34th International Workshop on Statistical Genetic Methods for Human Complex Traits - 2020

- Mike Neale (director) 
- Sarah Medland(co) 
- David Evans 
- Ben Neale 
- Hermine Maes 
- Conor Dolan 
- Lucia Colodro Conde 
- Nathan Gillespie 
- Brad Verhulst 
- Tim Bates 

- John Hewitt (host) 
- Jeff Lessem 
- Matt Keller 
- Stacey Cherny 
- Elizabeth Prom-Wormley 
- Nick Martin 
- Dorret Boomsma 
- Eveline de Zeeuw 
- Michel Nivard 
- Rob Kirkpatrick 
- Pete Hatemi 
- Mike Hunter 

The Causes of Variation

2020 International Workshop on Statistical Genetic Methods for Human Complex Traits

Boulder, CO.

Nick Martin

Queensland Institute of Medical Research

Brisbane, Australia

March 2, 2020

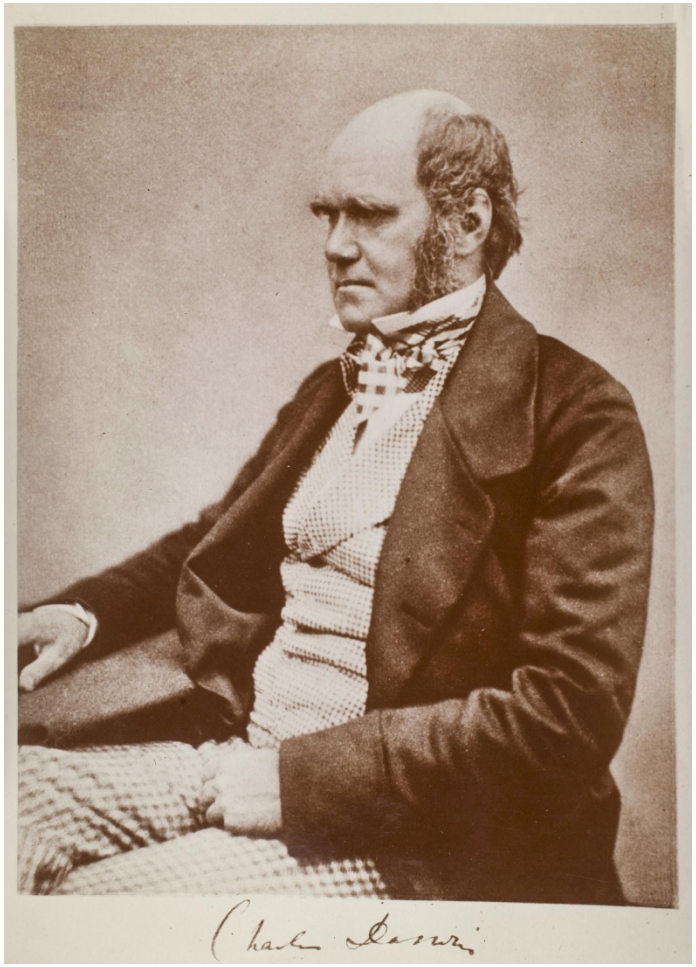
GOALS

- To outline some of the basic issues that we shall address this week
- To set some of the issues in their historical context in the story of genetics and its application to human variation.

Genetics

The Study of
Variation and Heredity

Charles Darwin (1809-1882)

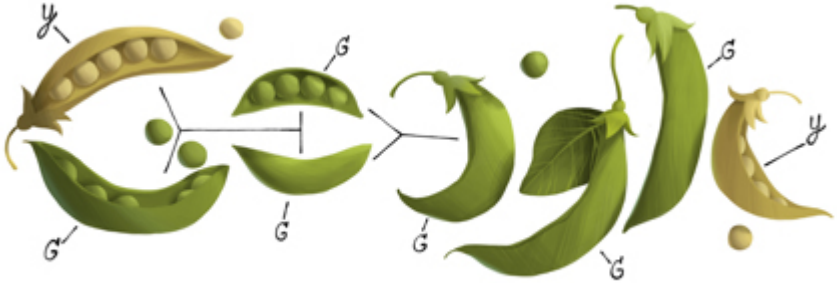


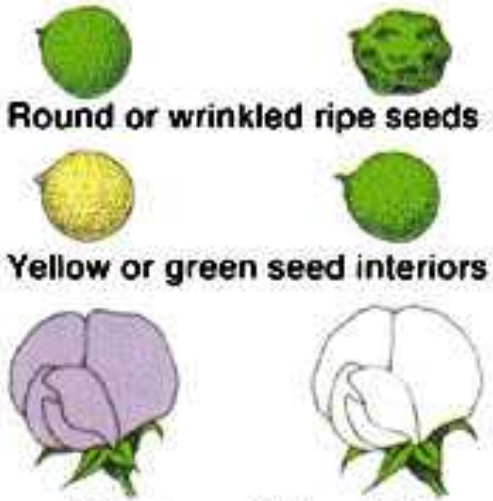
1859: On the Origin of Species

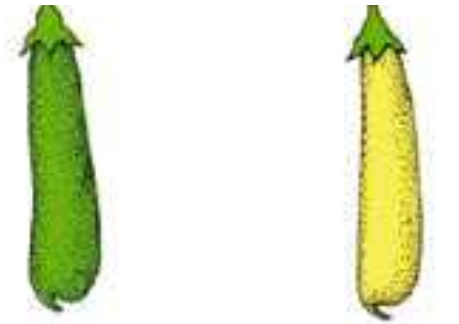
Gregor Mendel (1822-1885)

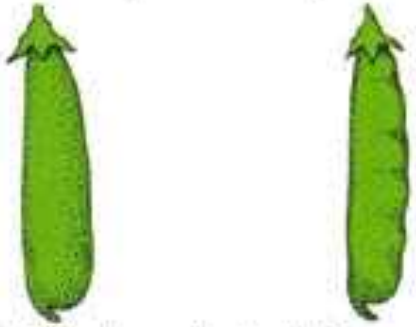
1865: Experiments in Plant Hybridisation

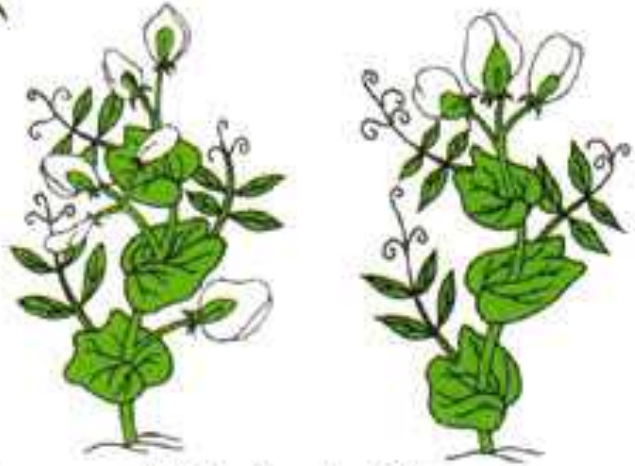
Mendel 1865 – genetics of discrete traits

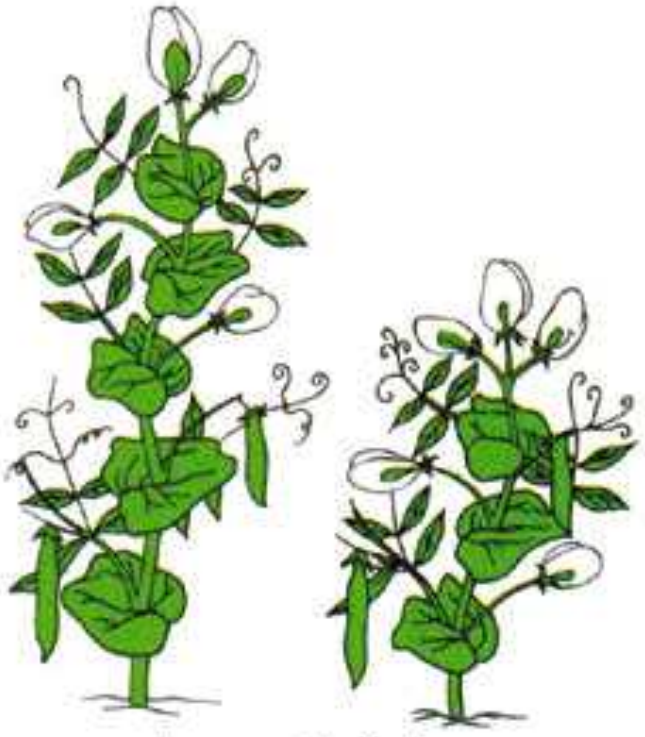




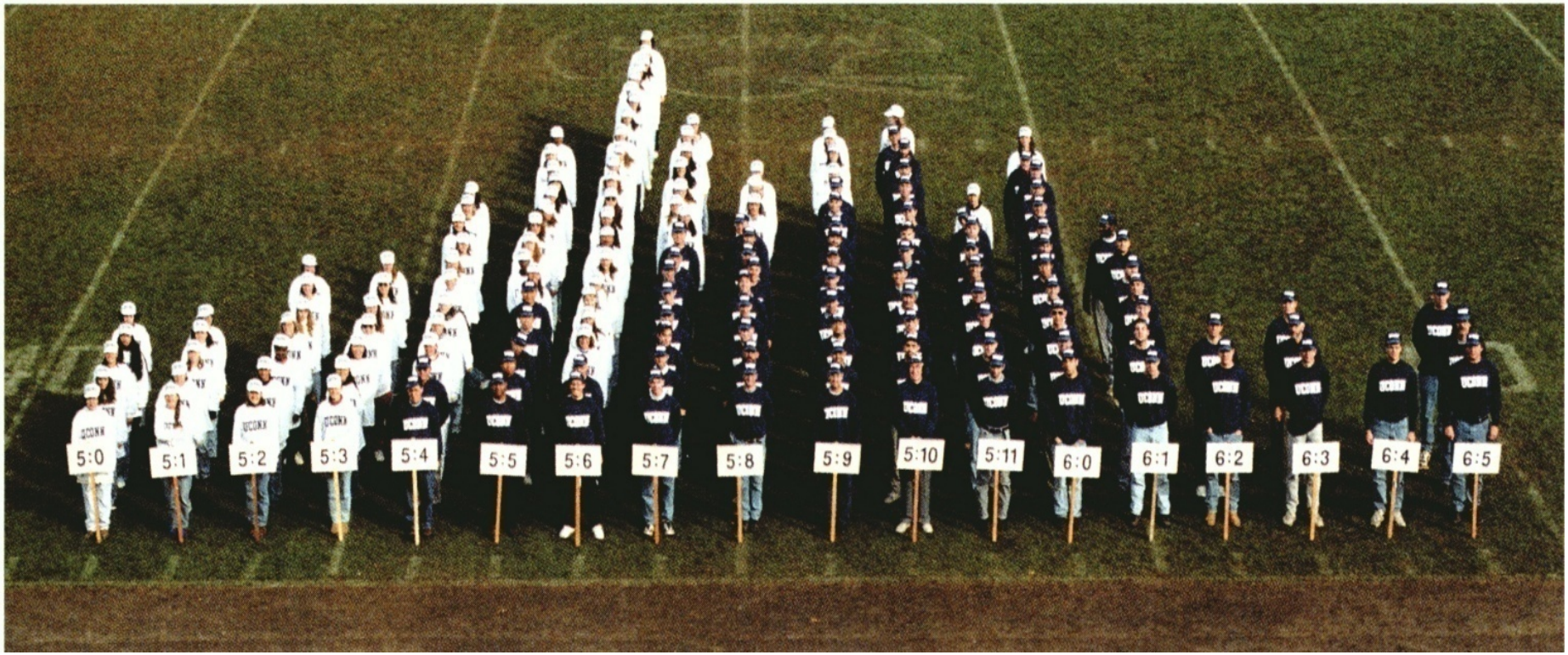








….the distribution of height



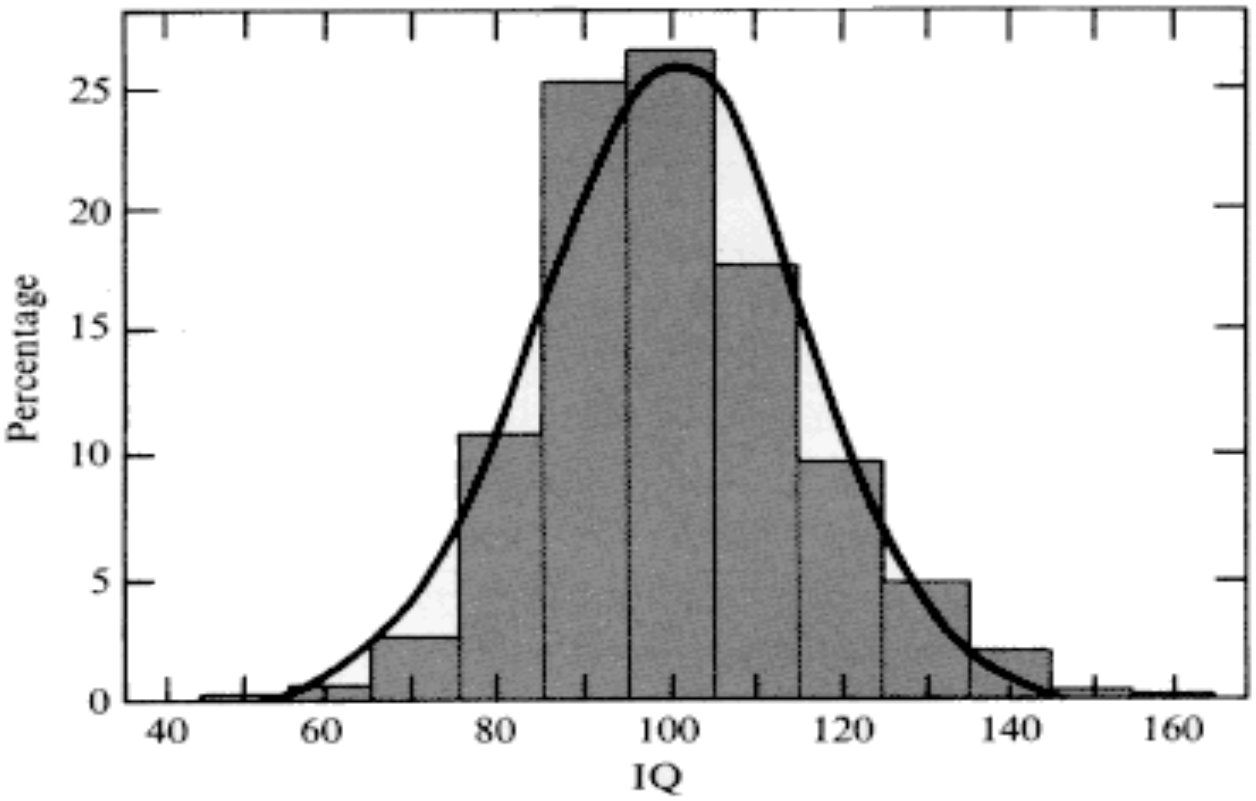
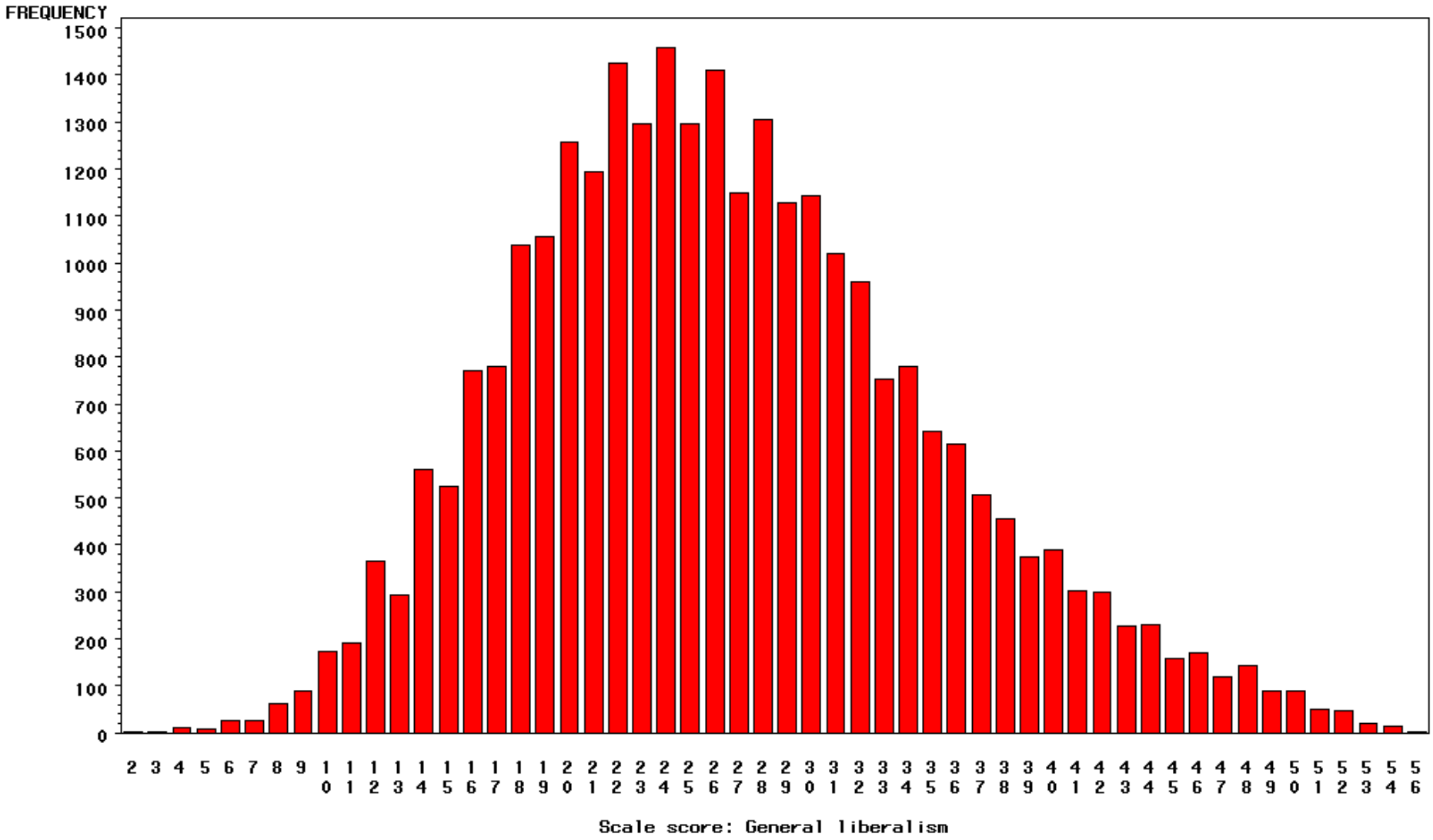


FIGURE 9.4
The distribution of IQ among the 14,963 children born in Scotland on February 1, May 1, August 1, and November 1, 1926. The shaded histogram shows the percentages of the group with IQ's in various ranges of 10 points. This grouping is artificial and is done solely for ease of representation: it does not imply any discontinuity in the values of IQ that children can show. The continuous curve shows the ideal distribution calculated from the observations and representing the statistical population of which the children actually observed are regarded as forming a sample. (Data from MacMeekan; from Mather 1964.)

