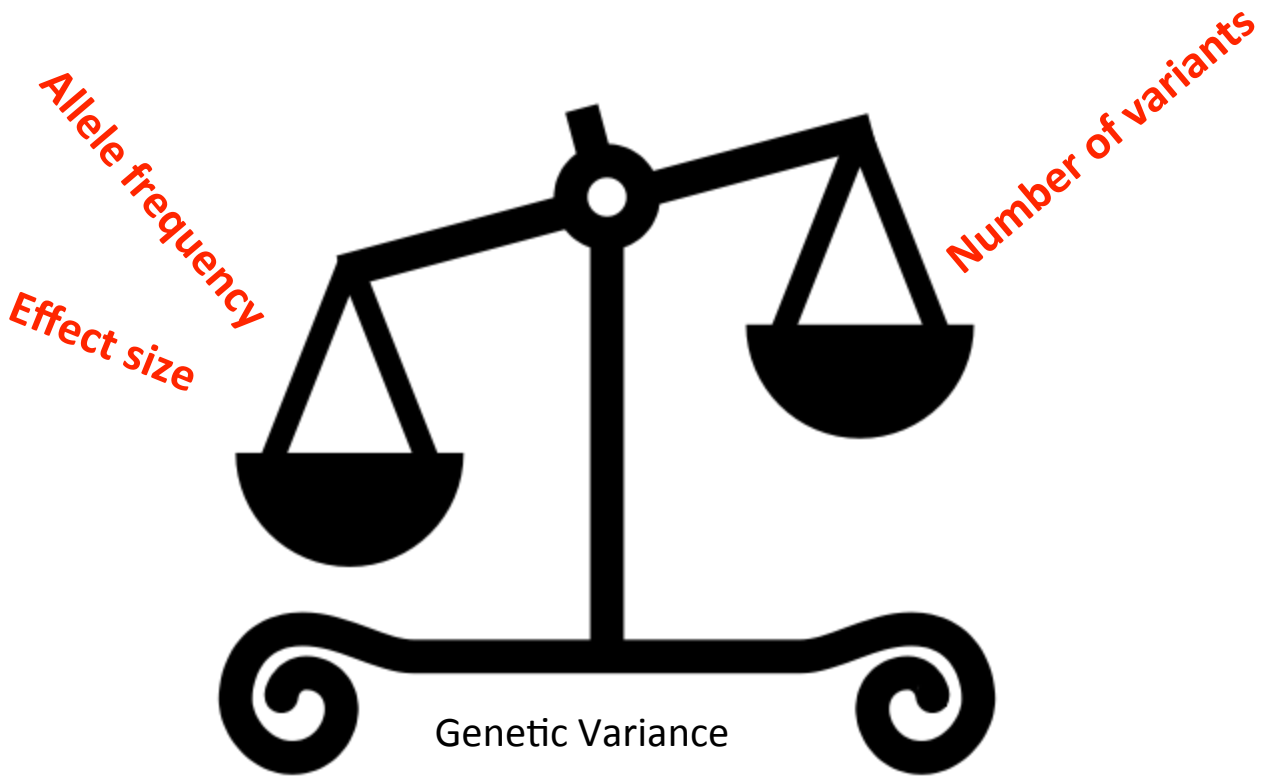
# Advances in genetic analyses of complex traits using GWAS data

