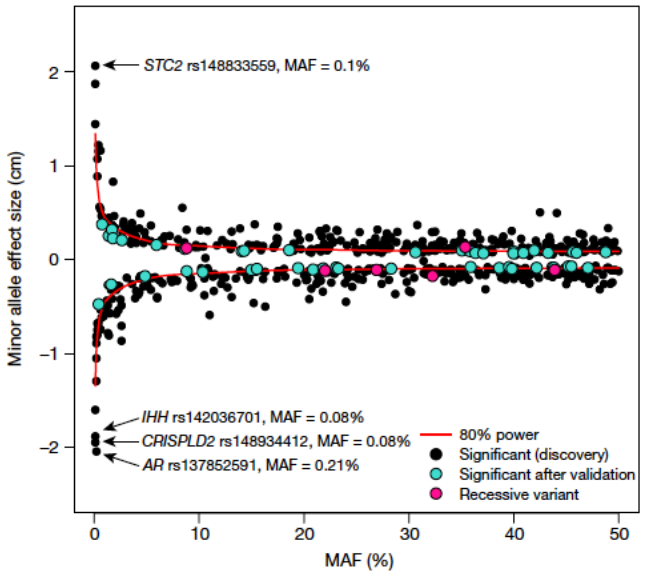
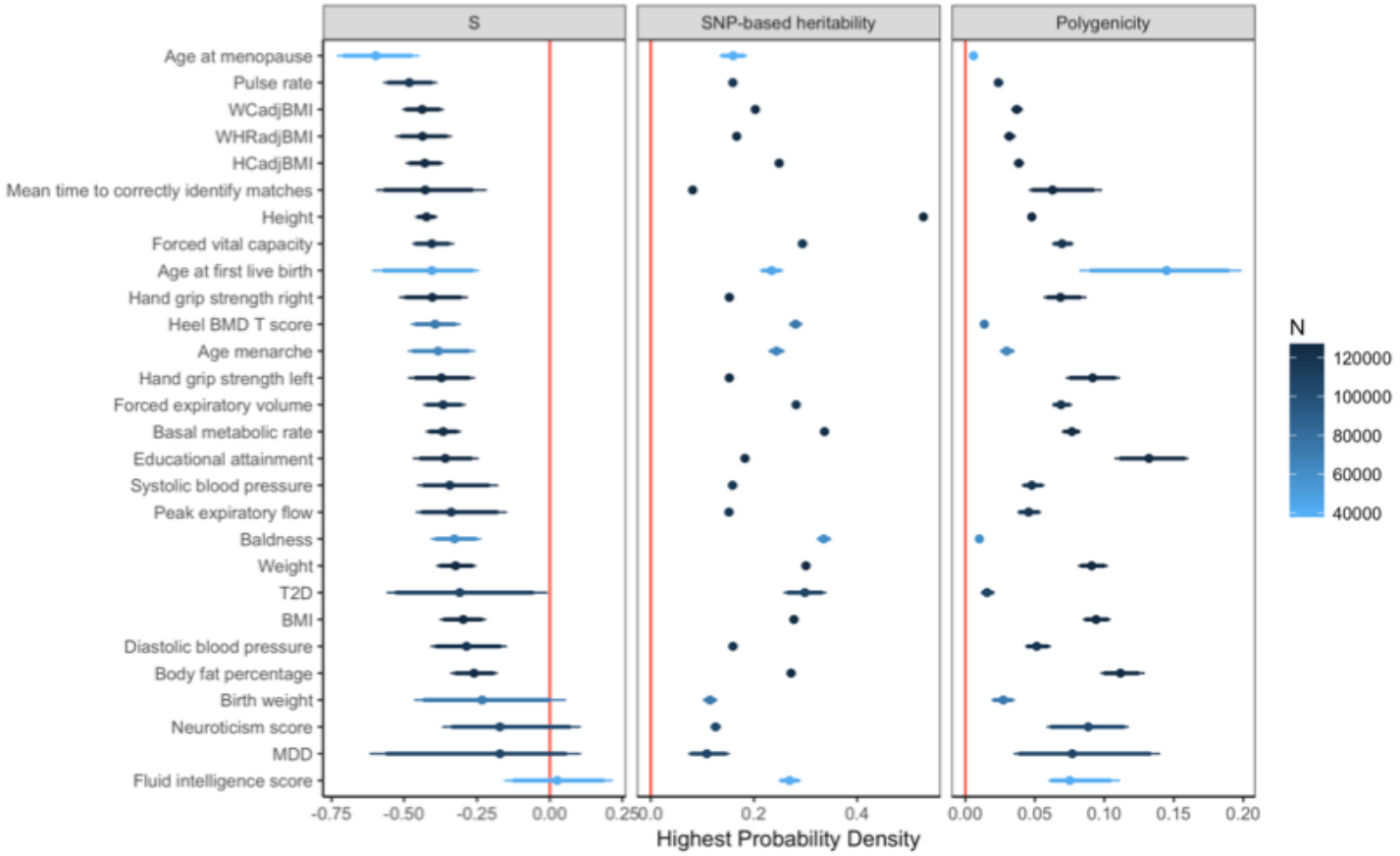
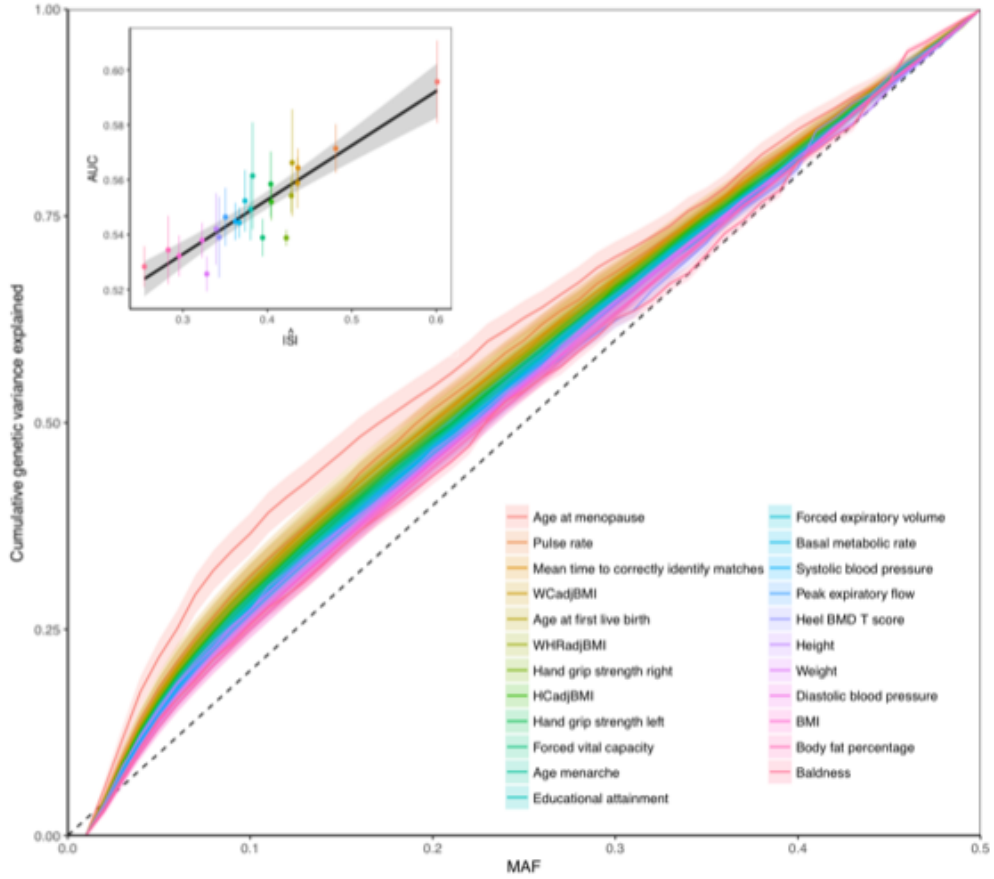