“Liberalism”


FREQUENCY
1500
1400
1300
1200
1100
1000
900
800
700
600
500
400
300
200
100
0
2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 56
Scale score: General liberalism

XV.—**The Correlation between Relatives on the Supposition of Mendelian Inheritance.** By **R. A. Fisher**, B.A. *Communicated by* Professor J. Arthur Thomson. (With Four Figures in Text.)

(MS. received June 15, 1918. Read July 8, 1918. Issued separately October 1, 1918.)

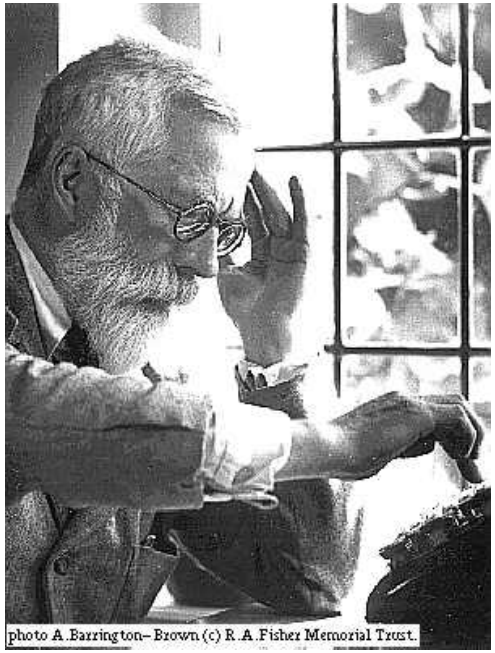
CONTENTS.

Several attempts have already been made to interpret the well-established results of biometry in accordance with the Mendelian scheme of inheritance. It is here attempted to ascertain the biometrical properties of a population of a more general type than has hitherto been examined, inheritance in which follows this scheme. It is hoped that in this way it will be possible to make a more exact analysis of the causes of human variability. The great body of available statistics show us that the deviations of a human measurement from its mean follow very closely the Normal Law of Errors, and, therefore, that the variability may be uniformly measured by the standard deviation corresponding to the square root of the mean square error. When there are two independent causes of variability capable of producing in an otherwise uniform population distributions with standard deviations σ_1 and σ_2, it is found that the distribution, when both causes act together, has a standard deviation $\sqrt{\sigma_1^2+\sigma_2^2}$. It is therefore desirable in analysing the causes of variability to deal with the square of the standard deviation as the measure of variability. We shall term this quantity the Variance of the normal population to which it refers, and we may now ascribe to the constituent causes fractions or percentages of the total variance which they together produce. It

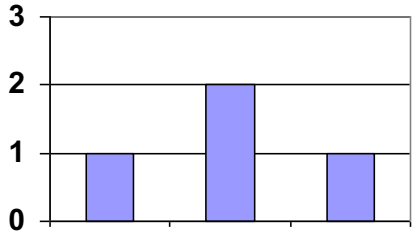
RA Fisher (1918). *Transactions of the Royal Society of Edinburgh* **52**: 399-433.

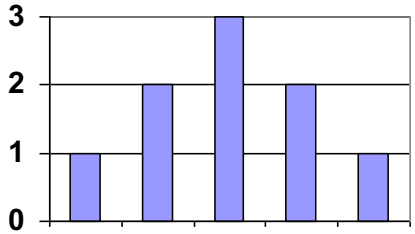
R.A. Fisher, 1918

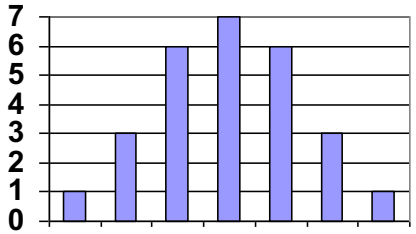
The explanation of quantitative inheritance in Mendelian terms

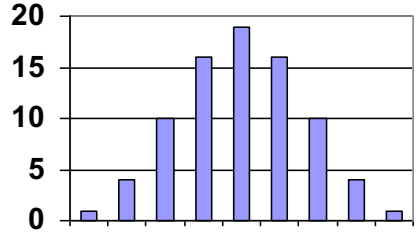


1 Gene	2 Genes	3 Genes	4 Genes
→ 3 Genotypes	→ 9 Genotypes	→ 27 Genotypes	→ 81 Genotypes
→ 3 Phenotypes	→ 5 Phenotypes	→ 7 Phenotypes	→ 9 Phenotypes





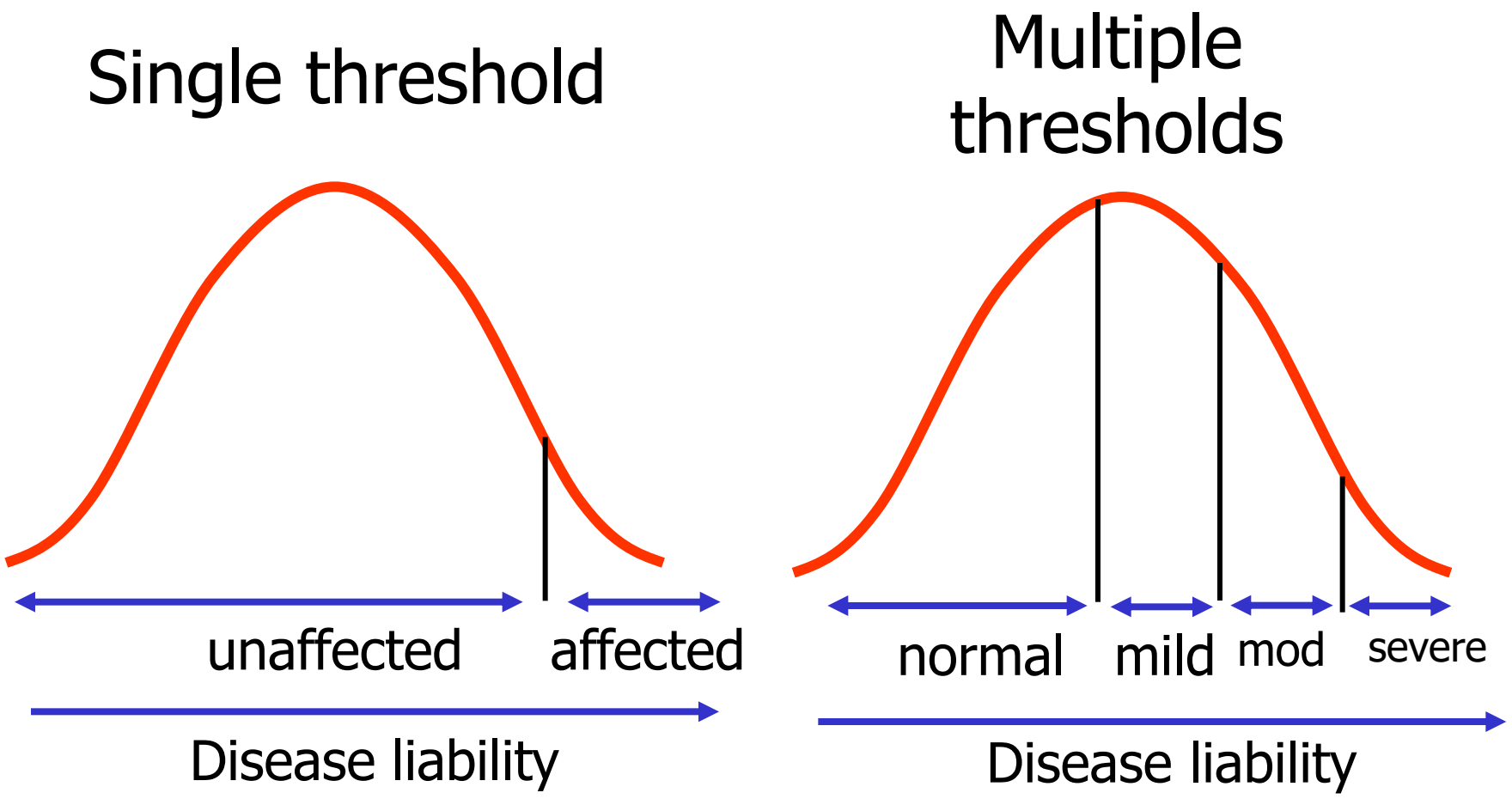




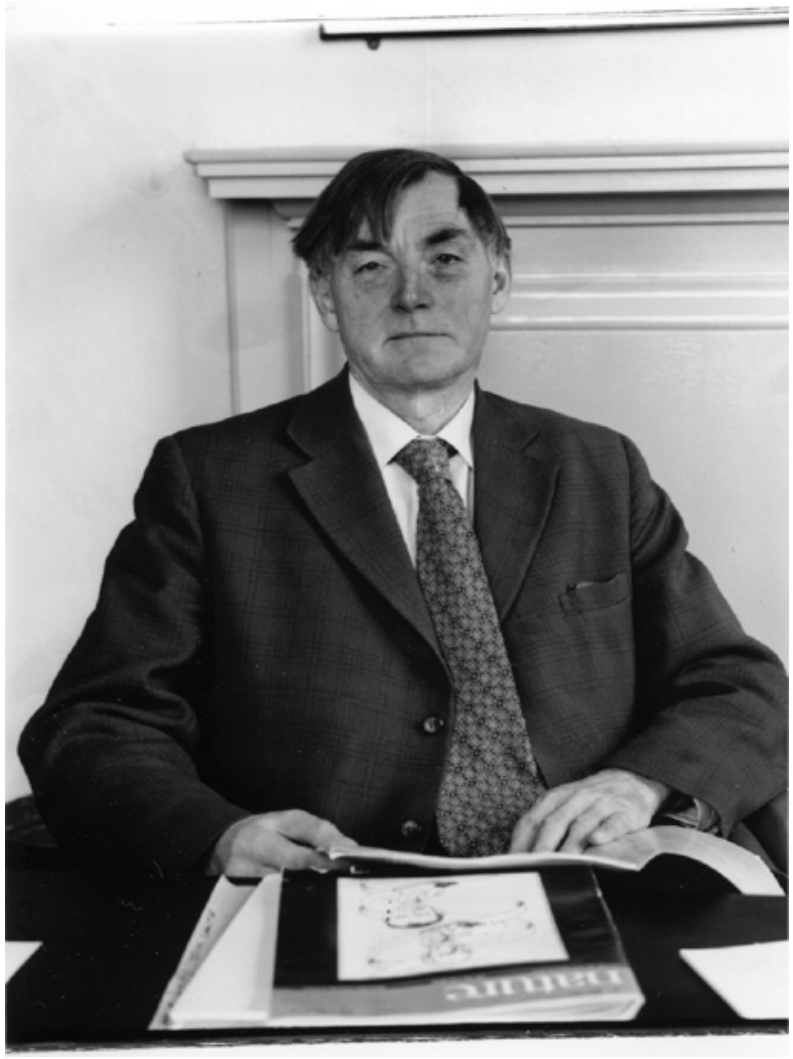
Fisher (1918): Basic Ideas

- Continuous variation caused by lots of genes ("polygenic inheritance")
- Each gene followed Mendel's laws
- Environment smoothed out genetic differences
- Genes may show different degrees of "dominance"
- Genes may have many forms ("multiple alleles")
- Mating may not be random ("assortative mating")
- Showed that correlations obtained by e.g. Pearson and Lee were explained well by polygenic inheritance
- Led to "Biometrical Genetics" (Mather, Jinks etc.)

Multifactorial Threshold Model of Disease – normally distributed "liability"



Douglas Scott Falconer, FRS, FRSE (1913-2004)



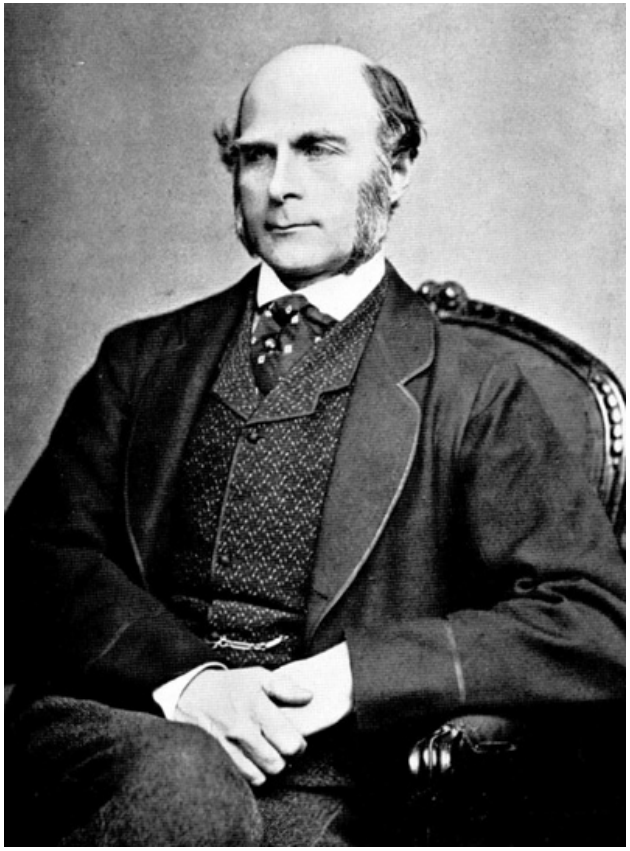
1965 Inheritance of liability to certain diseases estimated from incidence among relatives., *Ann. Hum. Genet.***29:**51ff.