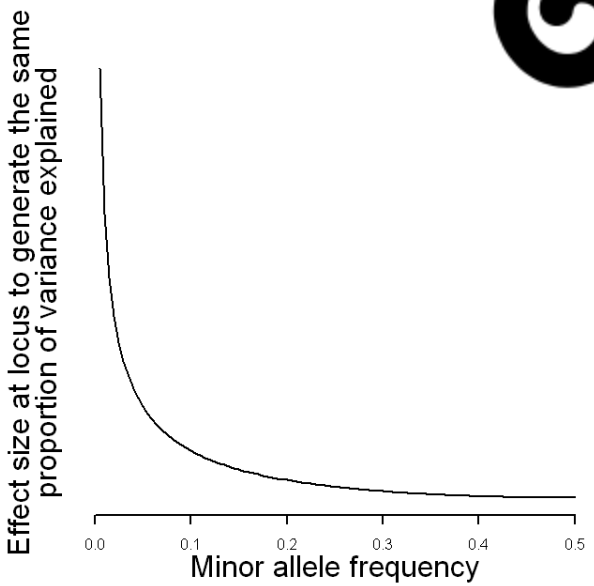
THE UNIVERSITY OF QUEENSLAND
AUSTRALIA

*Peter Visscher*

peter.visscher@uq.edu.au

# Genetic Architecture



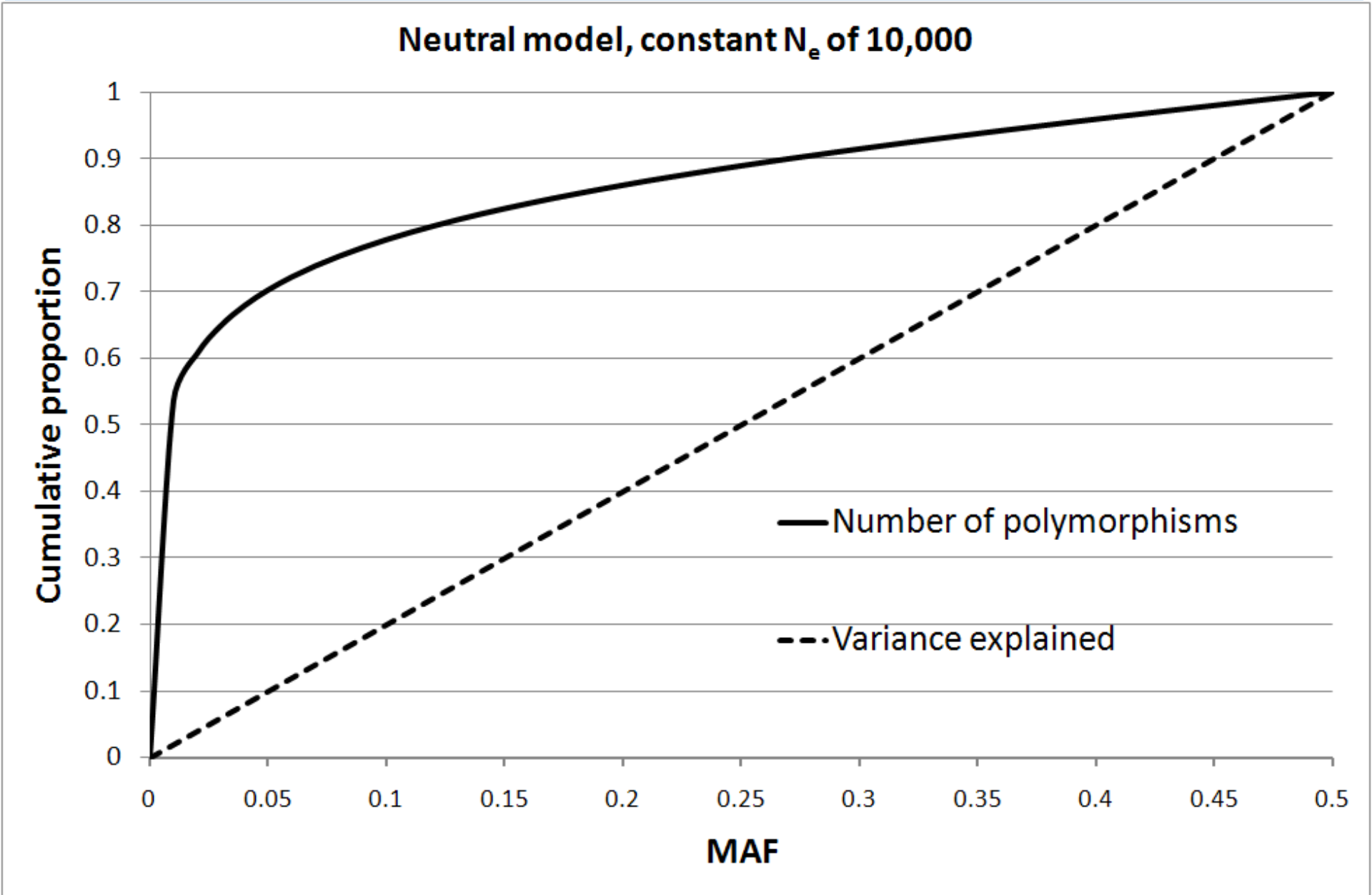




Variance explained for height

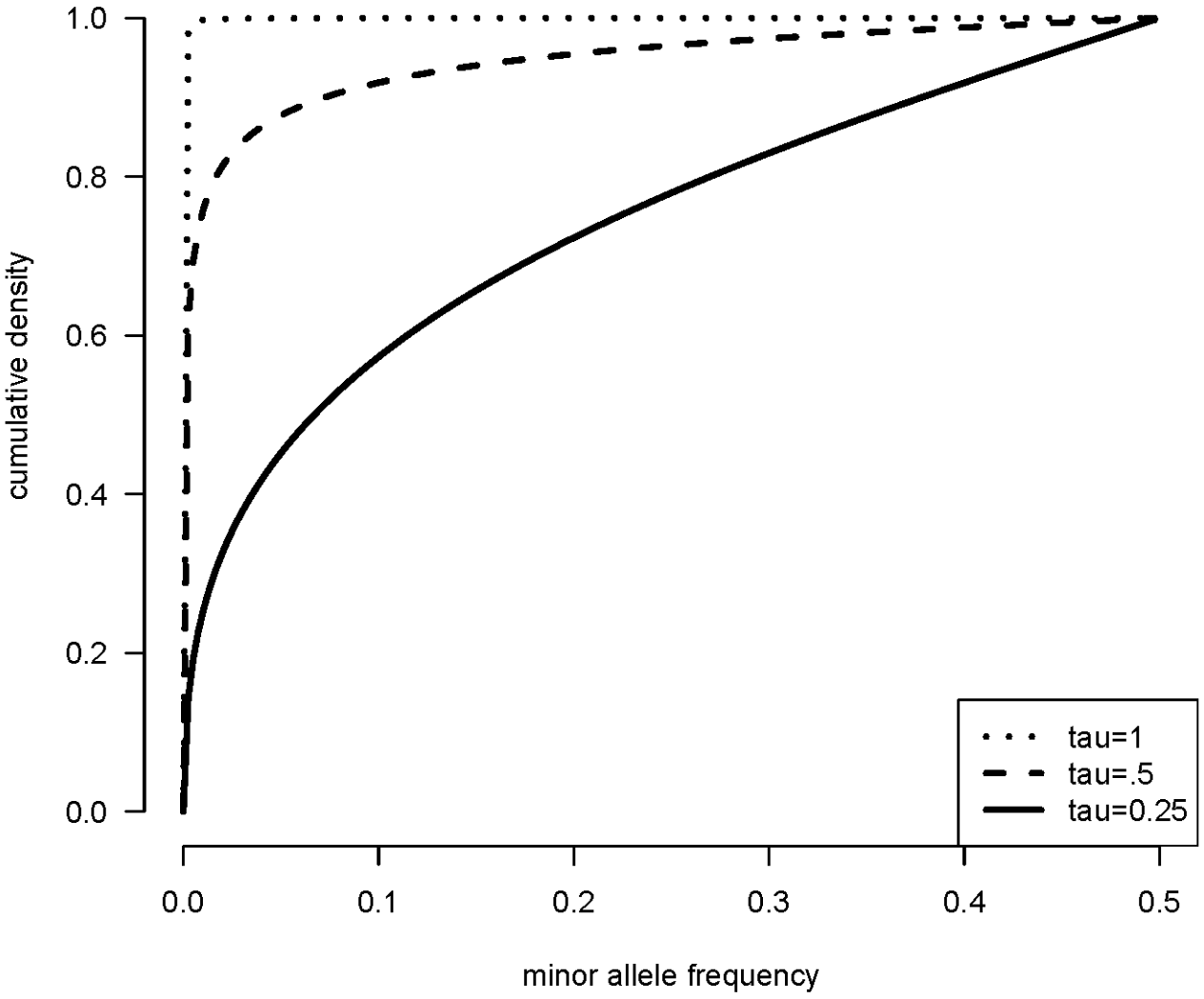
FBN1: ~0.3%

HMGA2: ~0.2%

# Neutral model: Most variants are rare. Most variation is due to common variants



# Natural selection: Most variation is due to rare variants for traits strongly correlated with fitness




[Eyre-Walker 2010, PNAS]

| Disease | Number of loci | Percent of Heritability Measure Explained | Heritability Measure |
|---|---|---|---|
| **Age-related macular degeneration** | 5 | 50% | Sibling recurrence risk |
| **Crohn's disease** | 32 | 20% | Genetic risk (liability) |
| **Systemic lupus erythematosus** | 6 | 15% | Sibling recurrence risk |
| **Type 2 diabetes** | 18 | 6% | Sibling recurrence risk |
| **HDL cholesterol** | 7 | 5.2% | Phenotypic variance |
| **Height** | 40 | 5% | Phenotypic variance |
| **Early onset myocardial infarction** | 9 | 2.8% | Phenot vari |
| **Fasting glucose** | 4 | 1.5% | Phenc varianc |

PLOS BIOLOGY

OPEN ACCESS Freely available online

## Rare Variants Create Synthetic Genome-Wide Associations

Samuel P. Dickson[1,2], Kai Wang[3], Ian Krantz[3,4,5], Hakon Hakonarson[3,4,5], David B. Goldstein[1*]



## Where is the Dark Matter?

Vol 461|8 October 2009|doi:10.1038/nature08494 nature

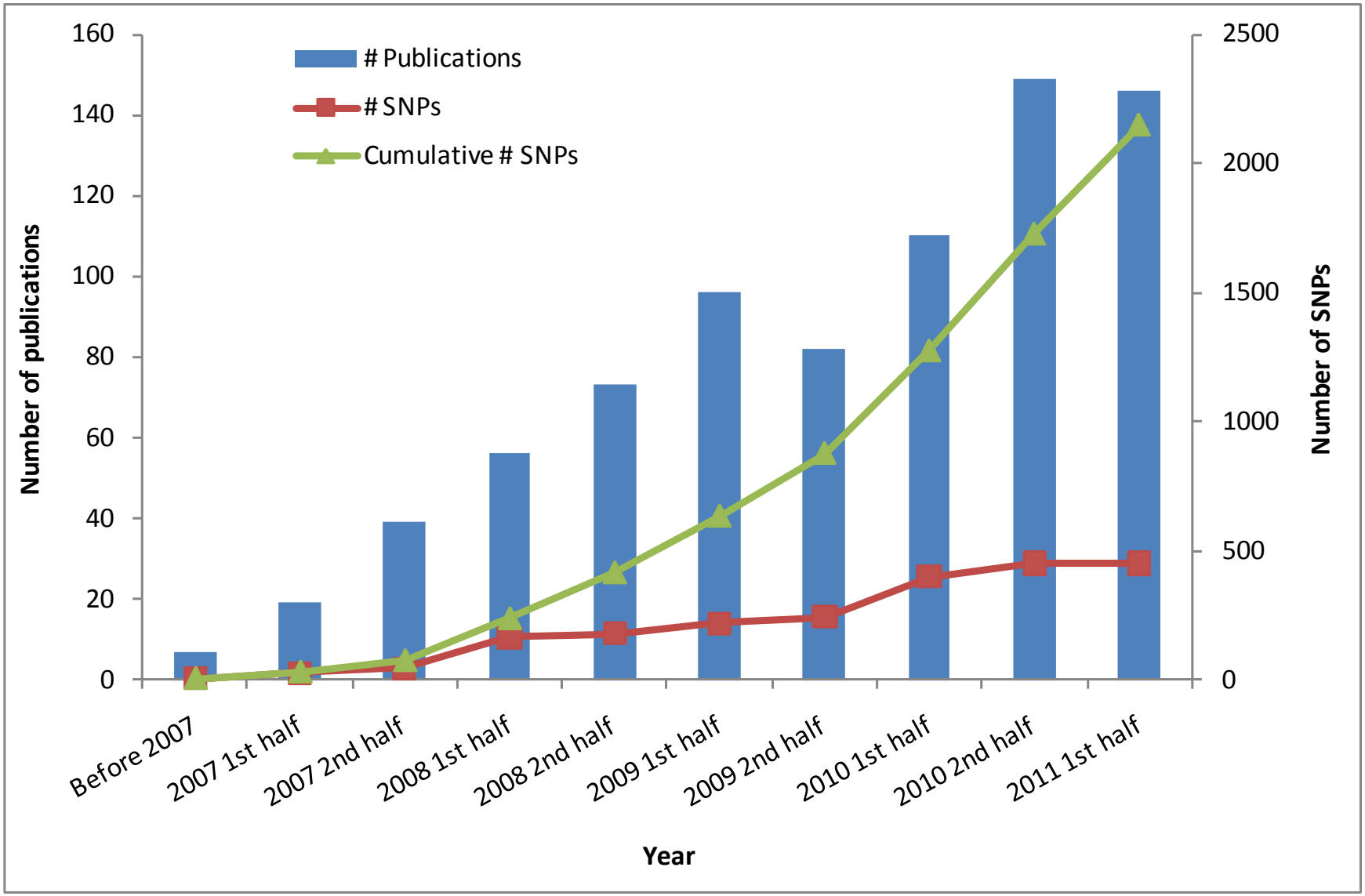
REVIEWS

## Finding the missing heritability of complex diseases

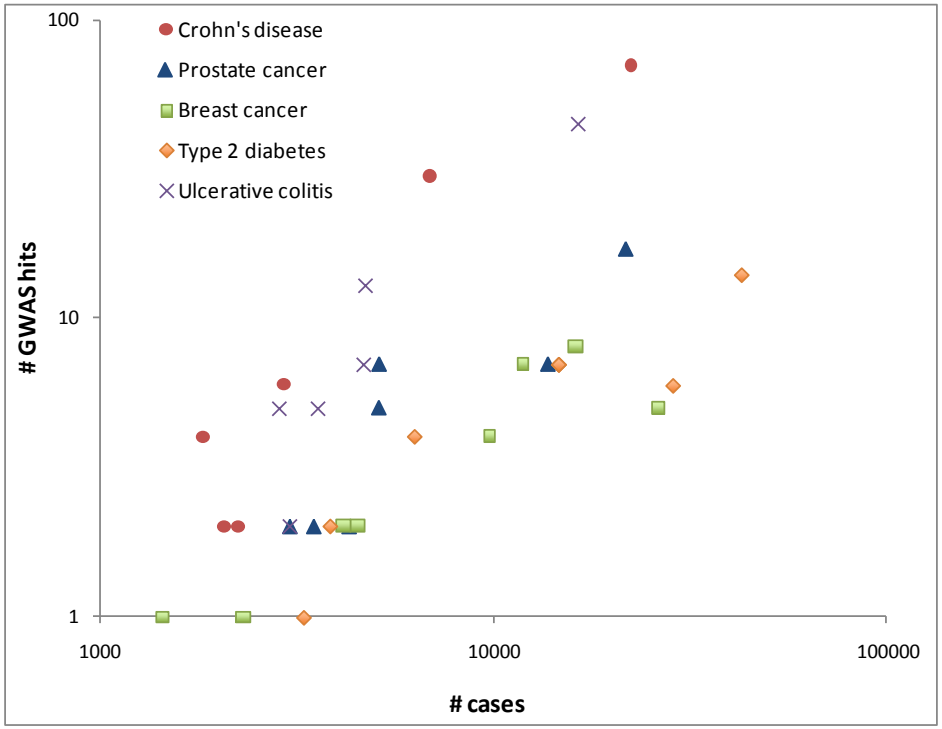
Teri A. Manolio[1], Francis S. Collins[2], Nancy J. Cox[3], David B. Goldstein[4], Lucia A. Hindorff[5], David J. Hunter[6], Mark I. McCarthy[7], Erin M. Ramos[5], Lon R. Cardon[8], Aravinda Chakravarti[9], Judy H. Cho[10], Alan E. Guttmacher[1], Augustine Kong[11], Leonid Kruglyak[12], Elaine Mardis[13], Charles N. Rotimi[14], Montgomery Slatkin[15], David Valle[9], Alice S. Whittemore[16], Michael Boehnke[17], Andrew G. Clark[18], Evan E. Eichler[19], Greg Gibson[20], Jonathan L. Haines[21], Trudy F. C. Mackay[22], Steven A. McCarroll[23] & Peter M. Visscher[24]