Figure 1 | Variants with a larger effect size on height variation tend to be rarer. An inverse relationship between the effect size (from the combined 'discovery and validation' analysis, in centimetres on the *y* axis) and the MAF for the height variants (*x* axis, from 0 to 50%) can be observed. Included in this figure are the 606 height variants with a $P < 2 \times 10^{-7}$.

[Marouli 2017 Nature]

Genetic architecture, selection and heritability





[Zeng et al. 2018 Nature Genetics]

Known unknowns

- Can we recover pedigree heritability from WGS data in a random sample from the population?
- How much trait variation is due to structural variation not captured by SNP chips and imputation?
- How much heritability is contributed by the X-chromosome?

Feasible studies in the near future

- Estimate and partition genetic variation using WGS with large sample sizes (> 50,000)
  - e.g. TOPMed, others
- Estimate genetic variance due to non-SNP variation
- Estimate genetic variance on the X chromosome
- Large family-based designs (e.g. 100,000 sibpairs; Young-Kong bioRxiv 2017)

Conclusions

- Complex traits are highly polygenic and pleiotropic
- Substantial proportion of genetic variance captured by SNPs arrays + imputation
- Not all traits are equal
- Evidence for selection on trait-associated loci
- WGS in combination with large sample sizes will provide currently missing information
- Large family studies needed to tease apart between and within family effects