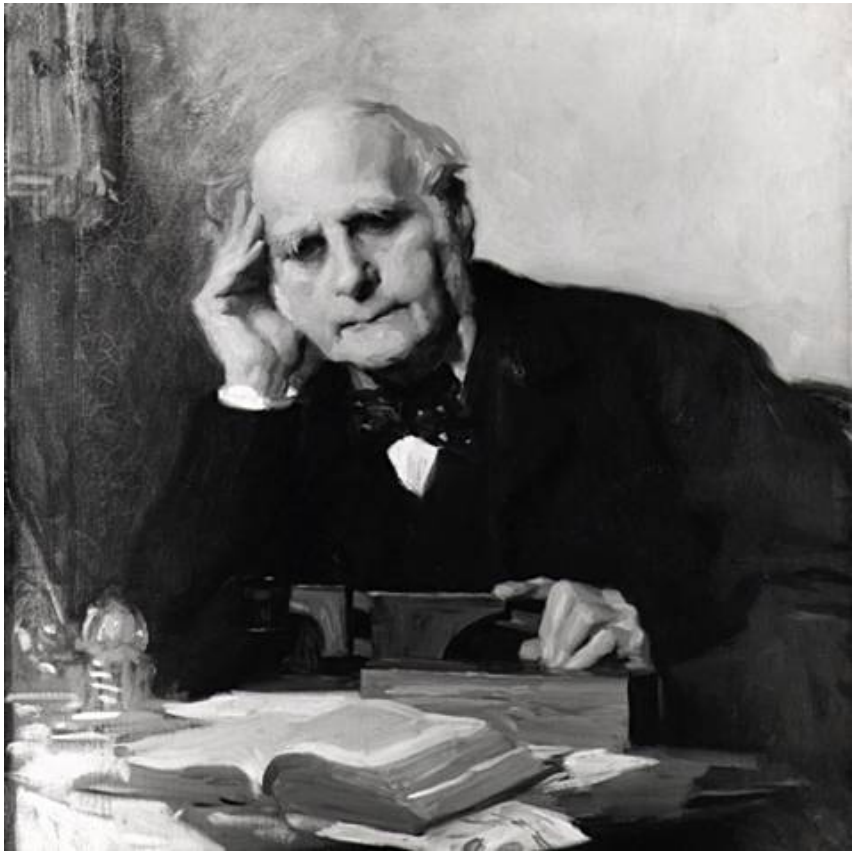
1960 *Introduction to Quantitative Genetics.* Edinburgh: Oliver and Boyd.

The (Really!) BIG Problem

Families are a mixture of genetic and social factors

Francis Galton (1822-1911)





1869: Hereditary Genius

1883: Inquiries into Human Faculty and its Development

1884-5: Anthropometic Laboratory at “National Health Exhibition”

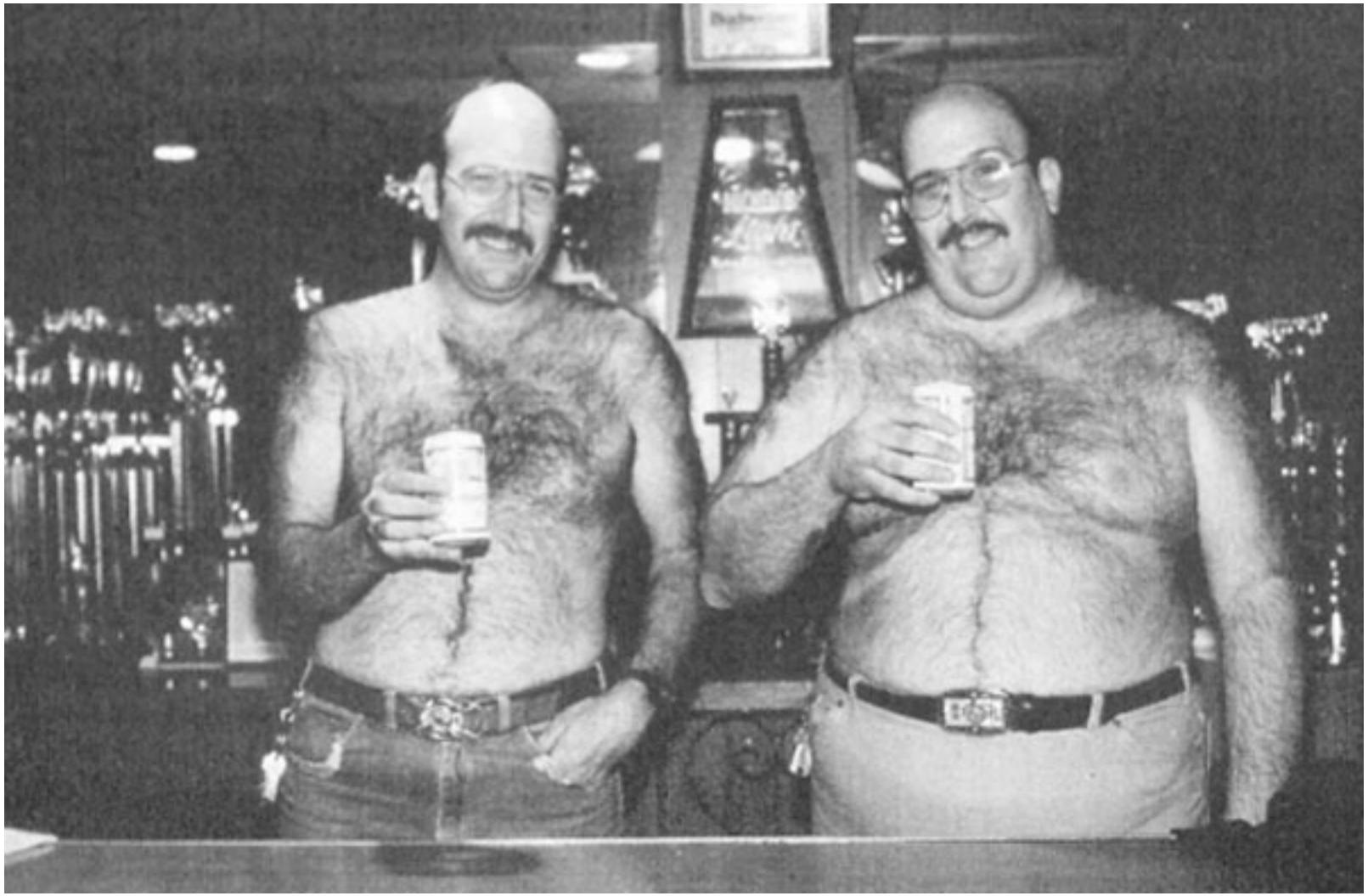
Galton's Solution:

Twins

(Though Augustine may have got there first – 5th cent.)

One (?ideal) solution

Twins separated at birth



MZ twins reared apart - note the same way of supporting their cans of beer

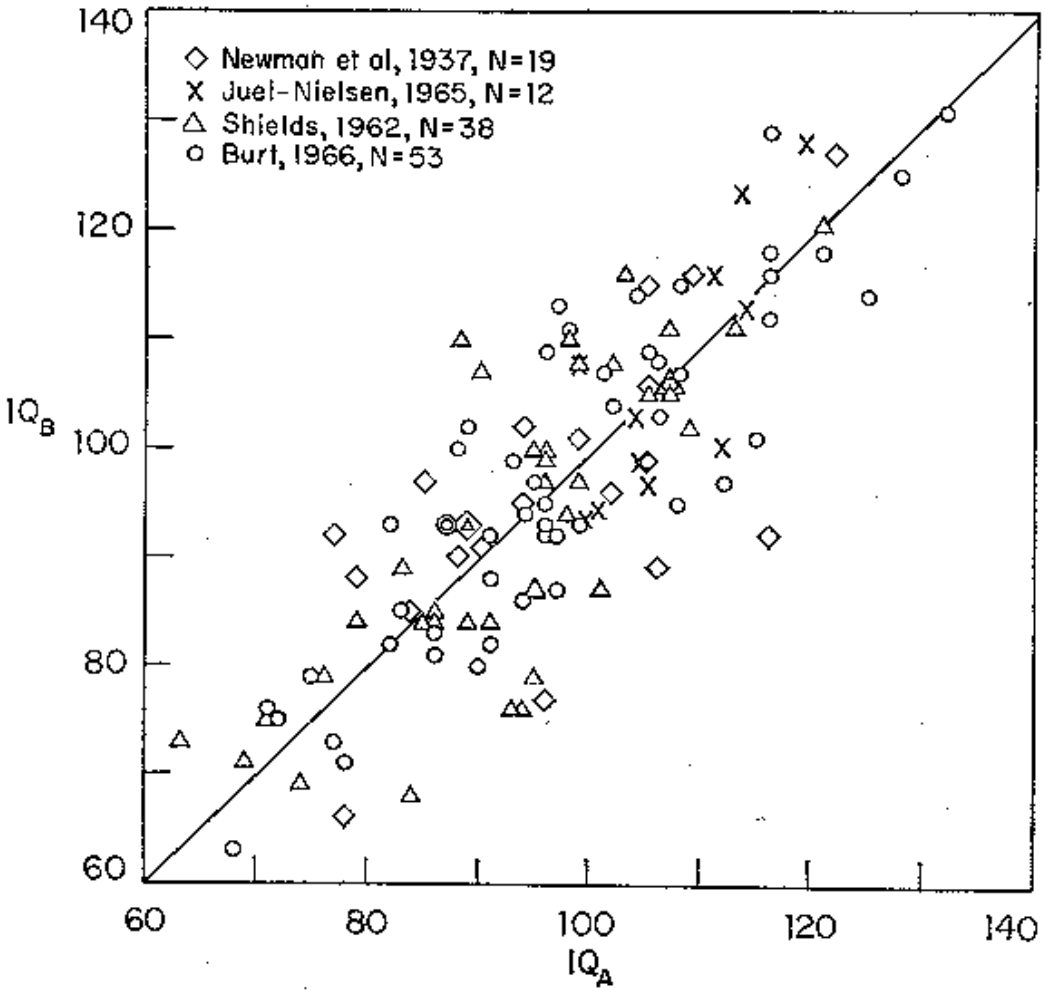


FIGURE 2. *Scatter diagram showing correlation between IQs of 122 sets of co-twins (A and B assigned at random). The obtained intra-class correlation (r_i) is 0·82. The diagonal line represents perfect correlation ($r_i = 1·00$).*

But separated MZs are rare

An easier alternative:

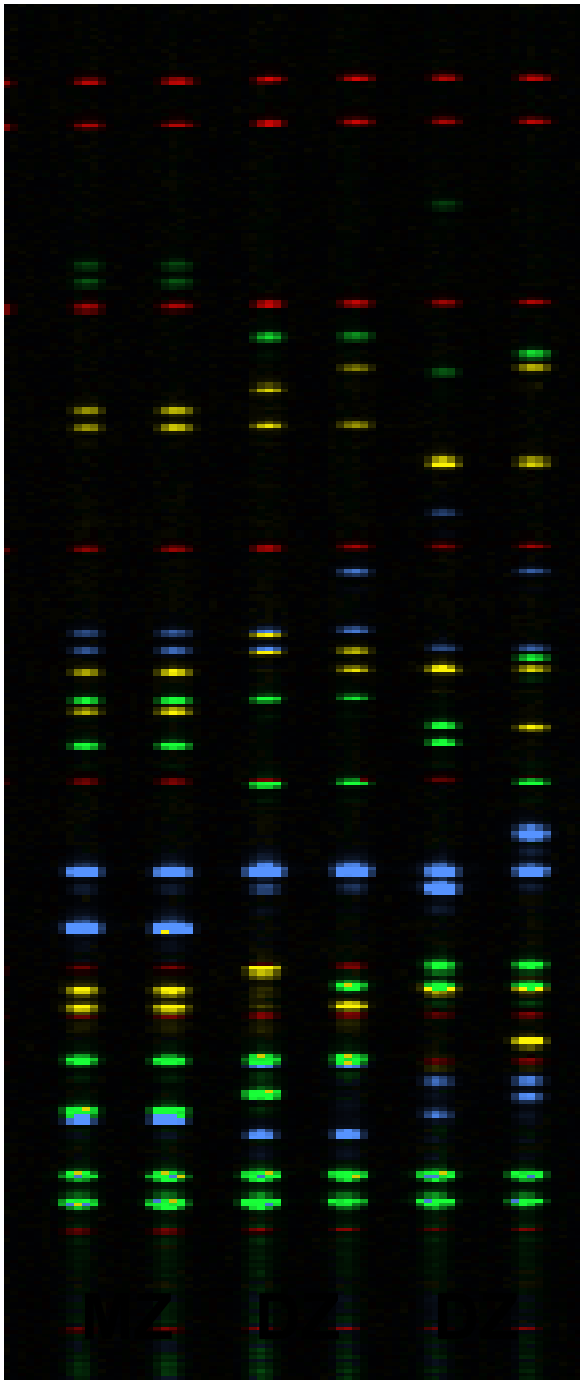
Identical and non-identical twins reared together
Galton (again!)……Or was it ?

Early Research on Human Genetics Using the Twin Method: Who Really Invented the Method?

Oliver Mayo
CSIRO Livestock Industries, Australia

Twin Research and Human Genetics 12: 237–245, 2009

Poll, H. (1914). Über Zwillingsforschung als Hilfsmittel menschlicher Erbkunde. *Zeitschrift für Ethnologie 46,* 87–105.