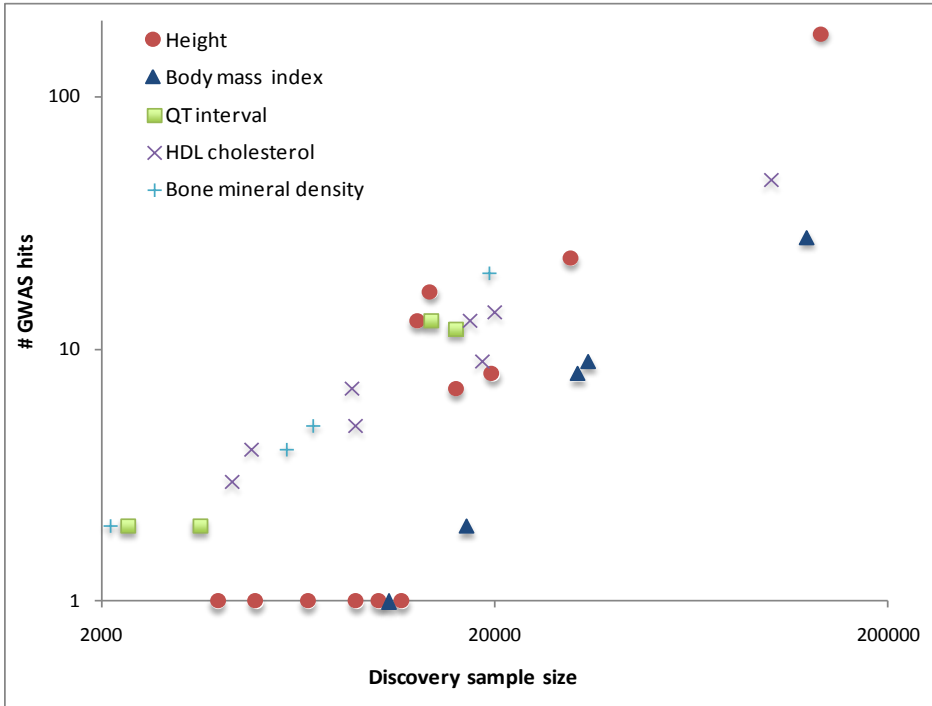
**The case of the missing heritability**

# An explosion of discoveries



# Bigger is better



# Possible explanations of missing heritability

- Heritability estimates from family data are wrong
- Causal variants are not tagged by SNP chips
- Effect sizes too small

- Interactions (GxE, GxG)
- Epigenetics
- Parent-of-origin effects
- …

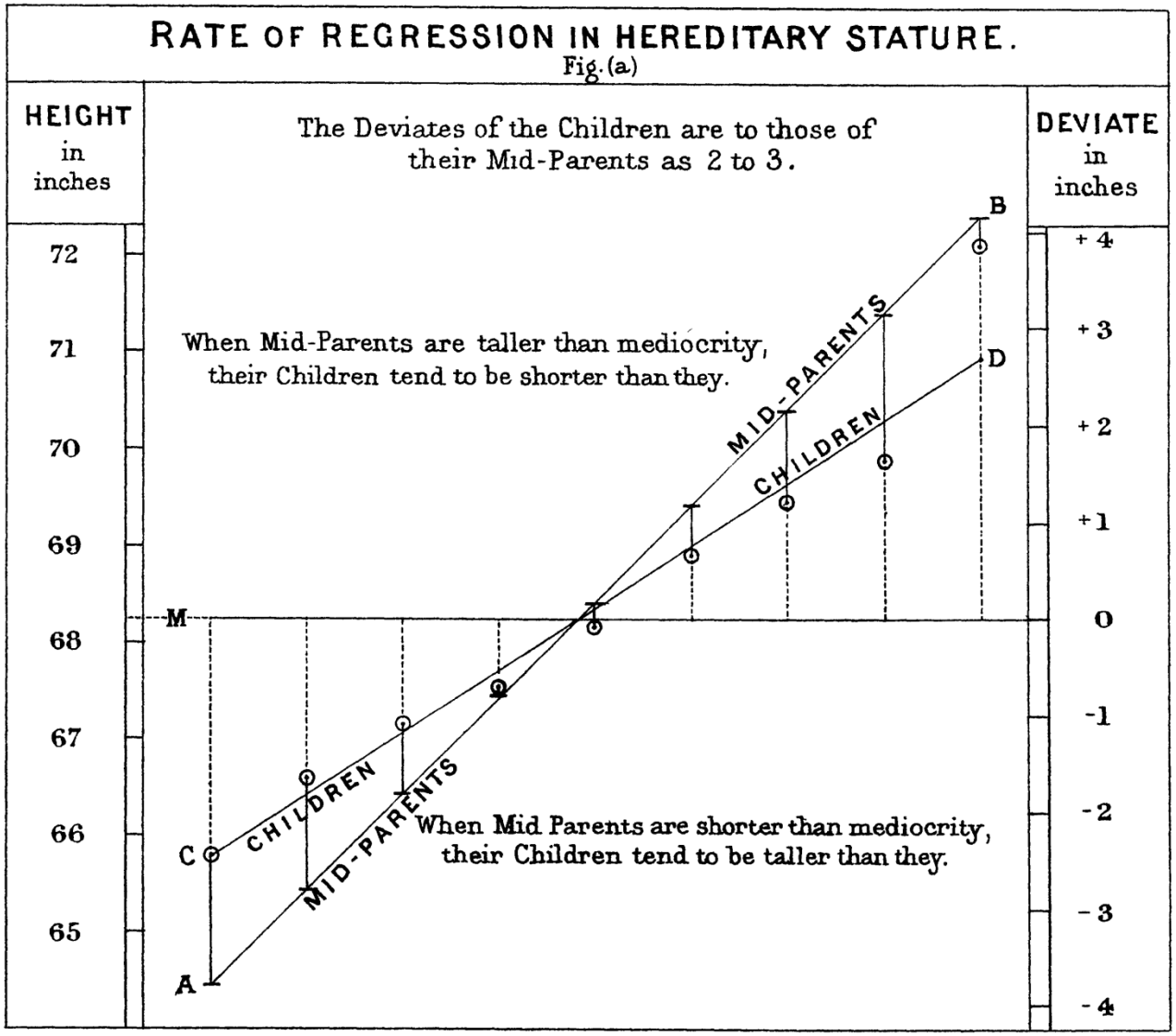
# Missing heritability explanations

Heritabilities are overestimated

1886

## Regression *towards* Mediocrity *in* Hereditary Stature.

By Francis Galton, F.R.S., &c.




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.*

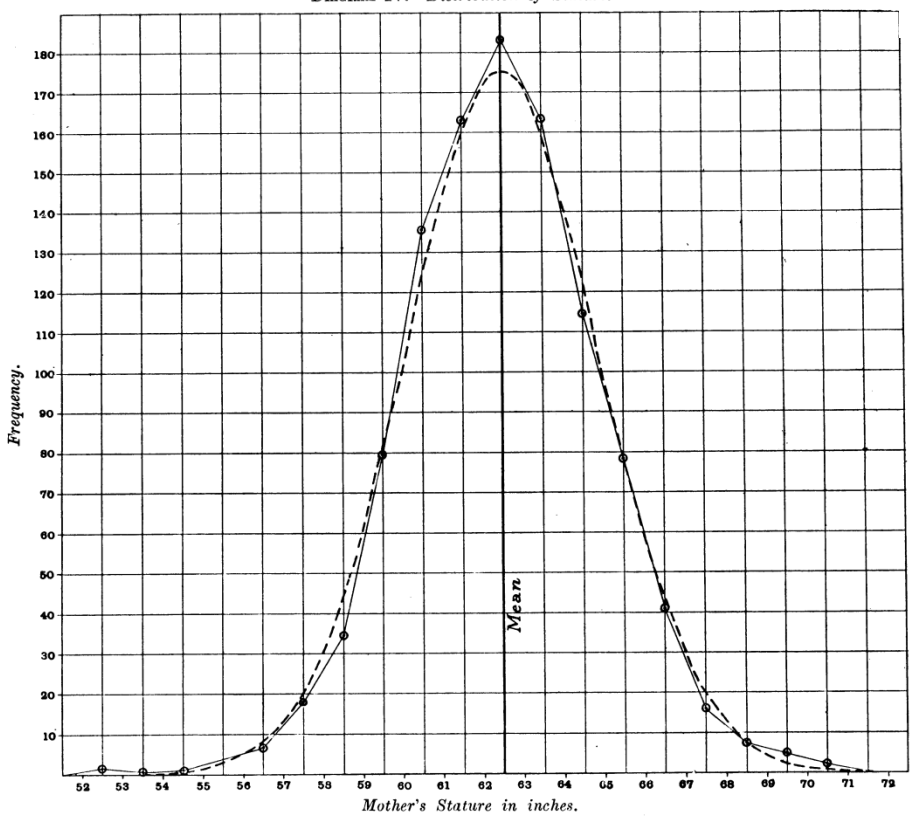
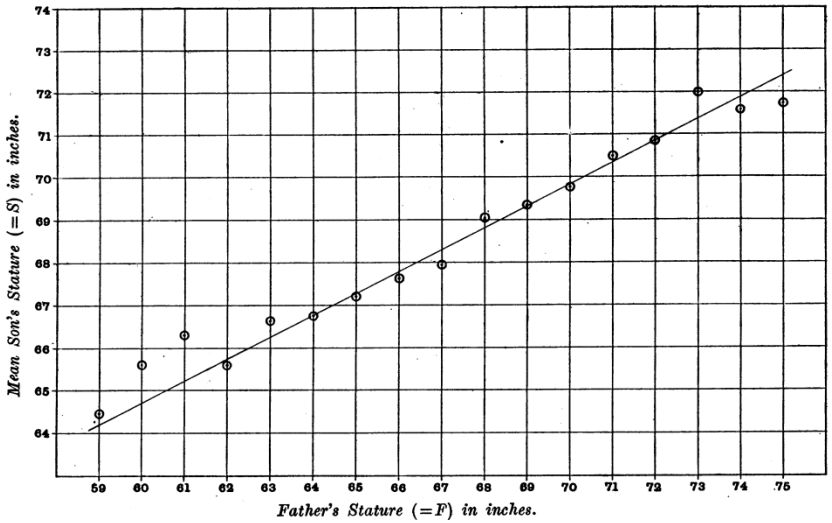