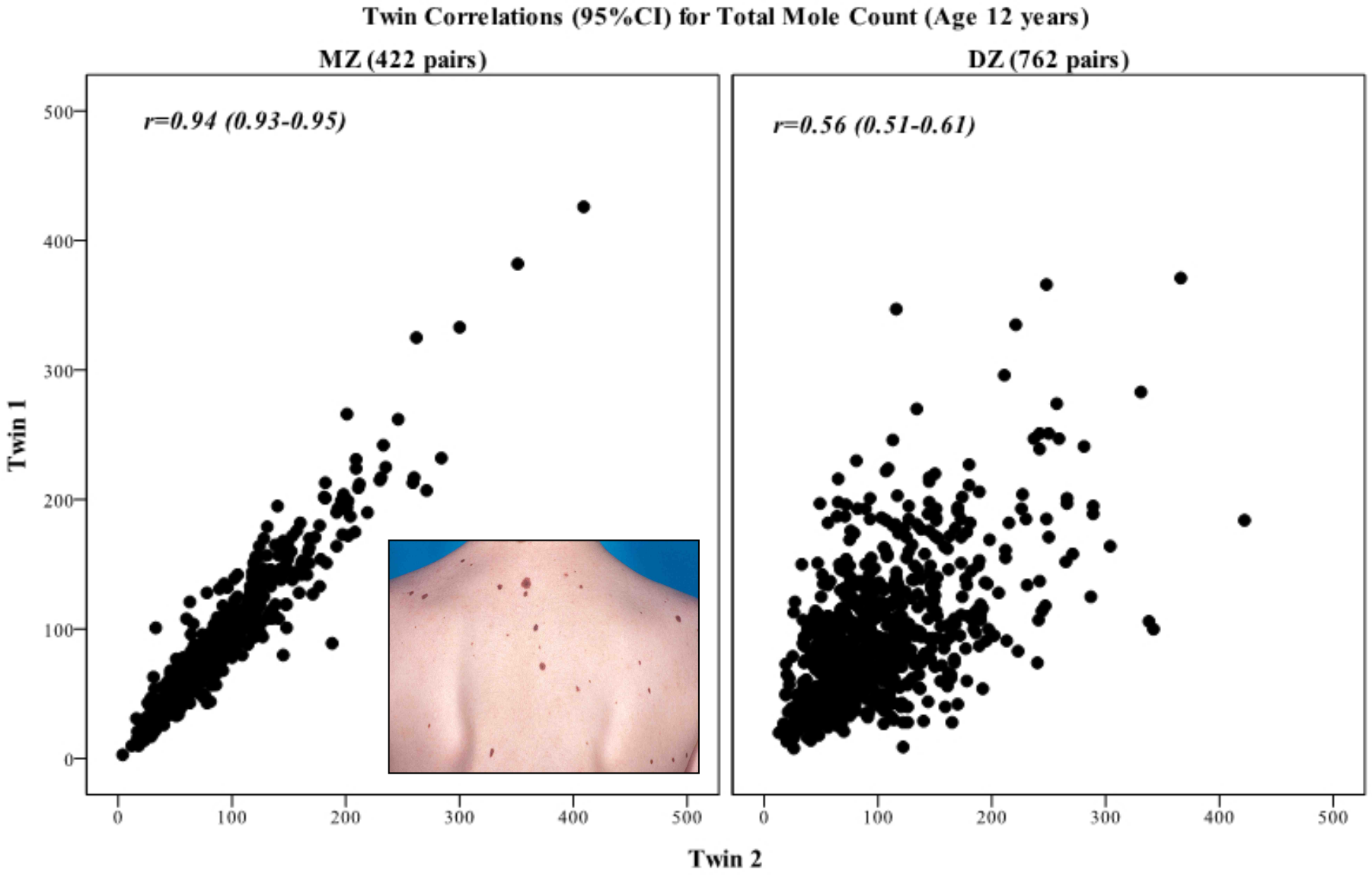


Identity at marker loci - except for rare mutation

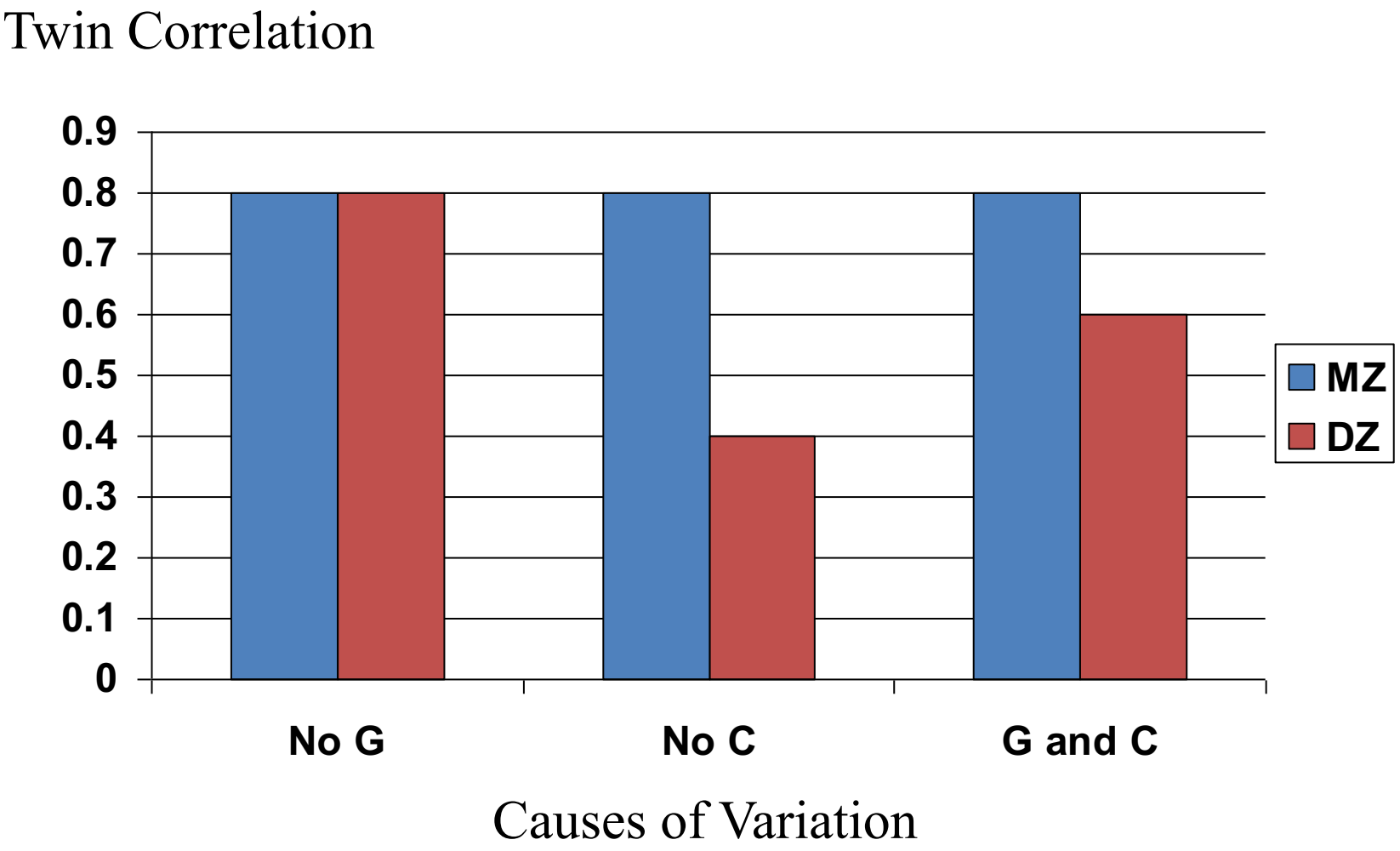
MZ and DZ twins: determining zygosity using ABI Profiler™ genotyping

(9 STR markers + sex)

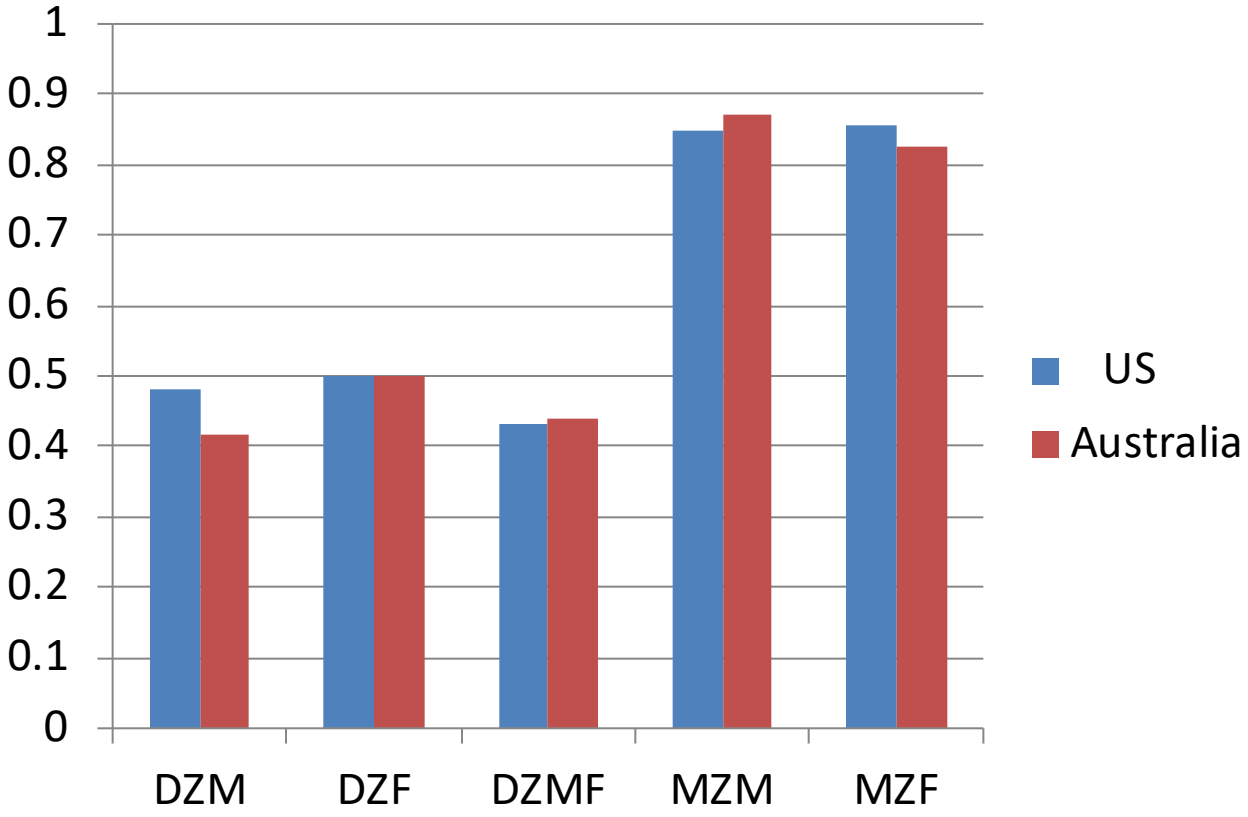
Twin correlations for total mole count



Three scenarios



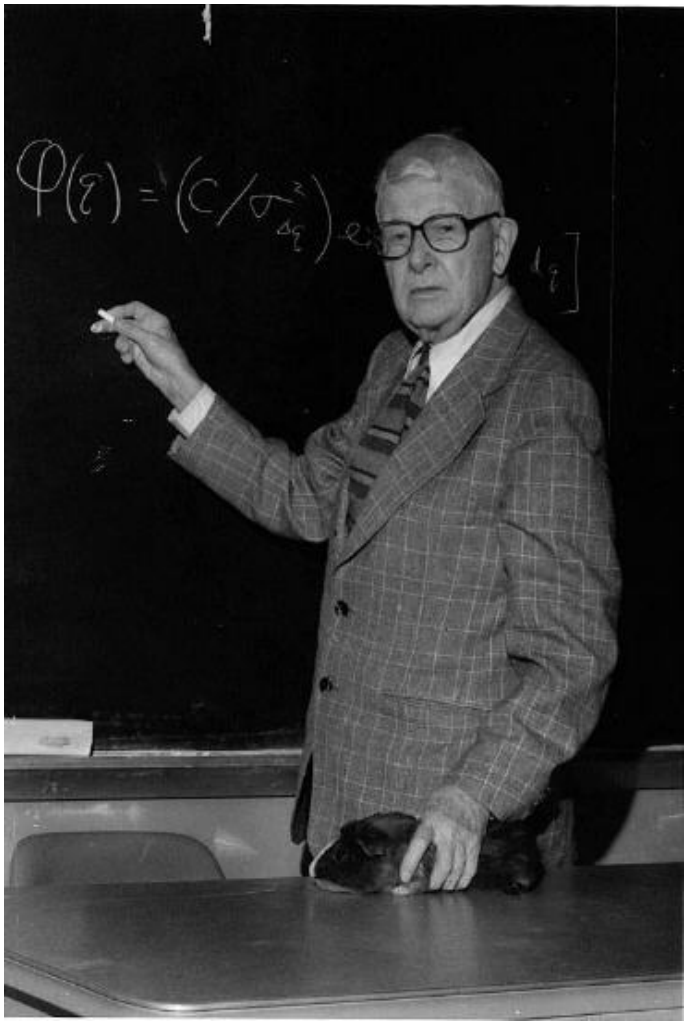
Twin Correlations for Adult Stature
(Virginia 30,000 and Australia 22,000)



Structural equation modeling

- Both continuous and categorical variables
- Systematic approach to hypothesis testing
- Tests of significance
- Can be extended to:
  - More complex questions
  - Multiple variables
  - Other relatives

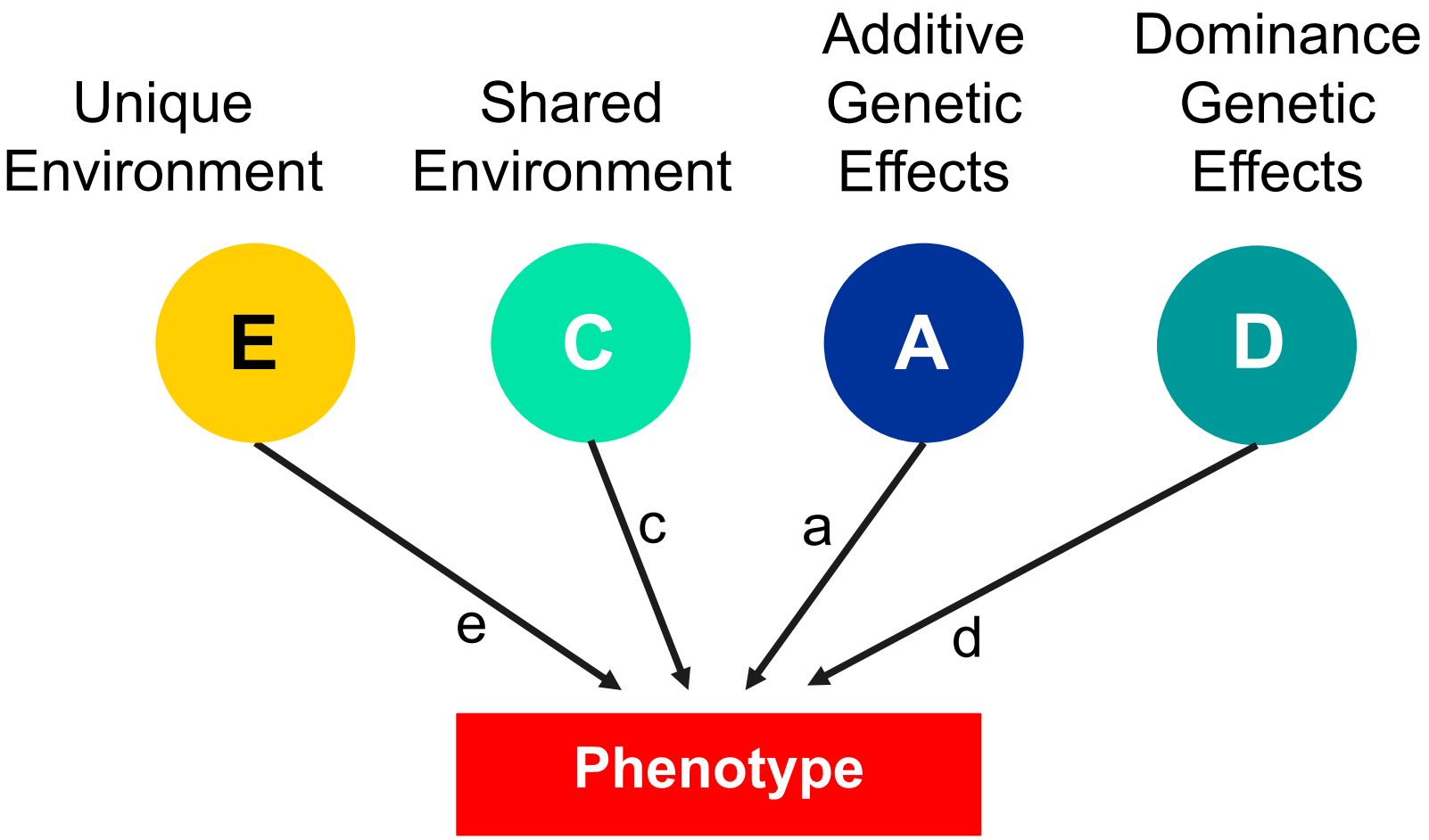
Sewall Wright (1889-1988)



1984. Evolution and the Genetics of Populations:
Genetics and Biometric Foundations (4 vols.)

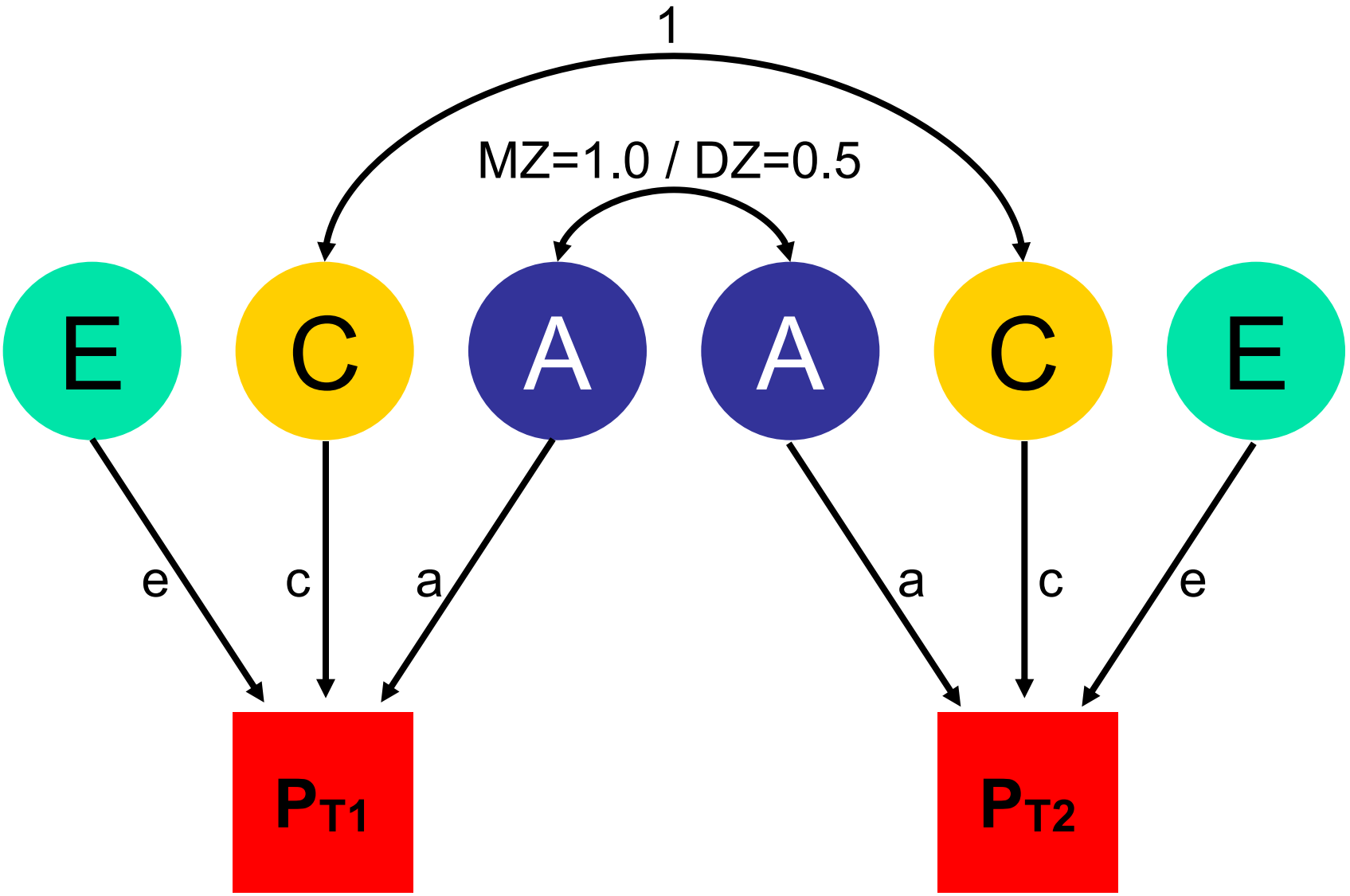
1934. The method of path coefficients. *Ann. Math. Staiist.* **5:** 161-215

Variance components

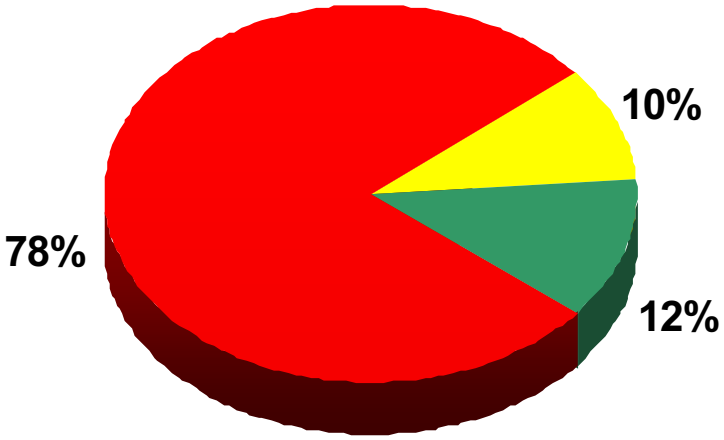


$$P = eE + aA + cC + dD$$

ACE Model for twin data



Sources of variation in exam (QCST) results



Additive genetic

Shared environment

Non-shared environment

Problems with the ACE model (in the classical twin design)

- C and D are negatively confounded and a model with ACDE is not identified – can only fit one or the other of C/D-
  - If rmz = rdz > pure C : CE model fits best
  - If rmz = 2rdz > pure A : AE model fits best
  - If rmz < 2rdz > A plus C : ACE model fits best
  - If rmz > 2rdz > A plus D : ADE model fits best
  - If rmz = rdz=0 > all E : no familial resemblance

Other problems (2)

- C can be inflated by age, sex differences; also batch effects; correct for these using Means model in Mx
- Low power to detect D
- D subsumes epistasis and other non-additive genetic effects
- E includes measurement error – value of test-retest data

Other problems (3)

- C can be inflated by phenotypic assortative mating (like marrying like)
  - Important to collect spouse data to estimate marital correlation
    - Height ~0.2
    - Education, smoking, drinking, politics ~0.6

CAN ACCOUNT FOR SOME/MOST/ALL OF "C" ESTIMATED FROM ACE MODEL IN TWIN DATA

How twin studies changed research agendas

1. Autism – “caused by cold mothers”

   10/11 MZ pairs concordant, 2/11 DZs concordant

2. ADHD – “caused by food dyes”

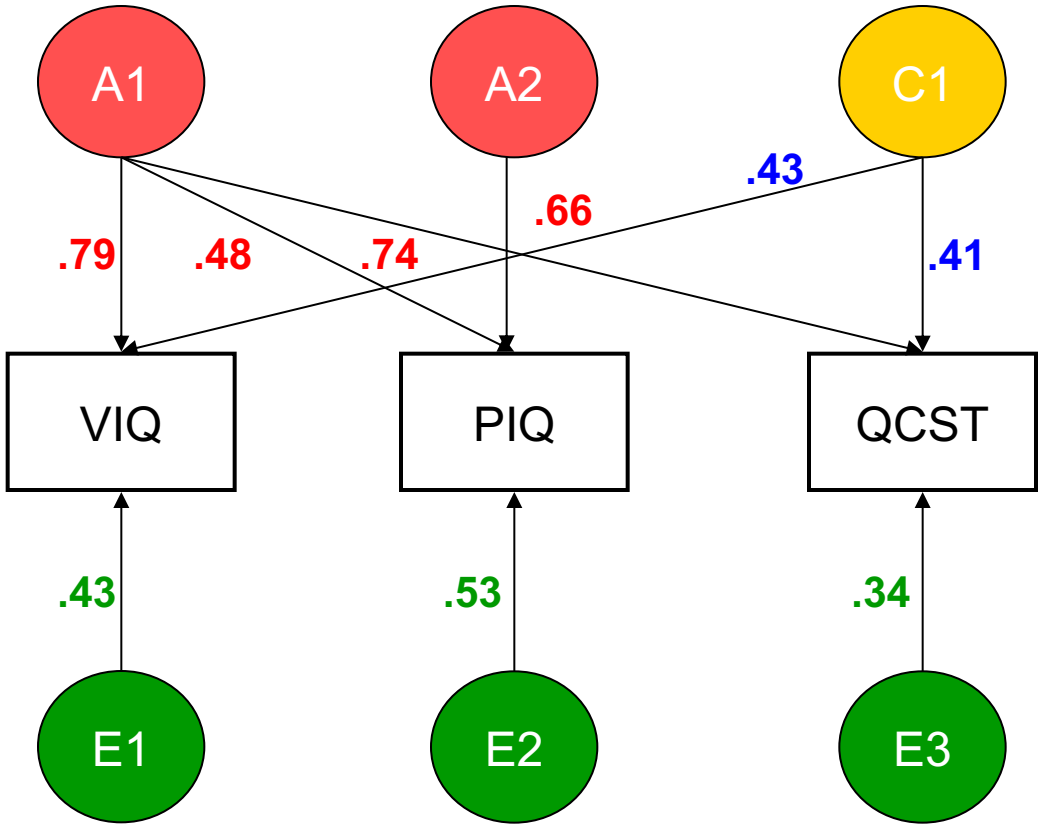
   Twin studies found h^2 ~0.8

3. Multiple sclerosis - “caused by a virus”

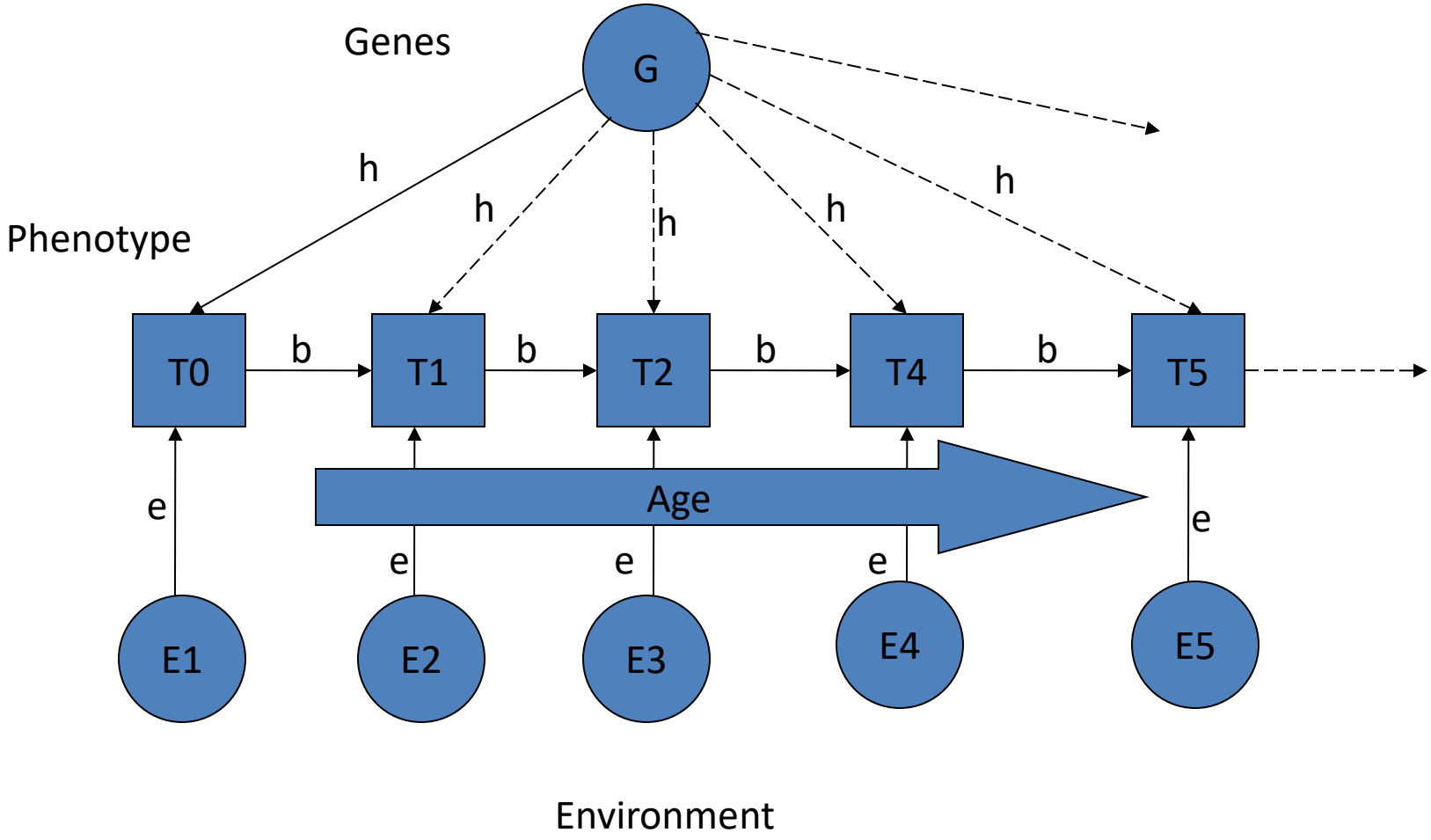
   MZ concordance 26%, DZ concordance 2%

Martin, Boomsma & Machin (1997) *Nature Genetics* 17: 387

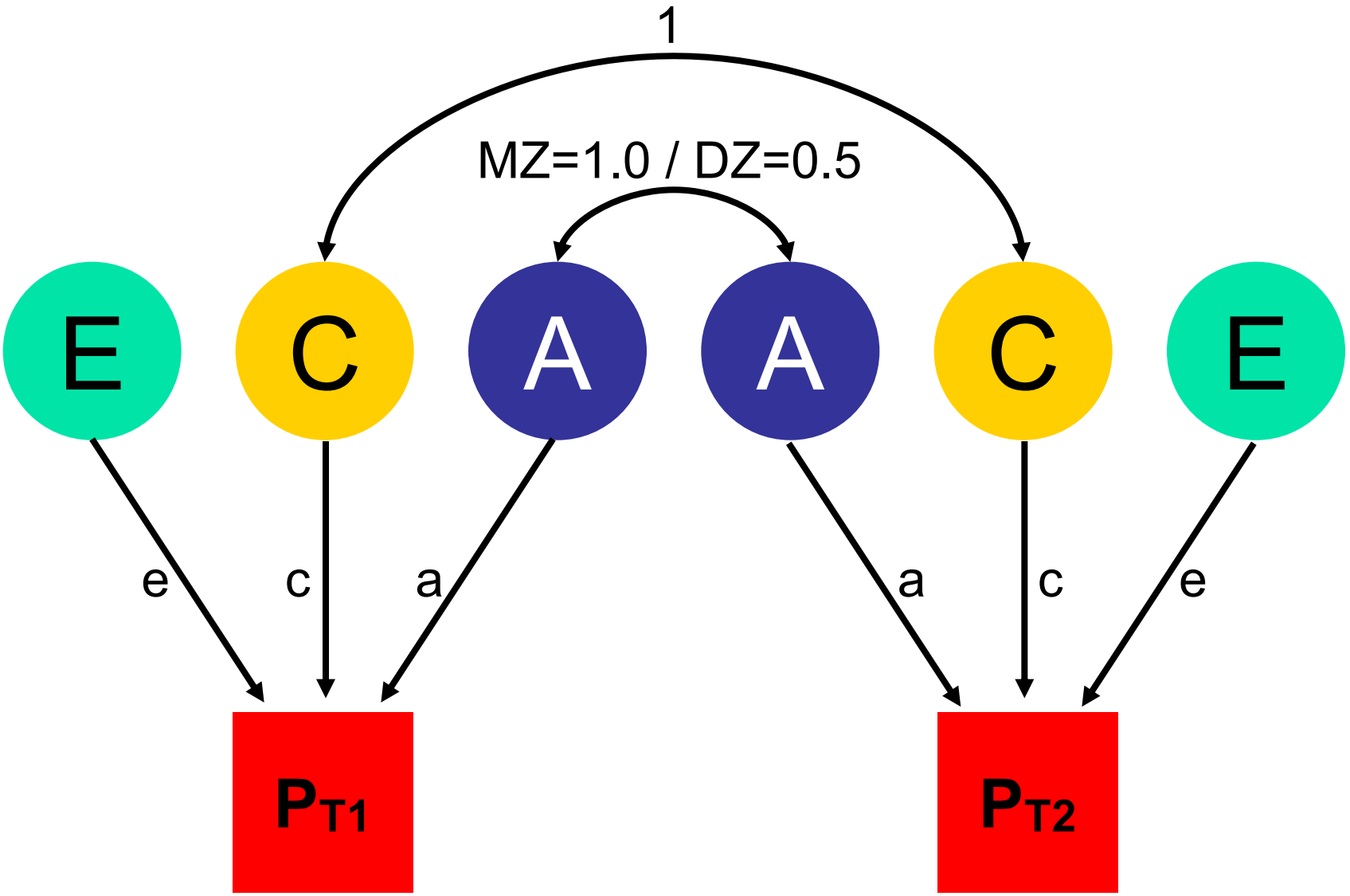
Are there common genetic and environmental factors influencing Verbal IQ, Performance IQ and QCST results?



Genetic variation in **developmental change**: time series with common genes and time-specific environmental "innovations"



ACE Model for twin data



Classical twin design revisited: Heritability estimation without MZ twins

Why do we use the average sib value of $r_a = 0.5$ when we can estimate the (almost) exact values for each sib pair from genetic marker data ?

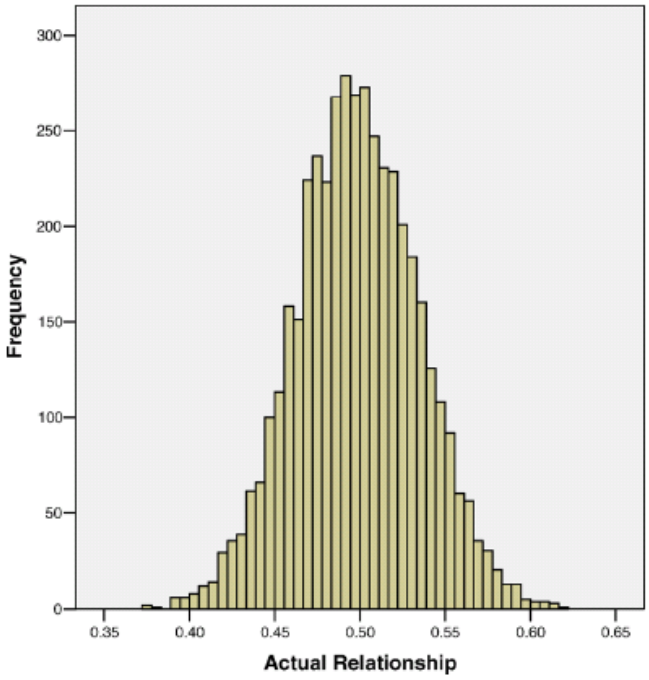


Figure 1. Empirical Distribution of Actual Additive Genetic Relationships of 4,401 Quasi-Independent Pairs of Full Sibs

Histogram of the genome-wide additive genetic relationships of full-sib pairs estimated from genetic markers.