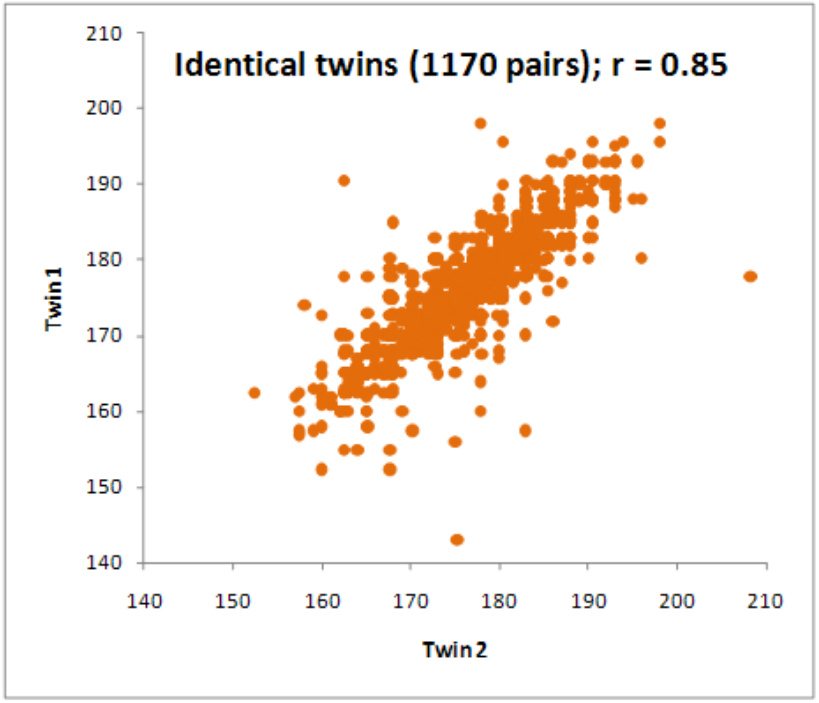

DIAGRAM I. *Probable Stature of Son for given Father's Stature.*

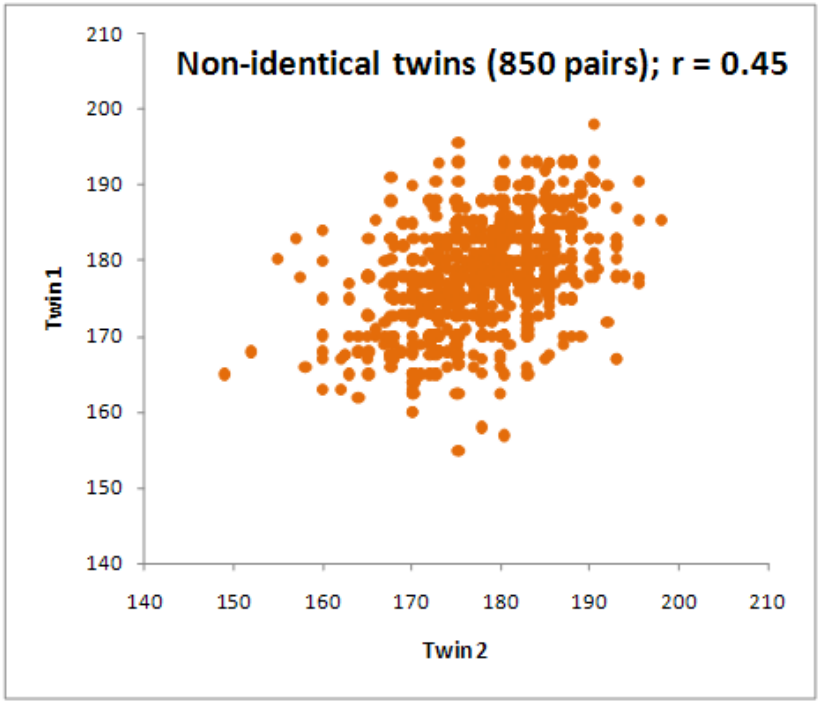
Regression Line: $S = 33{\cdot}73 + {\cdot}516\,F$. 1078 Cases.




| 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 |

# 100 years later
# Heritability of human height



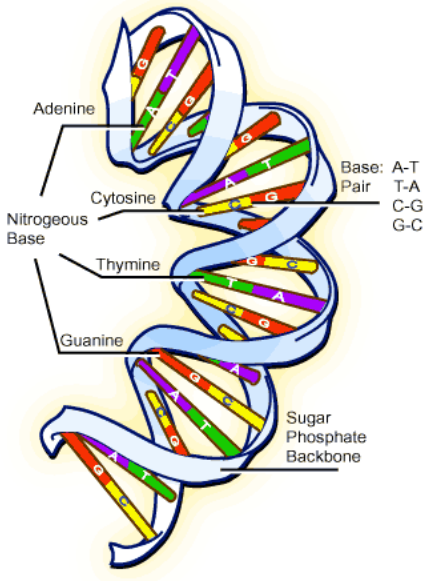




**$h^2 \sim 80\%$**

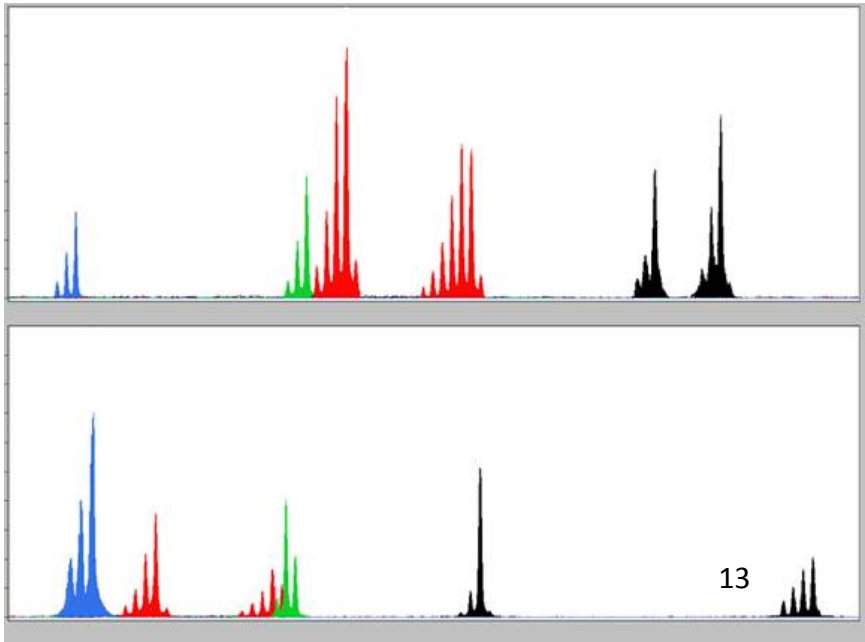
# Estimated heritability from family data can be wrong


## Solution: Estimating heritability within families

***Are fullsibs that share >50% of their genome by descent phenotypically more similar than those that share <50%?***




Need DNA samples

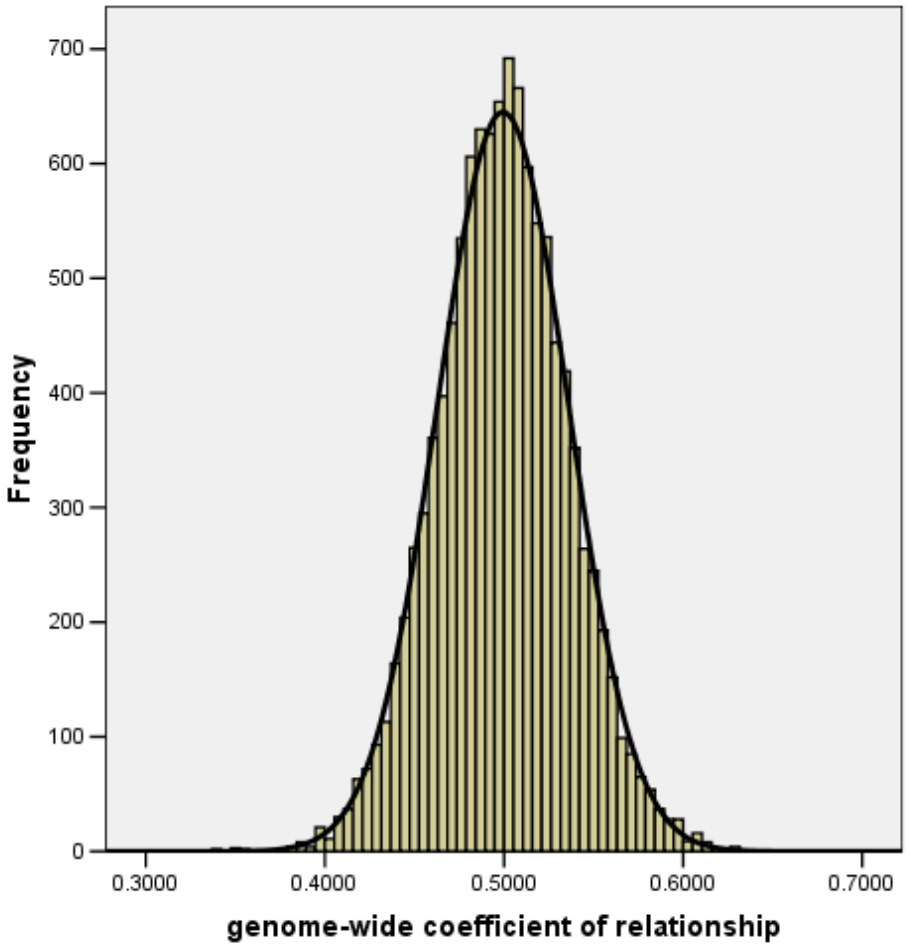


# Realised relationships

Mean 0.50, Range 0.31 – 0.64, SD 0.04

Brothers and sisters don't share exactly 50% of their genes from their mum and dad

Height (N = 11,214 pairs)
**$h^2$ = 86%**




OPEN ACCESS Freely available online PLOS GENETICS

Assumption-Free Estimation of Heritability from Genome-Wide Identity-by-Descent Sharing between Full Siblings

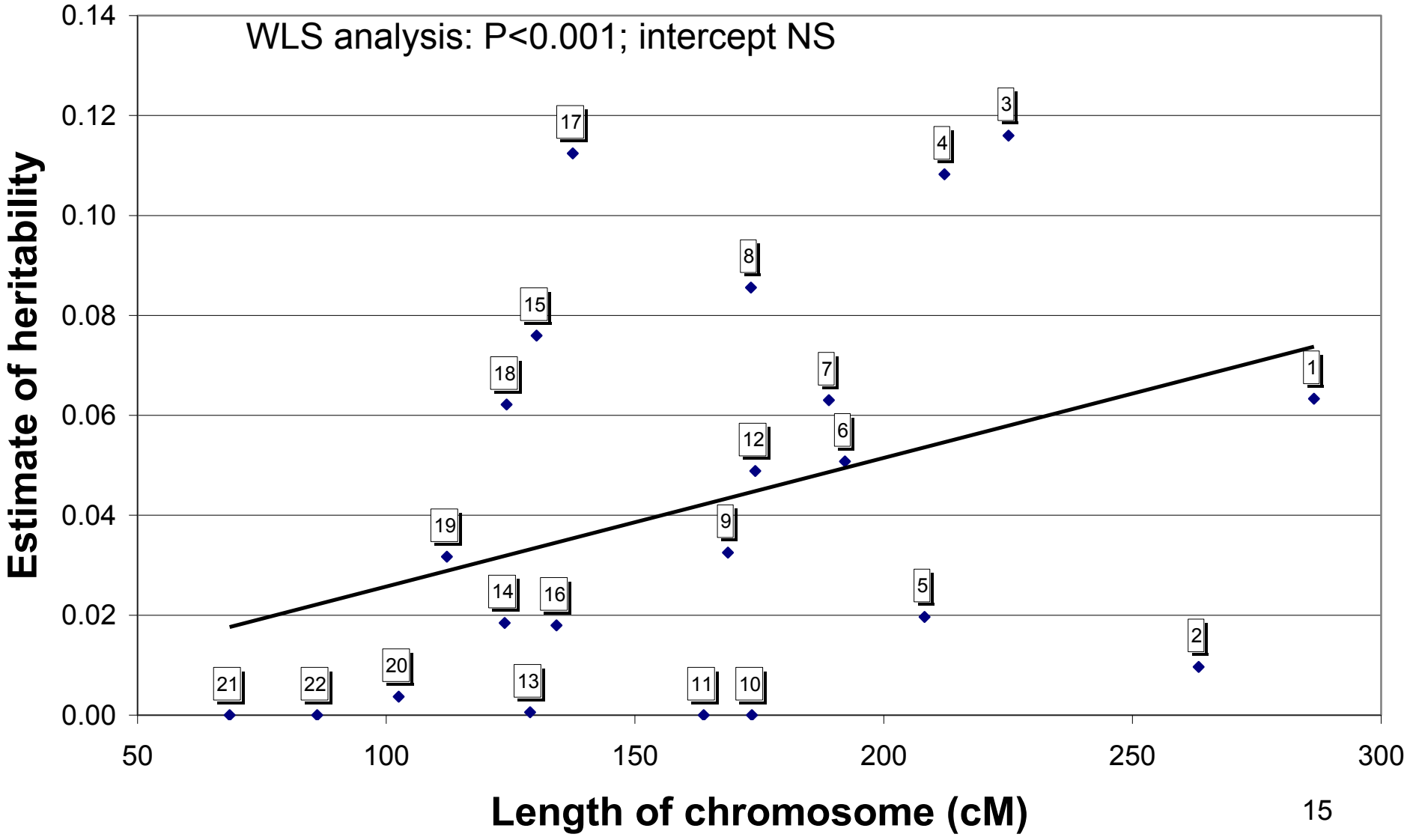
Peter M. Visscher*, Sarah E. Medland, Manuel A. R. Ferreira, Katherine I. Morley, Gu Zhu, Belinda K. Cornes, Grant W. Montgomery, Nicholas G. Martin

REPORT

Genome Partitioning of Genetic Variation for Height from 11,214 Sibling Pairs

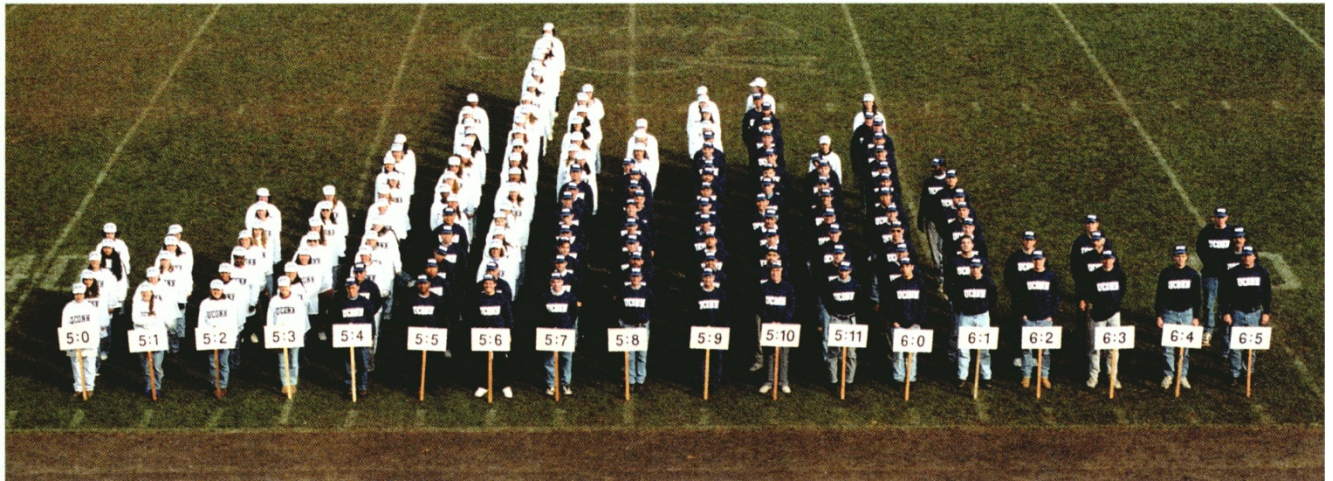
Peter M. Visscher, Stuart Macgregor, Beben Benyamin, Gu Zhu, Scott Gordon, Sarah Medland, William G. Hill, Jouke-Jan Hottenga, Gonneke Willemsen, Dorret I. Boomsma, Yao-Zhong Liu, Hong-Wen Deng, Grant W. Montgomery, and Nicholas G. Martin

## Longer chromosomes explain more additive genetic variance: ~0.03 per 100 cM



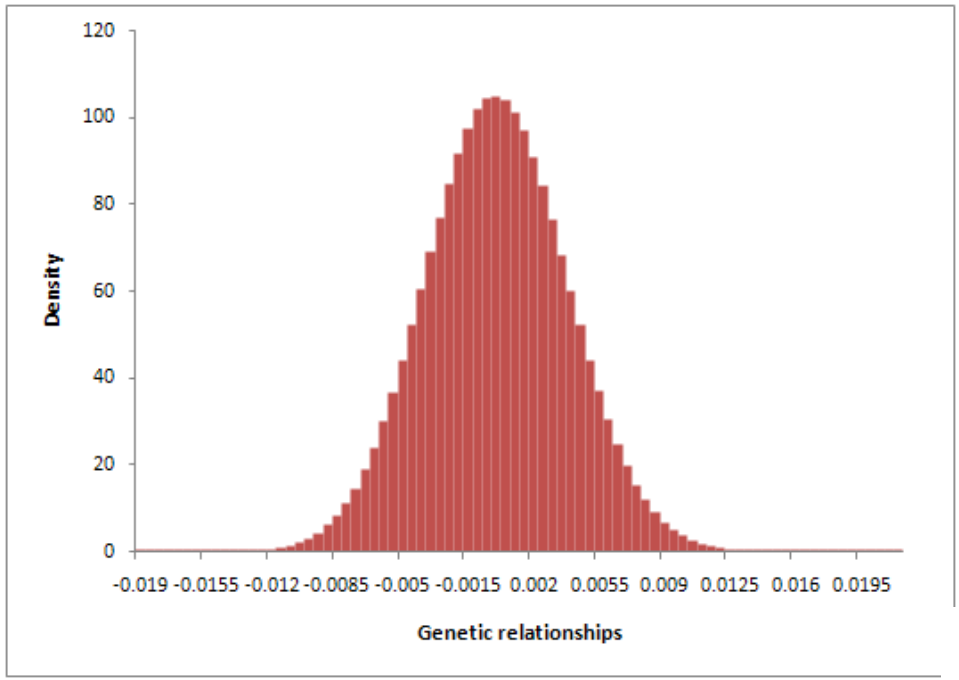
# One step further: Estimating heritability from 'unrelated' individuals

***Are very distant relatives that share more of their genome by descent phenotypically more similar than those that share less?***



# Data


- ~4000 ‘unrelated’ Australians
- Ancestry British Isles
- Measurement on height
- Measurement on 600,000 DNA markers




“Unrelated” pairs of individuals that are genetically similar tend to be more similar in height

$h^2$ ~ 50%
‘Chip heritability’

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]

# Genome partitioning

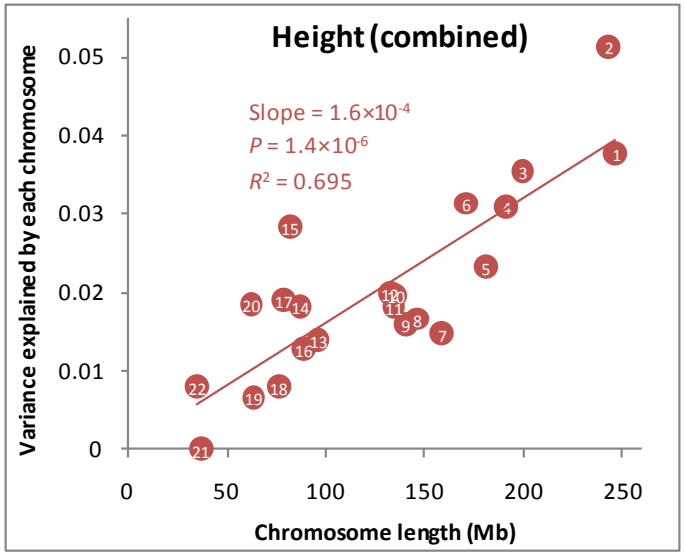
- Estimate genetic relationship matrix for each chromosome (or group of SNPs)
- Analyses of traits by fitting a relationship matrix for each chromosome
- Fit all autosomes simultaneously
  - Estimates of 22 variance components