DOI: 10.1371/journal.pgen.0020041.g001

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

Genetic Epidemiology Group, Queensland Institute of Medical Research, Brisbane, Australia

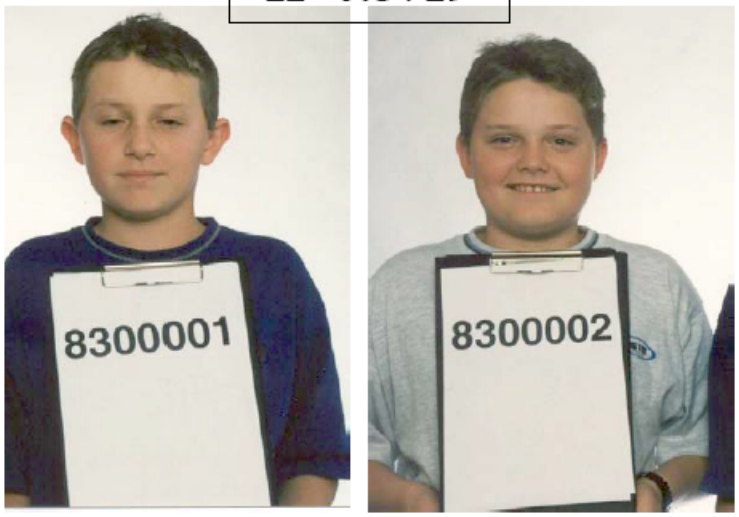
Table 2. ML Estimates of Heritability of Height from Genome-Wide IBD Sharing between Sib Pairs

Data	Model	Estimates (95% CI)	
		f^2	h^2
Adolescents (n = 931)	FAE	0.00 (0.00–0.43)	0.80 (0.00–0.90)
	FE	0.40 (0.34–0.45)	
Adults (n = 2,444)	FAE	0.00 (0.00–0.18)	0.80 (0.43–0.86)
	FE	0.39 (0.36–0.43)	
Combined (n = 3,375)	FAE	0.00 (0.00–0.17)	0.80 (0.46–0.85)
	FE	0.39 (0.36–0.42)	

8188001,02
H=0.5677

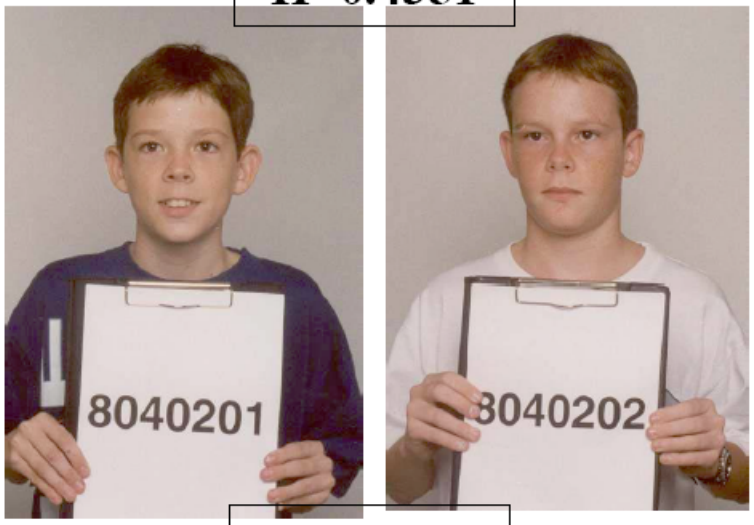
8473001,02
H=0.5577

8300001,02
H=0.5719



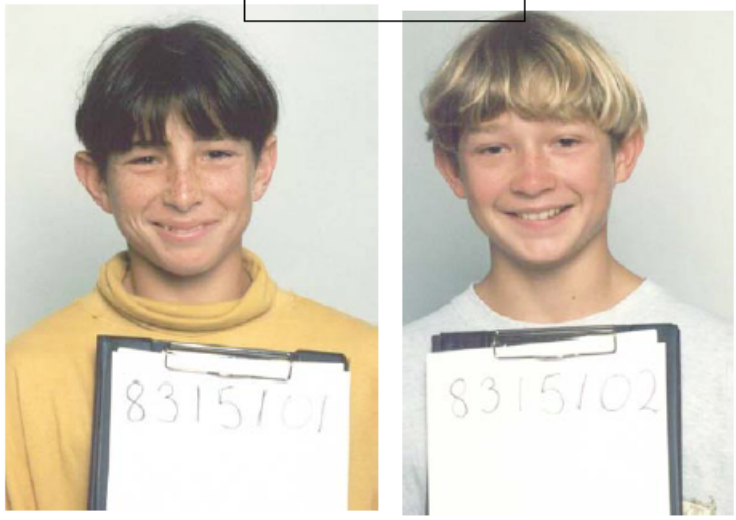
8582601,02
H=0.5640

8040201,02
H=0.4351

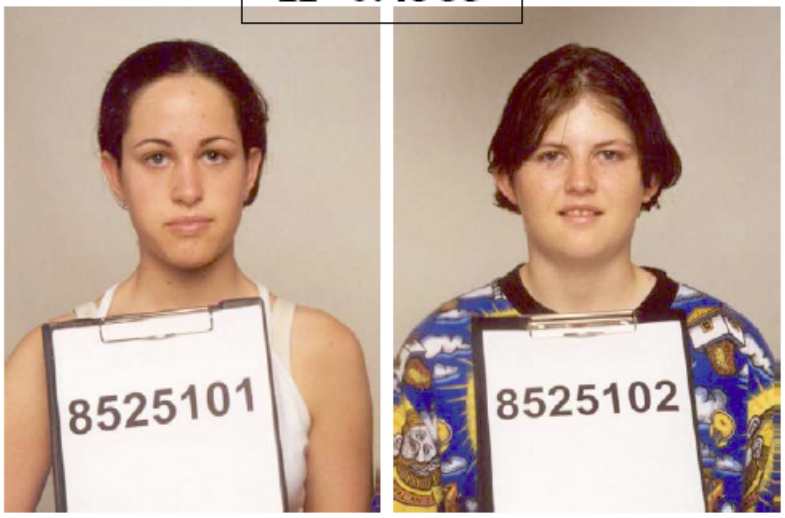


8069101,02
H=0.4291

8315101,02
H=0.4320

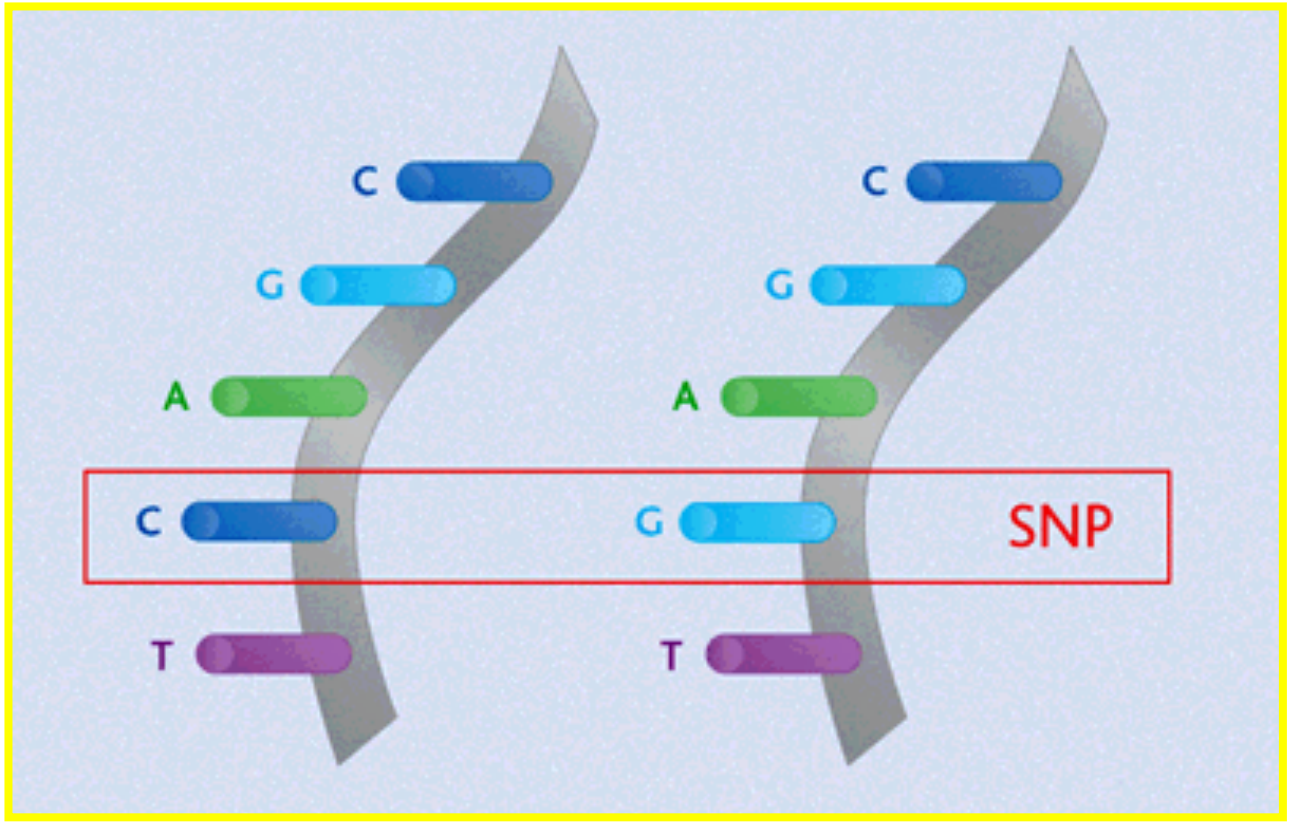


8525101,02
H=0.4385

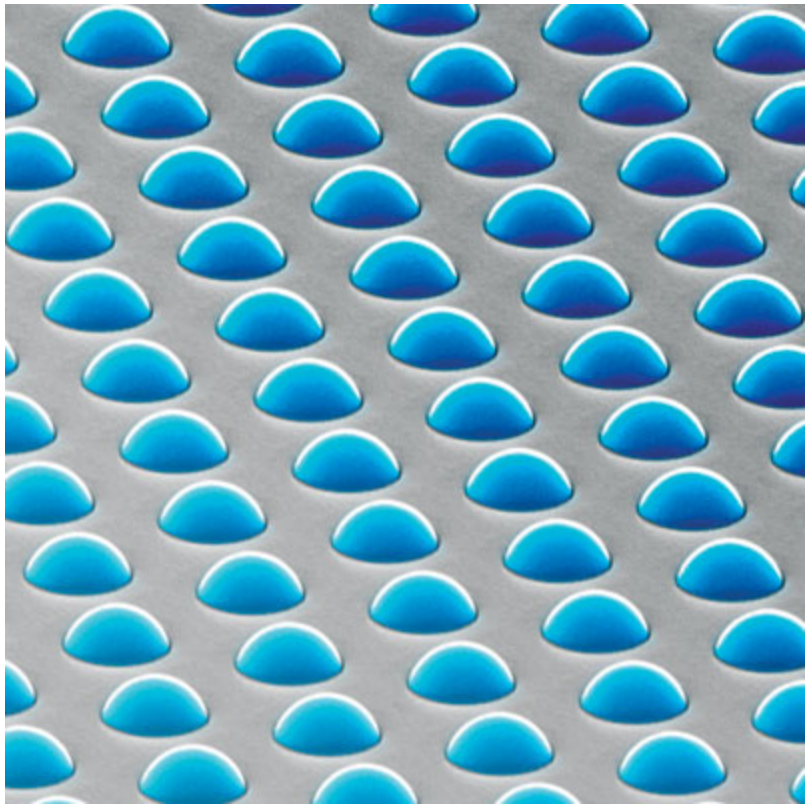


Finding the genes - association

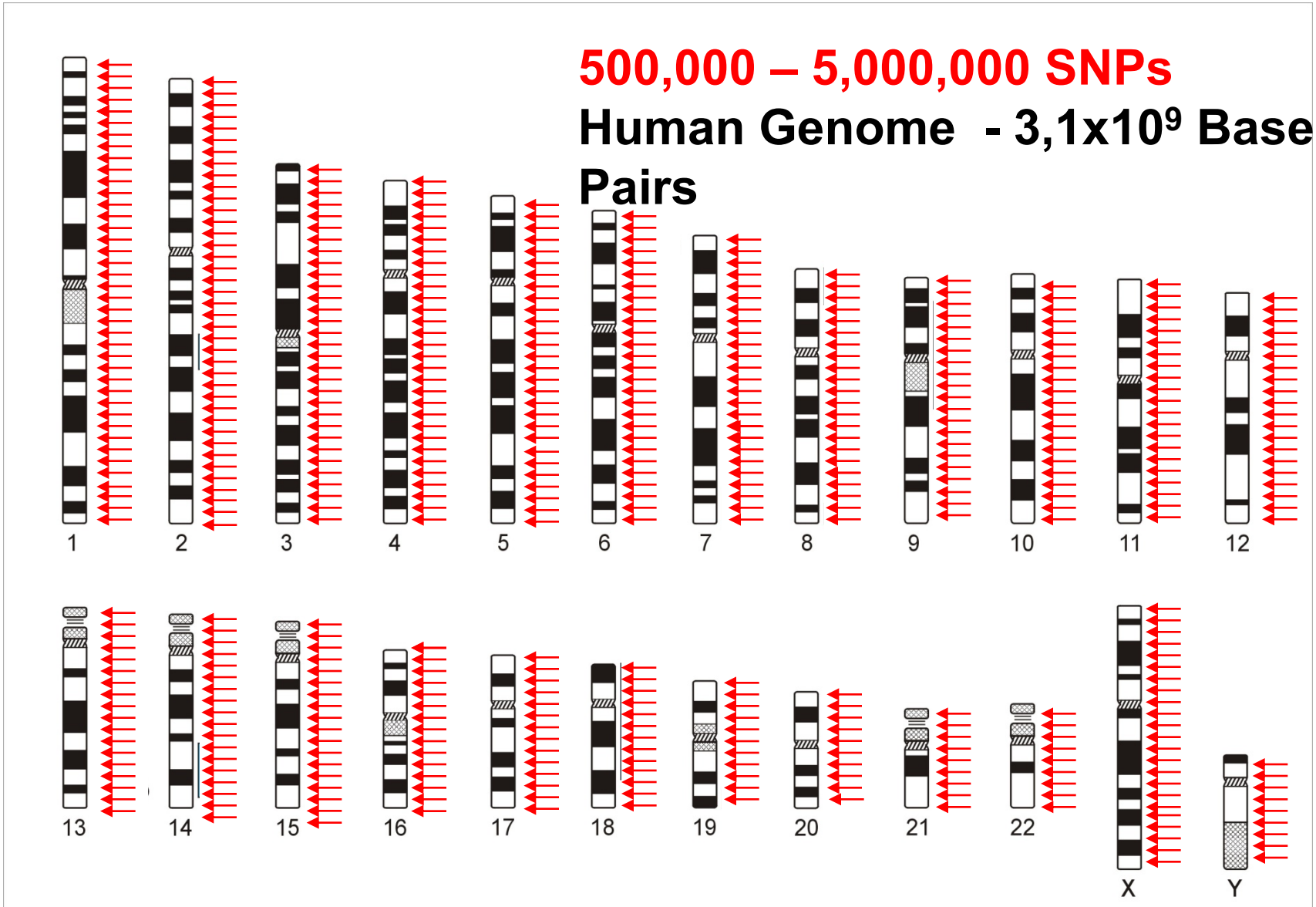
Looks for correlation between specific alleles and phenotype (trait value, disease risk) using single nucleotide polymorphisms (SNPs)



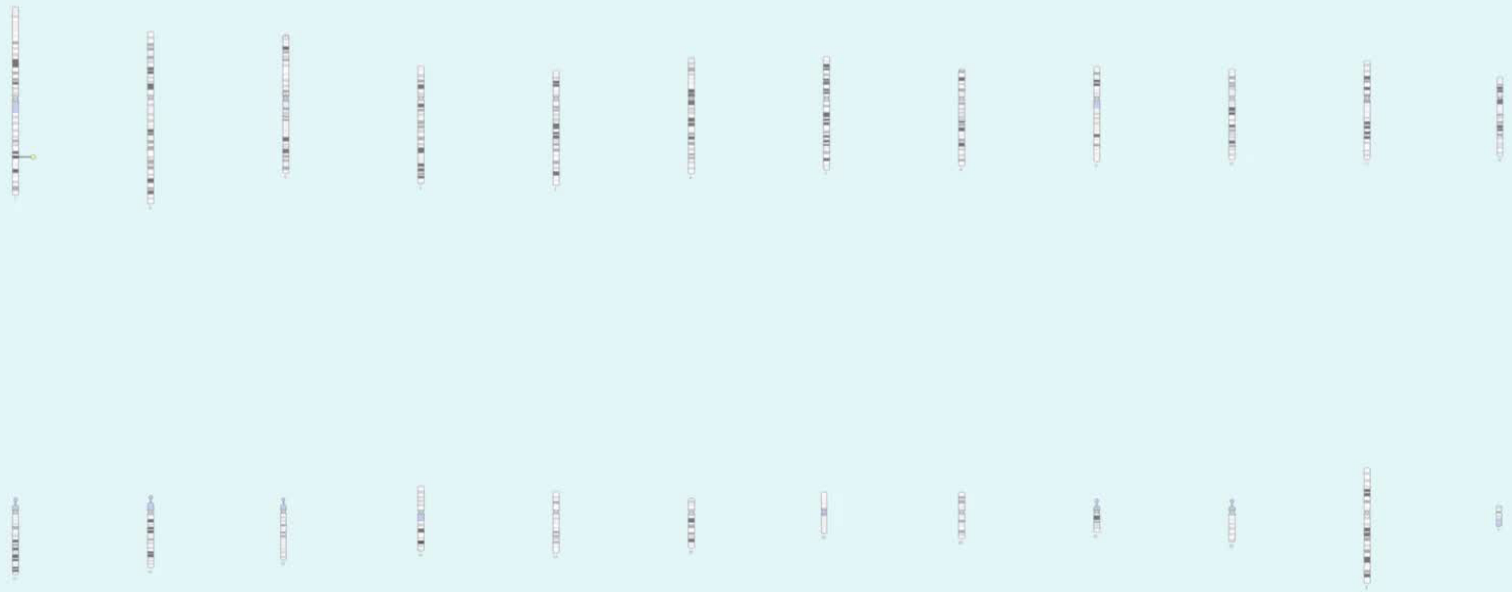
High density SNP arrays – up to 1 million SNPs