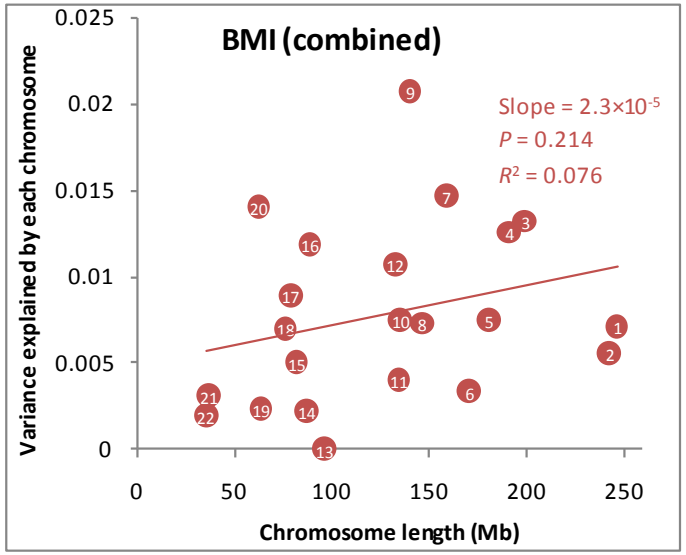
REPORT

GCTA: A Tool for Genome-wide Complex Trait Analysis

Jian Yang,[1,*] S. Hong Lee,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1]

# Genome-partitioning: longer chromosomes explain more variation

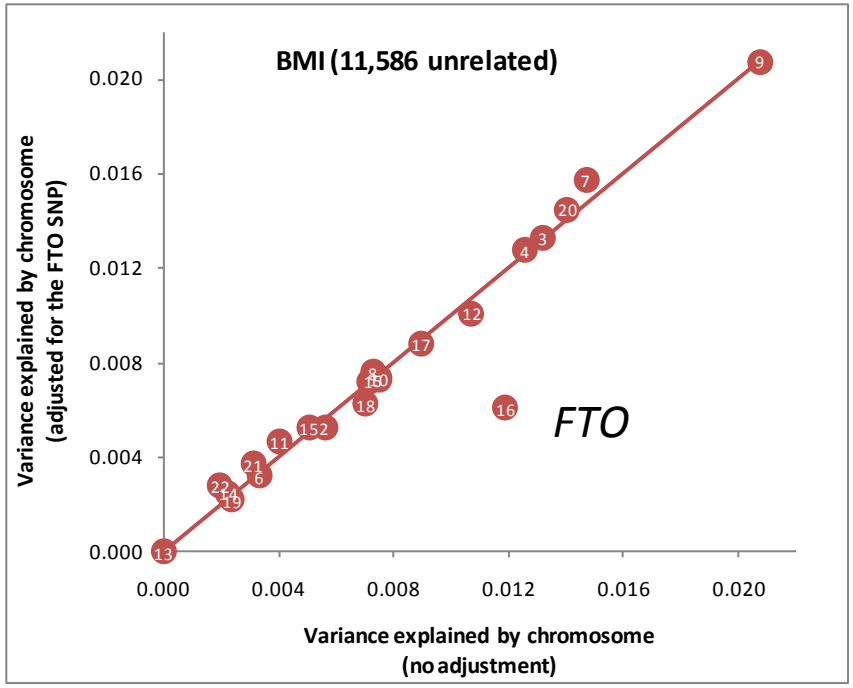


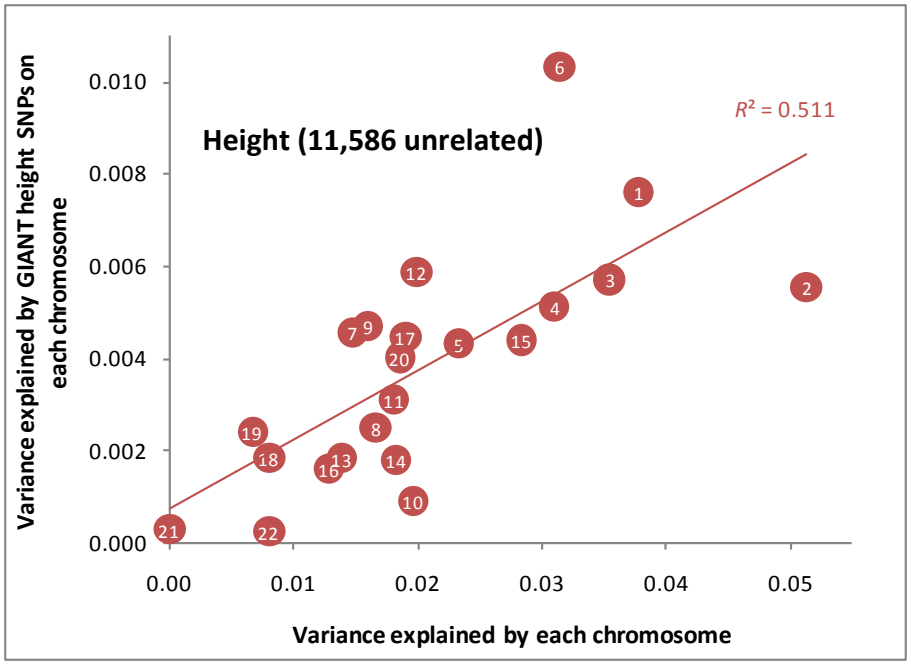




## Genome partitioning of genetic variation for complex traits using common SNPs

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]

# Results are consistent with reported GWAS



# Cognitive ability

Molecular Psychiatry (2011) 16, 996–1005

www.nature.com/mp

IMMEDIATE COMMUNICATION

**Genome-wide association studies establish that human intelligence is highly heritable and polygenic**

G Davies[1], A Tenesa[2,3], A Payton[4], J Yang[5], SE Harris[6,7], D Liewald[1,7], X Ke[8], S Le Hellard[9], A Christoforou[9], M Luciano[1,7], K McGhee[1], L Lopez[1,7], AJ Gow[1,7], J Corley[1], P Redmond[1], HC Fox[10], P Haggarty[11], LJ Whalley[10], G McNeill[10], ME Goddard[12,13], T Espeseth[14], AJ Lundervold[15], I Reinvang[14] A Pickles[16], VM Steen[9,17], W Ollier[4], DJ Porteous[6,7], M Horan[18], JM Starr[7,19], N Pendleton[18], PM Visscher[5,7,20] and IJ Deary[1,7,20]

Intelligence ‘looks like’ height…

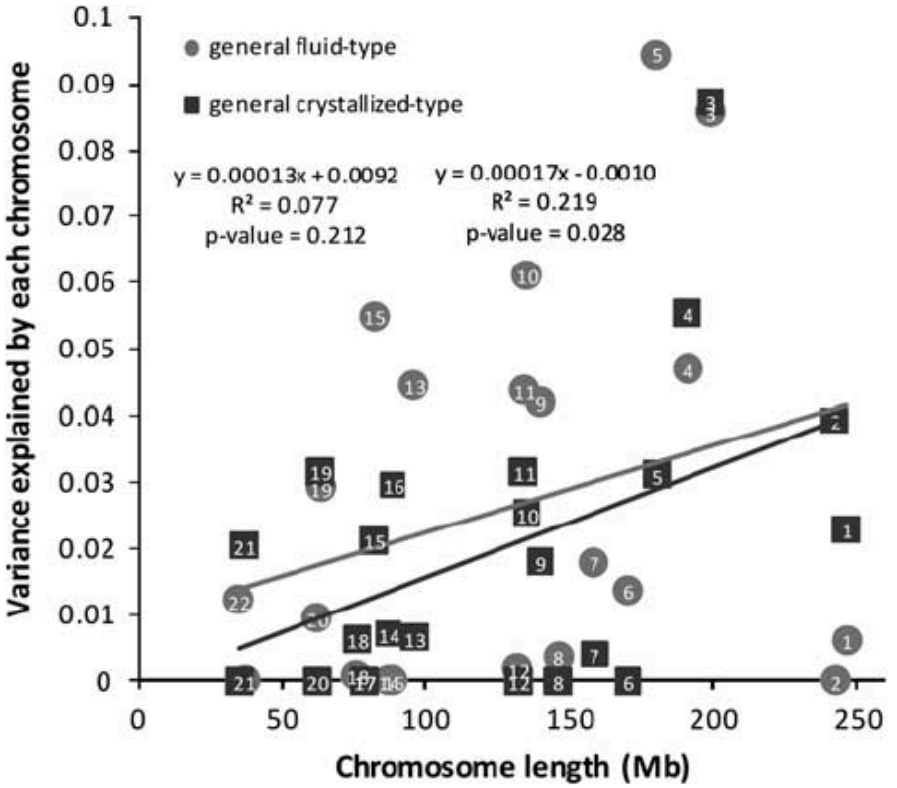



**Figure 2** Estimate of the proportion of variance explained by each chromosome for general fluid intelligence factor ($g_f$) and crystallized intelligence ($g_c$) in the combined dataset against chromosome length. The numbers in the circles and squares are the chromosome numbers.