Genome-Wide Association Studies



The success of GWAS



www.ebi.ac.uk/gwas

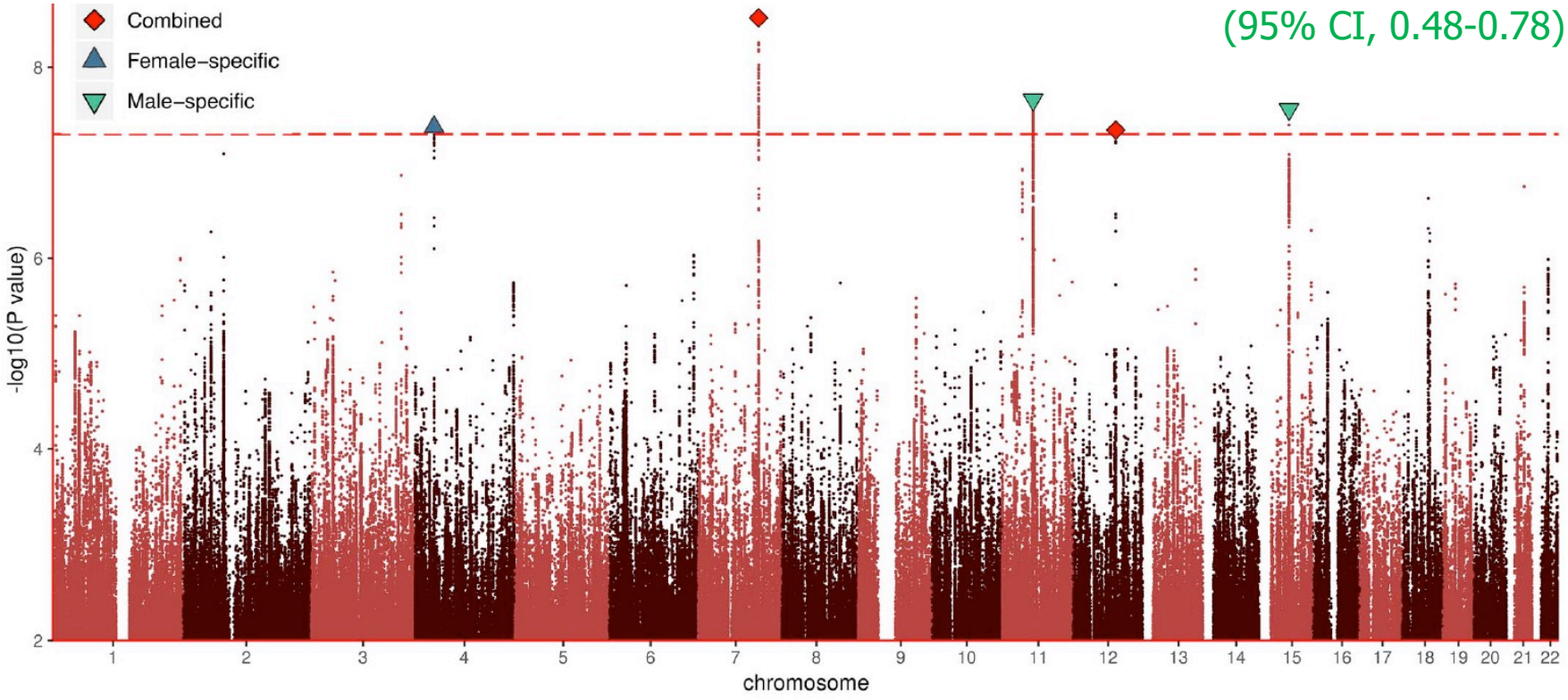
Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior

Total n = 477,522 (26,827 reporting same-sex sexual behavior)

Andrea Ganna[1,2,3,4]*, Karin J. H. Verweij[5]*, Michel G. Nivard[6], Robert Maier[1,2,3], Robbee Wedow[1,3,7,8,9,10,11], Alexander S. Busch[12,13,14], Abdel Abdellaoui[5], Shengru Guo[15], J. Fah Sathirapongsasuti[16], 23andMe Research Team[16], Paul Lichtenstein[4], Sebastian Lundström[17], Niklas Långström[4], Adam Auton[16], Kathleen Mullan Harris[18,19], Gary W. Beecham[15], Eden R. Martin[15], Alan R. Sanders[20,21], John R. B. Perry[12]†, Benjamin M. Neale[1,2,3]†, Brendan P. Zietsch[22]†‡

GCTA h^2 = 32.4%
(95% CI 10.6 - 54.3)

Across sex Rg = 0.63
(95% CI, 0.48-0.78)

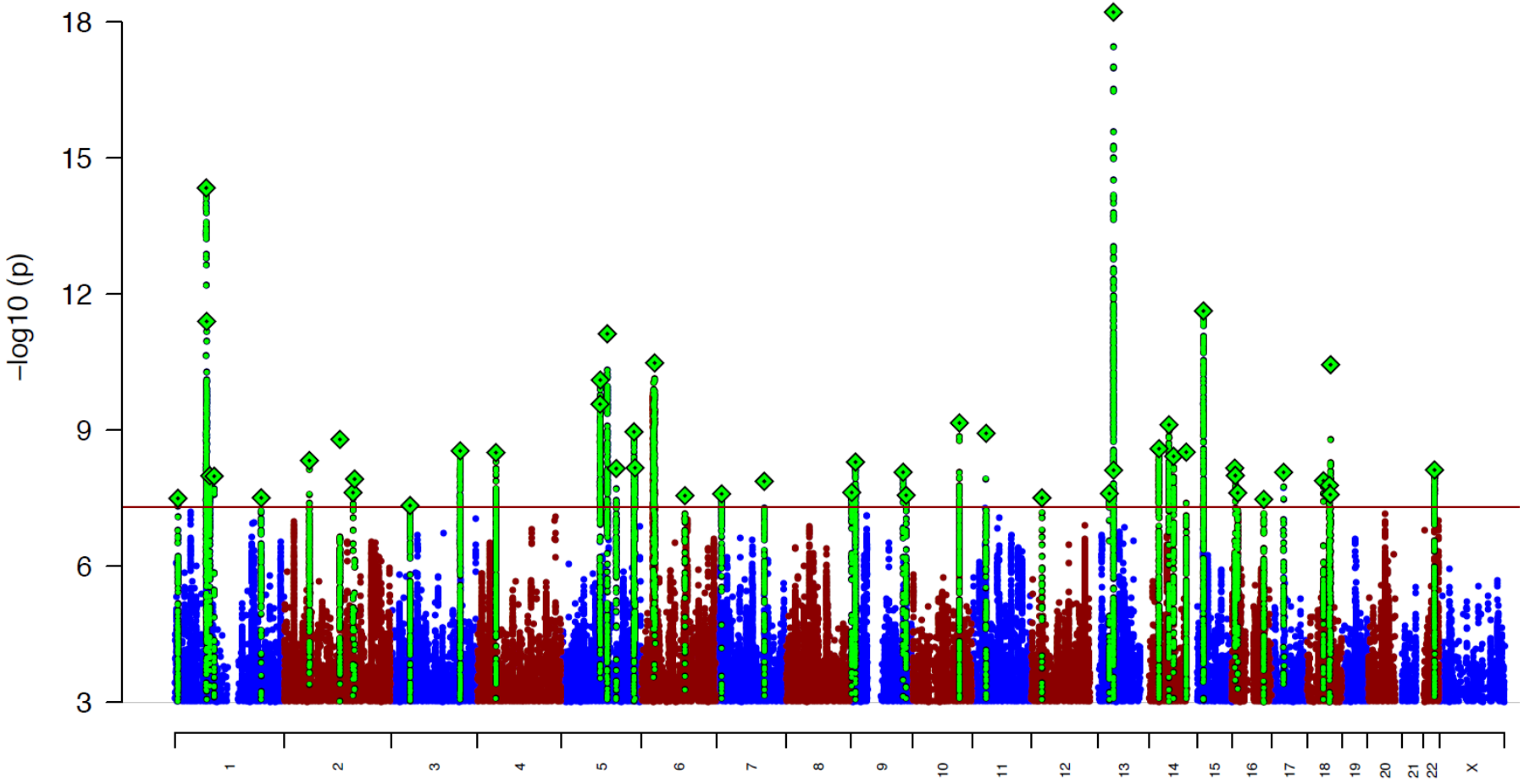


Ganna *et al.*, *Science* **365**, 882 (2019) 30 August 2019

Major Depressive Disorder

130,664 MDD cases , 330,470 controls → 44 independent loci (P < 5e-8)

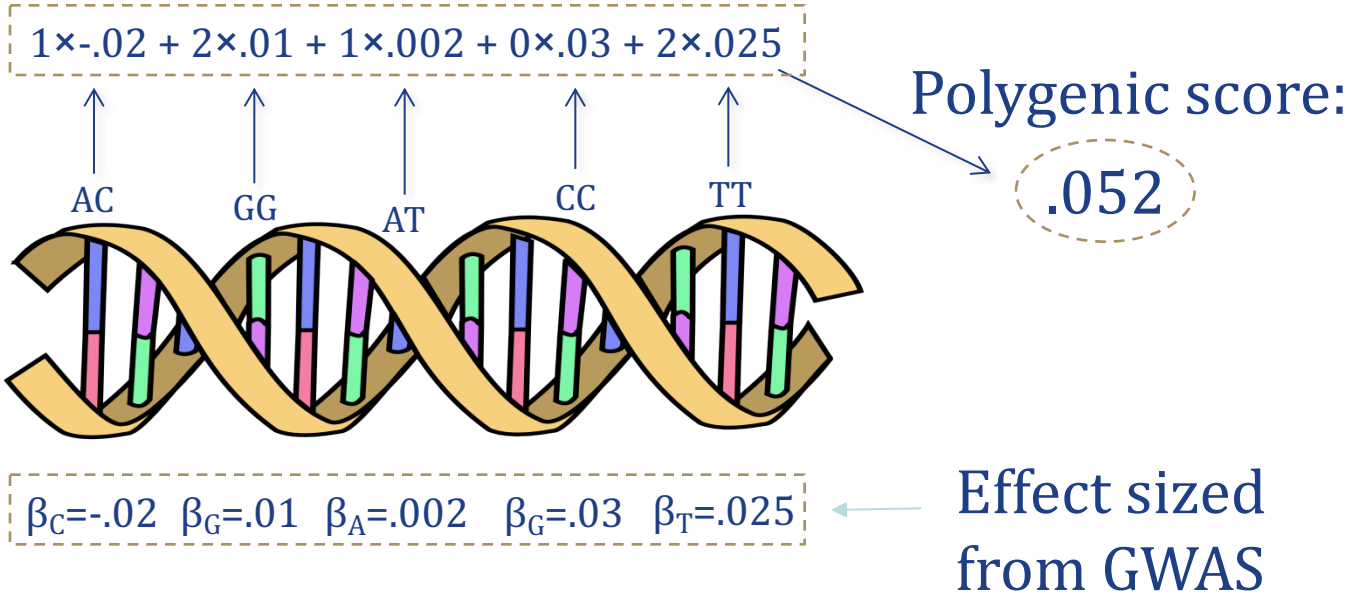
Genes that are targets of antidepressant medications were strongly enriched for MDD association signals (P=8.5e-10)



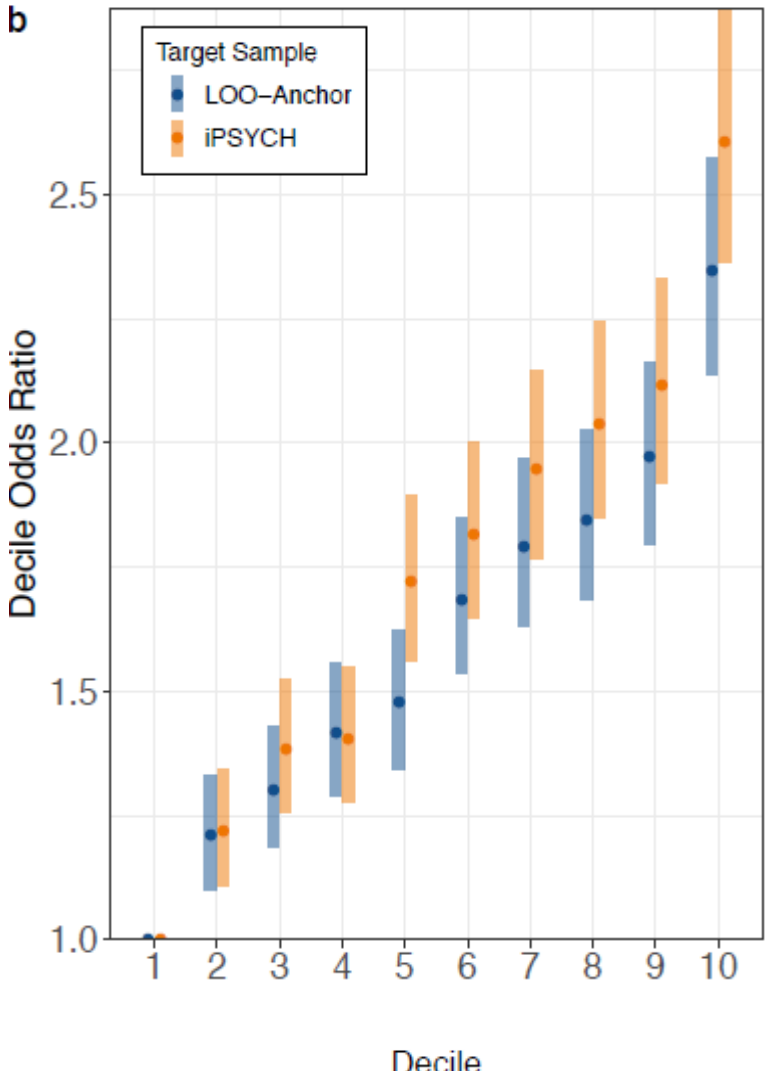
NR Wray….PF Sullivan: bioRxiv Jul. 24, 2017; *Nature Genetics*, in press

Polygenic Risk Scores

Polygenic Risk Scores capture (part of) someone's genetic "risk" by summing all risk alleles weighted by the effect sizes estimated in a Genome-Wide Association Study (GWAS)



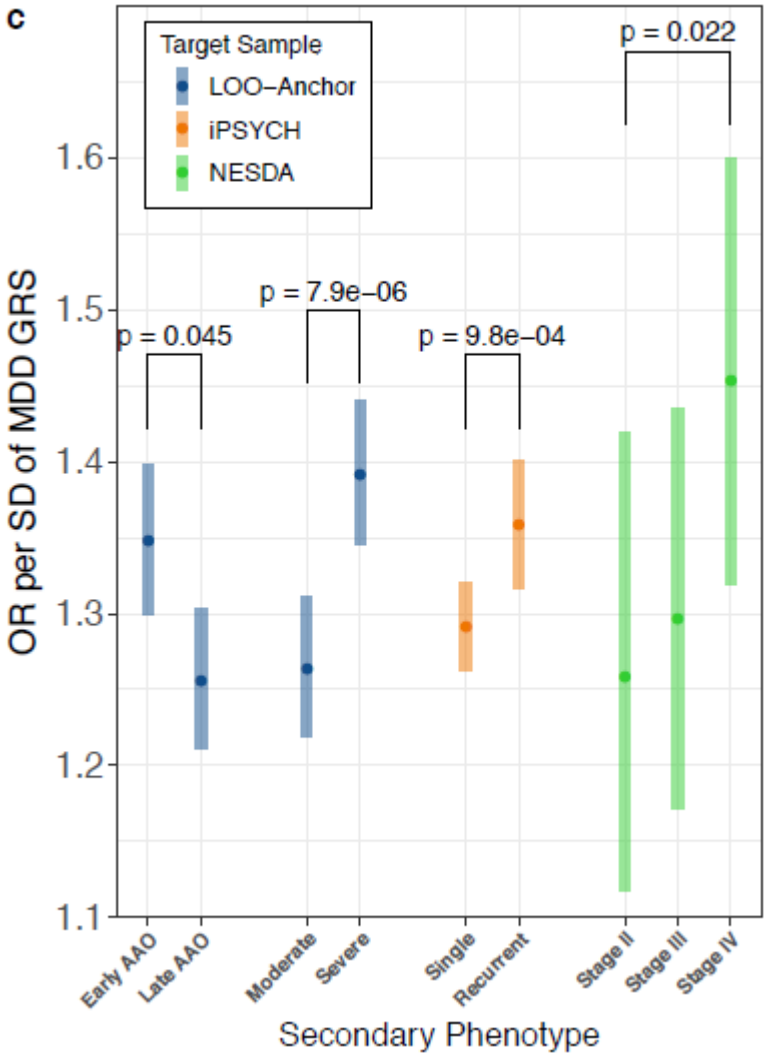
MDD Polygenic Risk Score predicts risk in independent samples



Odd ratios of MDD per PRS decile relative to the first decile for iPSYCH and anchor cohorts.