# Mimicking a conditional & joint association analysis from meta-analysis data

- Standard GWAS (or MA) analysis is one SNP at a time and does not use the correlation (LD) between SNPs
- A multiple SNP association analysis implicitly uses the observed correlation (LD) between SNPs
- Mimic multiple SNP analysis by using the LD correlation structure from a reference sample

# Data requirement and assumptions

- Minimum data requirements
  - Summary statistics: effect size (β) and standard error (SE) from the discovery set (e.g. MA)
    - & allele frequency (p), but this could come from reference sample
  - A reference sample with individual-level genotype data

- Assumptions
  - HWE
  - The LD structure in the discovery set is similar to that in the reference sample
  - SNPs > d Mb (e.g. d = 20) away from each other are independent (LD $r^2 = 0$)

# Conditional & joint association analysis

## Height examples

GIANT discovery data (summary statistics of ~133000 samples and 2.8M SNPs). LD structure mimicked by QIMR data (~4000 samples)

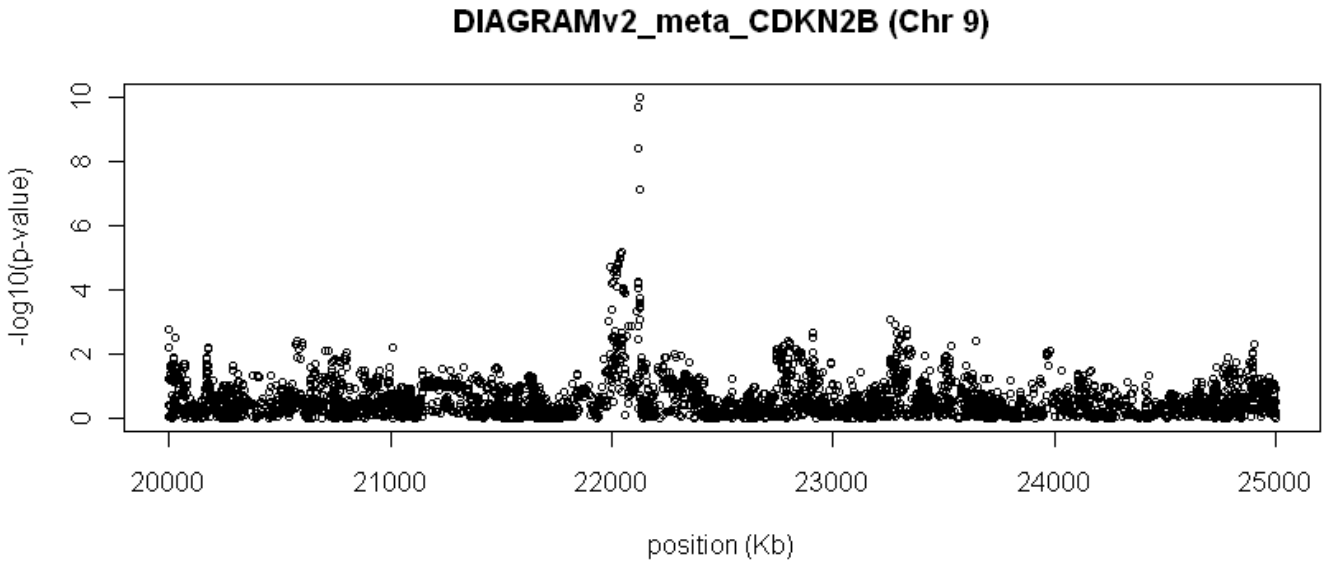
254 hits with $p_J < 5e\text{-}8$

Evidence for multiple signals within 1-Mb loci

| | | | | GIANT MA summary statistics | | | | | Joint analysis (LD from QIMR data) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Chr | SNP | bp | refA | freq | b | se | p | N | freq | $b_J$ | se | p |
| 1 | rs6684205 | 216,676,325 | A | 0.714 | -0.0328 | 0.0041 | 1.6E-15 | 145782 | 0.719 | -0.0347 | 0.0042 | 7.3E-17 |
| 1 | rs11118171 | 217,114,492 | A | 0.631 | 0.0248 | 0.0039 | 2.2E-10 | 141216 | 0.640 | 0.0241 | 0.0039 | 9.1E-10 |
| 1 | rs11118346 | 217,810,342 | T | 0.464 | -0.0264 | 0.0037 | 1.7E-12 | 146866 | 0.459 | -0.0240 | 0.0037 | 6.6E-11 |
| 2 | rs6436071 | 219,306,285 | A | 0.533 | 0.0258 | 0.0038 | 6.1E-12 | 139125 | 0.525 | 0.0221 | 0.0038 | 5.7E-09 |
| 2 | rs12470505 | 219,616,613 | T | 0.902 | 0.0483 | 0.0063 | 1.8E-14 | 142384 | 0.904 | 0.0426 | 0.0065 | 5.0E-11 |
| 2 | rs16859517 | 219,657,428 | T | 0.036 | 0.0734 | 0.0107 | 5.8E-12 | 124840 | 0.042 | 0.0655 | 0.0101 | 7.5E-11 |
| 5 | rs6450922 | 32,725,475 | C | 0.752 | 0.0399 | 0.0058 | 3.9E-12 | 79719 | 0.739 | 0.0377 | 0.0060 | 3.2E-10 |
| 5 | rs3811968 | 32,801,246 | A | 0.302 | 0.0291 | 0.0041 | 8.8E-13 | 141025 | 0.302 | 0.0344 | 0.0041 | 5.8E-17 |
| 5 | rs1173727 | 32,866,278 | T | 0.394 | 0.0356 | 0.0038 | 1.5E-20 | 145074 | 0.403 | 0.0359 | 0.0038 | 7.0E-21 |
| 5 | rs11745439 | 33,265,791 | A | 0.288 | -0.0281 | 0.0042 | 1.3E-11 | 138115 | 0.271 | -0.0261 | 0.0042 | 7.2E-10 |
| 5 | rs4620037 | 170,807,702 | A | 0.801 | 0.0322 | 0.0046 | 3.6E-12 | 148018 | 0.790 | 0.0334 | 0.0046 | 3.1E-13 |
| 5 | rs1529701 | 170,933,582 | T | 0.292 | -0.0228 | 0.0042 | 4.9E-08 | 137051 | 0.299 | -0.0253 | 0.0042 | 1.4E-09 |
| 5 | rs12153391 | 171,136,043 | A | 0.255 | -0.0329 | 0.0045 | 2.4E-13 | 130140 | 0.267 | -0.0306 | 0.0045 | 1.3E-11 |
| 5 | rs4868126 | 171,216,074 | T | 0.396 | -0.0305 | 0.0042 | 3.3E-13 | 118495 | 0.393 | -0.0297 | 0.0042 | 1.7E-12 |
| 6 | rs12204421 | 33,736,841 | A | 0.738 | 0.0303 | 0.0044 | 5.6E-12 | 133438 | 0.748 | 0.0279 | 0.0045 | 5.6E-10 |
| 6 | rs2780226 | 34,307,070 | T | 0.925 | -0.079 | 0.0075 | 3.4E-26 | 127971 | 0.916 | -0.0766 | 0.0072 | 4.2E-26 |
| 6 | rs2814985 | 34,656,274 | T | 0.149 | 0.0474 | 0.0053 | 6.3E-19 | 140457 | 0.155 | 0.0398 | 0.0053 | 4.6E-14 |
| 6 | rs6915007 | 35,030,171 | T | 0.022 | -0.1038 | 0.0137 | 3.1E-14 | 123266 | 0.022 | -0.0901 | 0.0141 | 1.6E-10 |
| 6 | rs975496 | 35,341,421 | A | 0.150 | -0.0409 | 0.0053 | 8.4E-15 | 139992 | 0.150 | -0.0304 | 0.0055 | 3.1E-08 |
| 9 | rs10512248 | 97,299,524 | T | 0.660 | -0.0329 | 0.004 | 1.5E-16 | 139340 | 0.665 | -0.0297 | 0.0041 | 5.6E-13 |
| 9 | rs817300 | 97,420,043 | A | 0.065 | -0.082 | 0.0096 | 1.1E-17 | 89014.6 | 0.060 | -0.0618 | 0.0100 | 7.4E-10 |
| 9 | rs12343296 | 97,438,598 | T | 0.141 | 0.0352 | 0.0055 | 1.3E-10 | 136388 | 0.144 | 0.0318 | 0.0055 | 7.0E-09 |
| 9 | rs2025151 | 98,201,333 | C | 0.810 | -0.0406 | 0.005 | 3.1E-16 | 129850 | 0.807 | -0.0391 | 0.0049 | 2.7E-15 |
| 15 | rs4932199 | 87,142,054 | A | 0.526 | -0.0265 | 0.0042 | 2.2E-10 | 113677 | 0.522 | -0.0290 | 0.0042 | 3.7E-12 |
| 15 | rs16942341 | 87,189,909 | T | 0.031 | -0.1335 | 0.0131 | 2.2E-24 | 97910.9 | 0.030 | -0.1130 | 0.0134 | 2.9E-17 |
| 15 | rs2280470 | 87,196,630 | A | 0.334 | 0.0394 | 0.004 | 1.5E-22 | 140401 | 0.327 | 0.0361 | 0.0040 | 4.4E-19 |
| 20 | rs7274811 | 31,796,842 | T | 0.231 | -0.0402 | 0.0045 | 7.9E-19 | 138955 | 0.263 | -0.0354 | 0.0043 | 3.0E-16 |
| 20 | rs2223553 | 32,173,105 | A | 0.600 | -0.0043 | 0.0038 | 2.7E-01 | 144215 | 0.580 | -0.0229 | 0.0040 | 1.2E-08 |
| 20 | rs6579167 | 32,496,576 | T | 0.795 | -0.0384 | 0.0048 | 7.2E-16 | 133158 | 0.780 | -0.0474 | 0.0051 | 6.7E-21 |
| 20 | rs143384 | 33,489,170 | A | 0.581 | -0.0639 | 0.0041 | 0 | 122167 | 0.597 | -0.0474 | 0.0050 | 4.8E-21 |
| 20 | rs224343 | 33,500,113 | A | 0.761 | -0.0532 | 0.0045 | 4.1E-32 | 135913 | 0.774 | -0.0335 | 0.0056 | 2.2E-09 |

# T2D example

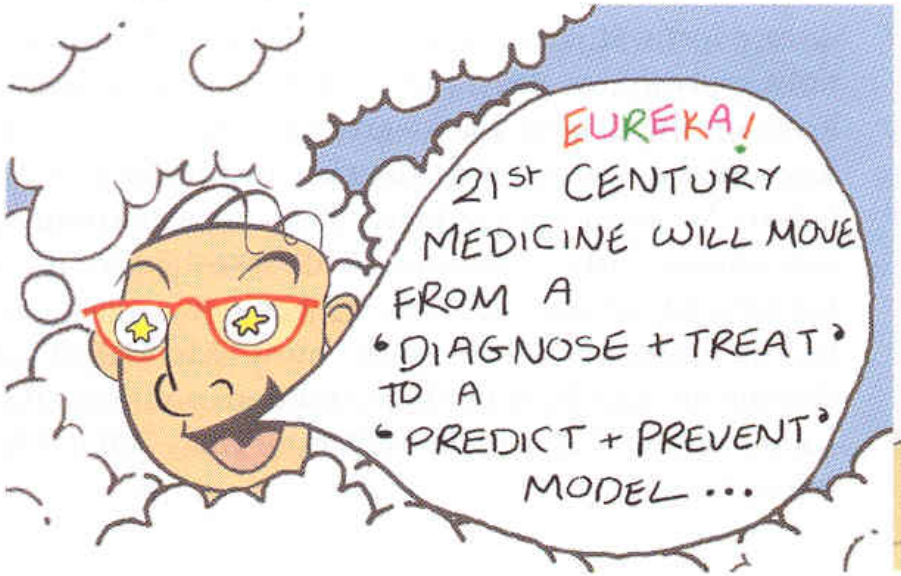
## Conditional & joint association analysis




| | | | | DIAGRAM meta-analysis | | | Conditional & joint analysis QIMR data (3924 unrelated) | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Chr | SNP | bp | Risk Allele | logOR | se | p | freq | logOR | se | p |
| 9 | rs10965250 | 22,123,284 | G | 0.181 | 0.028 | 1.0E-10 | 0.828 | 0.332 | 0.033 | 1.7E-23 |
| 9 | rs10757282 | 22,123,984 | C | 0.097 | 0.027 | 3.1E-04 | 0.373 | 0.269 | 0.032 | 3.3E-17 |

Pearson correlation between rs10965250 and rs10757282 is -0.59 in the QIMR data.

# Caveat on risk prediction

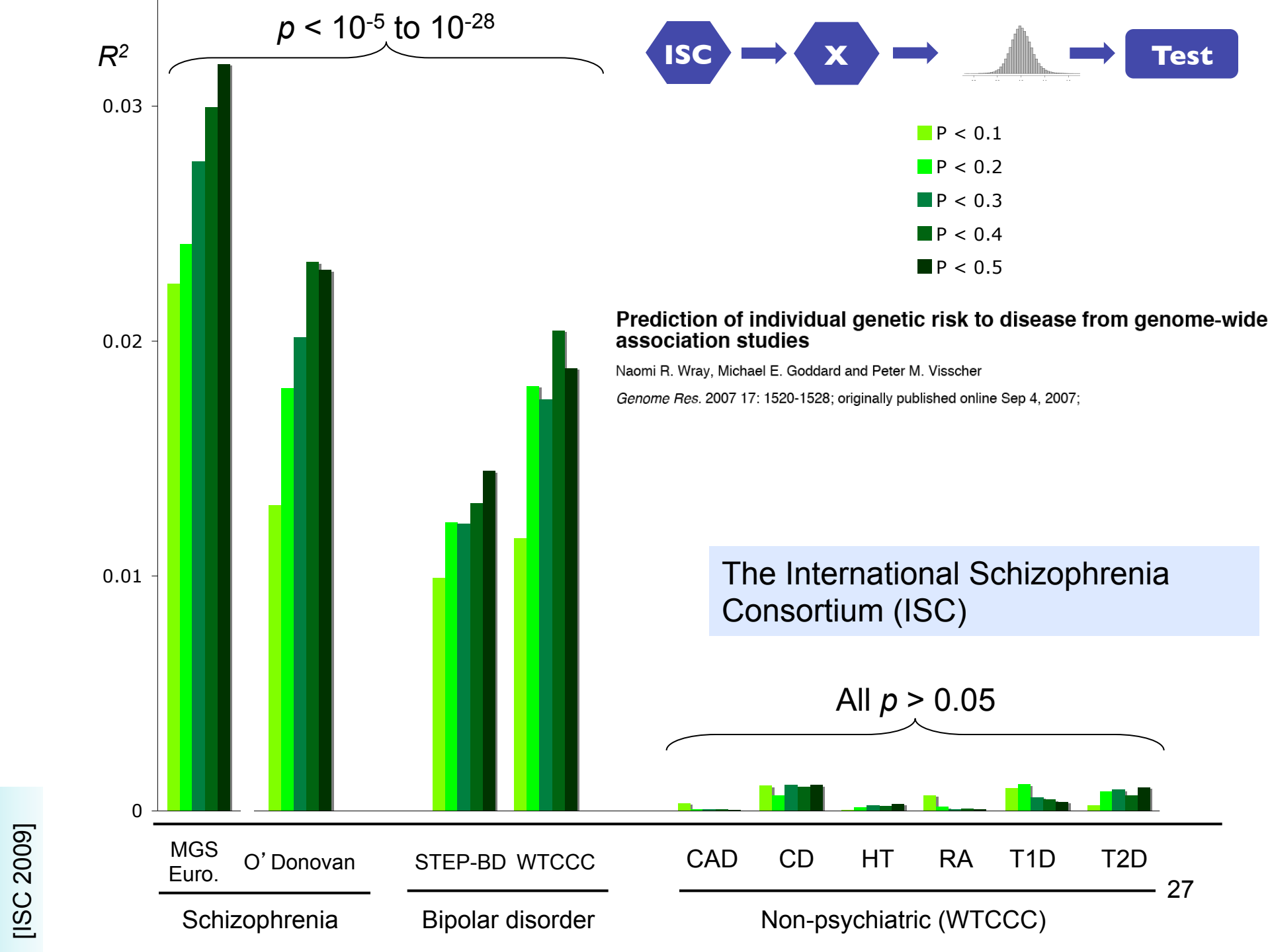



Head in the clouds


Head in the sand



Strachan & Read Human Molecular Genetics 3.


$p < 10^{-5}$ to $10^{-28}$
$R^2$
0.03
0.02
0.01
0
ISC
X
Test
P < 0.1
P < 0.2
P < 0.3
P < 0.4
P < 0.5
Prediction of individual genetic risk to disease from genome-wide association studies
Naomi R. Wray, Michael E. Goddard and Peter M. Visscher
Genome Res. 2007 17: 1520-1528; originally published online Sep 4, 2007;
The International Schizophrenia Consortium (ISC)
All $p > 0.05$
MGS Euro.
O' Donovan
STEP-BD
WTCCC
CAD
CD
HT
RA
T1D
T2D
Schizophrenia
Bipolar disorder
Non-psychiatric (WTCCC)
27
[ISC 2009]

# Measures of how well a predictor works

- "Accuracy" (animal breeding)
  - Correlation between true genome-wide genetic value and its predictor
- $R^2$ from a regression of outcome on predictor (human genetics)
- Area-under-curve from ROC analyses (disease classification)

# Limitations

Heritability determines

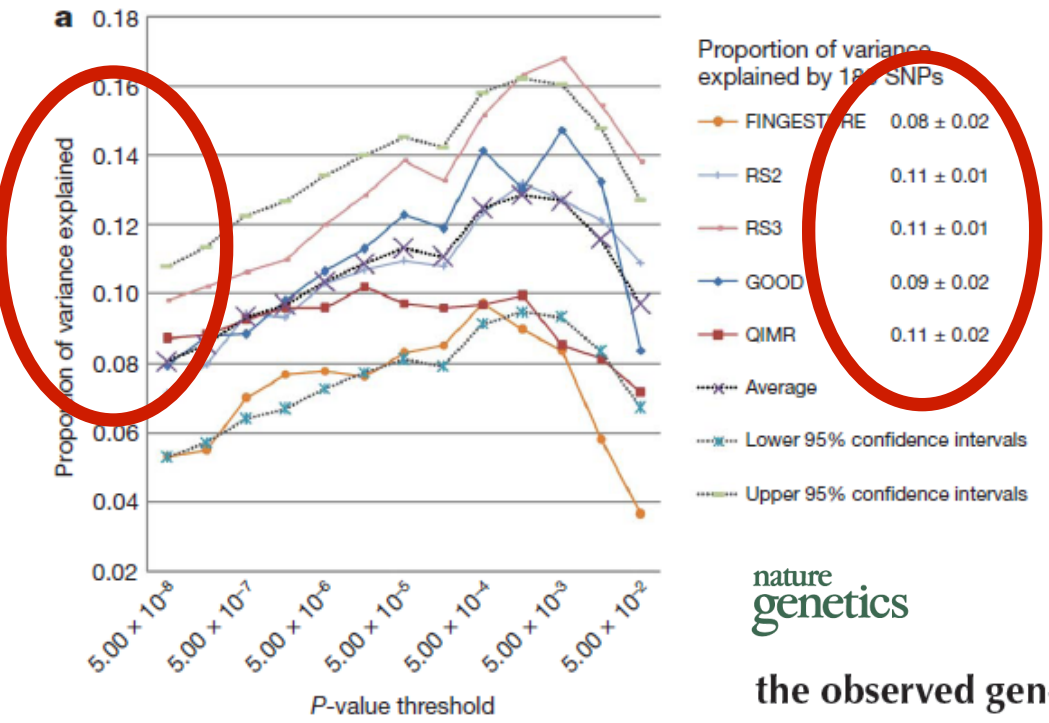
- Maximum regression $R^2$
- Precision of disease classification from genomic profile

A perfect predictor of an individual's genotypic value can be a lousy predictor of his/her phenotype

LETTER

doi:10.1038/nature09410

# Hundreds of variants clustered in genomic loci and biological pathways affect human height




10% vs 45% are not inconsistent

Estimation of variance contributed by (all) loci is not the same as prediction accuracy

unless the effect sizes are estimated without error

nature genetics

**the observed genotype data. We show tha 45% f variance can be explained by considering all SNPs si ultaneously. Thus,**

## 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]

# Conclusions

- From empirical data on height, BMI, schizophrenia and other complex traits
  - Large proportion of genetic variation is tagged by SNP chips
  - Large proportion of genetic variation is polygenic
  - Variation can be assigned to chromosomes or chromosome segments

- Much heritability is not missing, merely hiding
- GWAS experimental design & larger sample sizes will find new variants and explain more variation
- Conditional analysis can uncover more genetic variation

- Jian Yang
- Beben Benyamin
- Hong Lee
- Joseph Powell
- Allan McRae
- Anna Vinkhuyzen
- Guo-Bo Chen

***Quantitative genetics***

- Mike Goddard
- Bill Hill
- Naomi Wray

***QIMR***

- Nick Martin
- Grant Montgomery
- Dale Nyholt

***GIANT***

- Joel Hirschhorn
- Tim Frayling
- Many others

***GENEVA***

- Teri Manolio
- Bruce Weir
- Many others

***PGC-SCZ***

- Pat Sullivan
- Pamela Sklar
- Shaun Purcell
- Pablo Gejman
- Many others