Interdecile risk ~2.5

MDD Polygenic Risk Score predicts age at onset, recurrence, and severity in independent samples



MDD PRS (from out-of-sample discovery sets) were significantly higher in MDD cases with:

- *earlier age at onset; more severe MDD symptoms (based on number of criteria endorsed)*
- *recurrent MDD compared to single episode*
- *chronic/unremitting MDD ("Stage IV" compared to "Stage II", first-episode MDD)*

Error bars represent 95% CI

nature genetics
Published online 14/8/18
LETTERS
https://doi.org/10.1038/s41588-018-0183-z

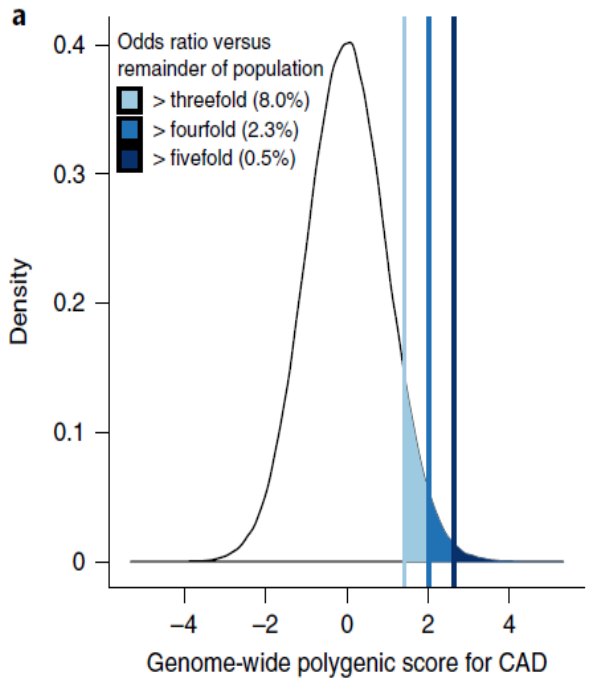
Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations

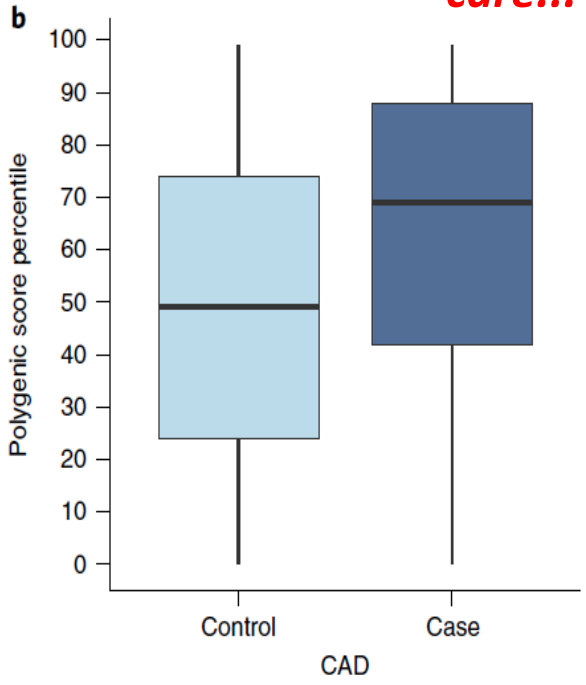
Amit V. Khera[1,2,3,4,5], Mark Chaffin[4,5], Krishna G. Aragam[1,2,3,4], Mary E. Haas[4], Carolina Roselli[4], Seung Hoan Choi[4], Pradeep Natarajan[2,3,4], Eric S. Lander[4], Steven A. Lubitz[2,3,4], Patrick T. Ellinor[2,3,4] and Sekar Kathiresan[1,2,3,4*]

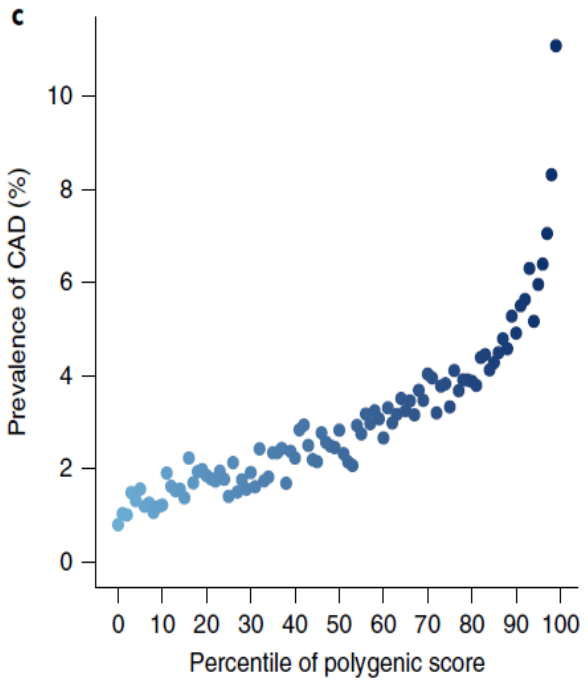
% Population at >3fold increased risk

- CAD 8.0%,
- atrial fibrillation 6.1%
- type 2 diabetes 3.5%
- IBD 3.2%
- breast cancer 1.5%

"We propose that it is time to contemplate the inclusion of polygenic risk prediction in clinical care... "



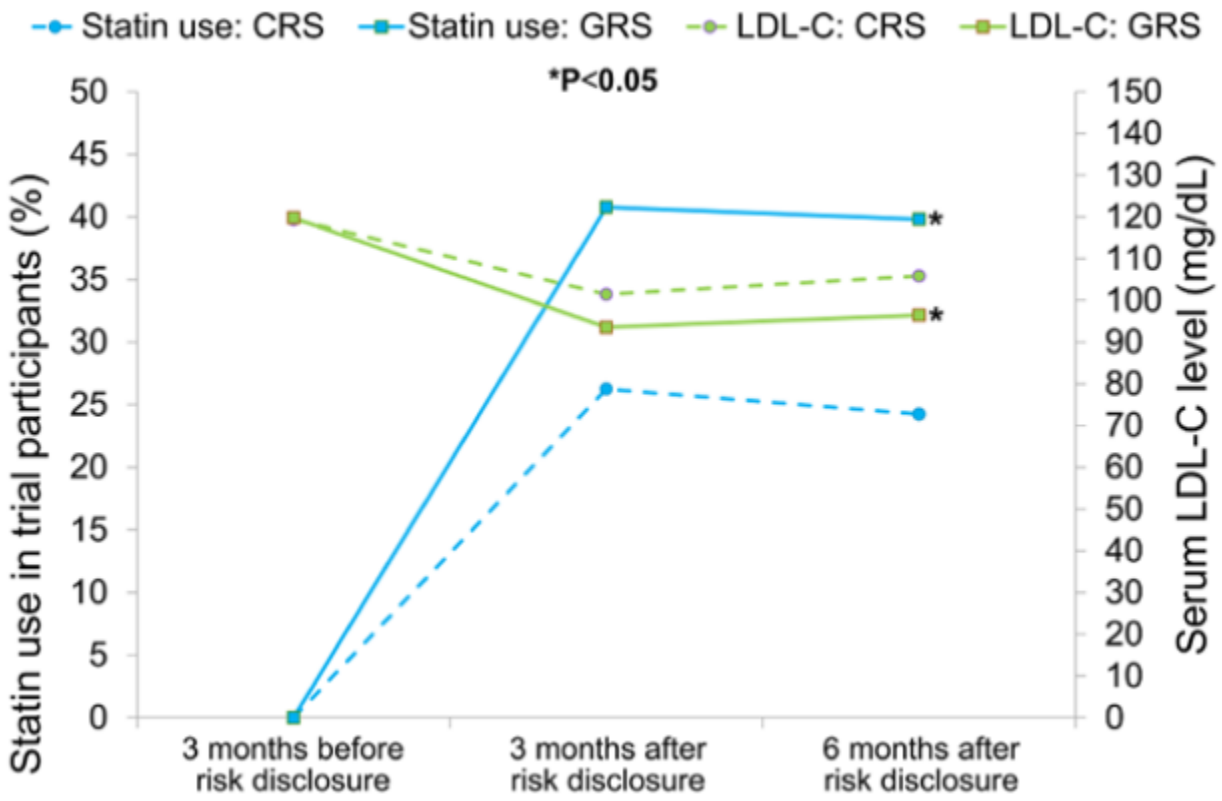




Effect of Disclosing Genetic Risk for Coronary Heart Disease on Information Seeking and Sharing

The MI-GENES Study (Myocardial Infarction Genes)

Sherry-Ann N. Brown, MD, PhD; Hayan Jouni, MD; Tariq S. Marroush, MD; Iftikhar J. Kullo, MD



Statin use significantly higher in patients given genetic risk score than conventional risk score

Circ Cardiovasc Genet. 2017;10:e001613.

GxE Interaction and Correlation

- GxE: SENSITIVITY to E controlled by G

- rGE: EXPOSURE to E correlated with ("depends on") G

Lots of good plant and animal models for both

DIATHESIS-STRESS MODEL IN DEPRESSION

Depression = Diathesis + Stress + Diathesis * Stress

(Predisposition, Vulnerability)

(Disruption of psychological equilibrium)

Hypothesised contribution to risk

Our model:

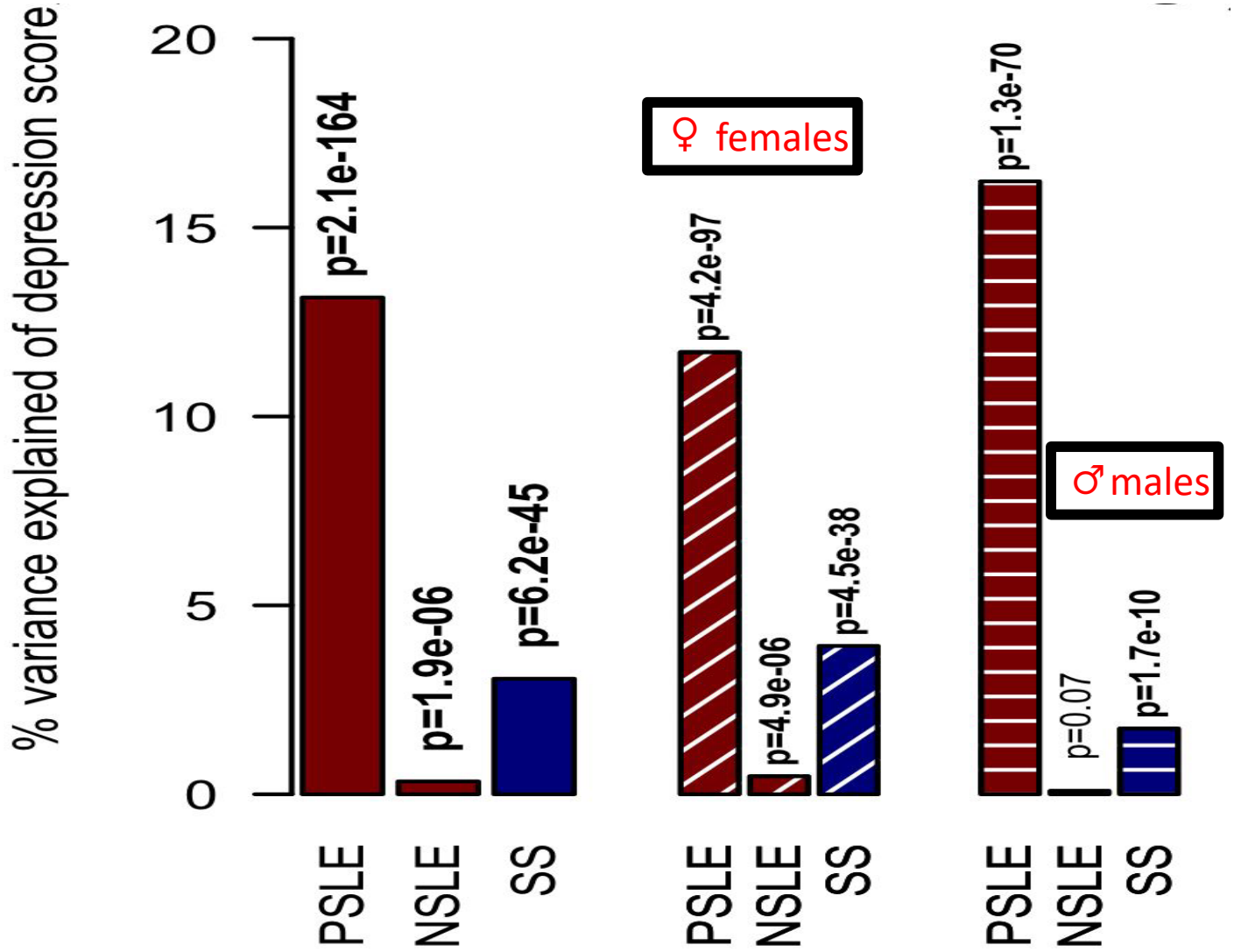
Depression = PRS + PSLE + NSLE + SS + PRS * PSLE + PRS*NSLE + PRS*SS

- PRS: Polygenic Risk Score
- PSLE: Personal Stressful Life Event, NSLE: Network Stressful life event, SS Social Support
- Familial relatedness modelled using a kinship matrix or a genetic relatedness matrix calculated from SNPs.
- All analyses controlled for age, age^2, sex, age*sex and age^2*sex interactions, study, array, and the first four genetic principal components
- Scores regressed for covariates to avoid false positive interactions

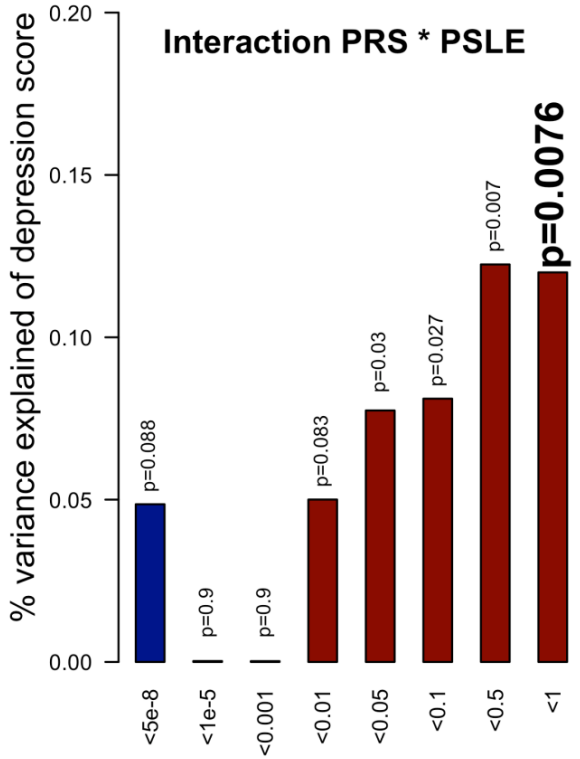
Mol Psychiatry 23:1590-1596, 2018

MAIN EFFECTS OF STRESS SCORES

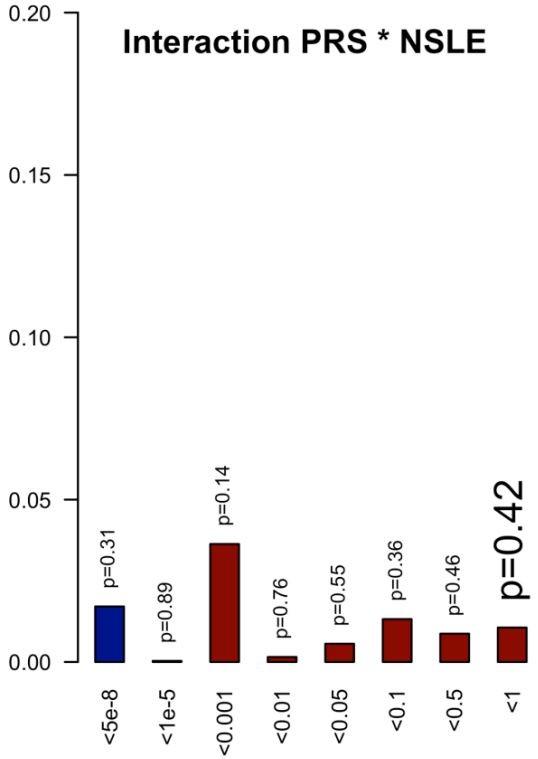
- Depression score associated with Personal (PSLE) and network (NSLE) stressful life events and perceived social support (SS). Note sex differences.



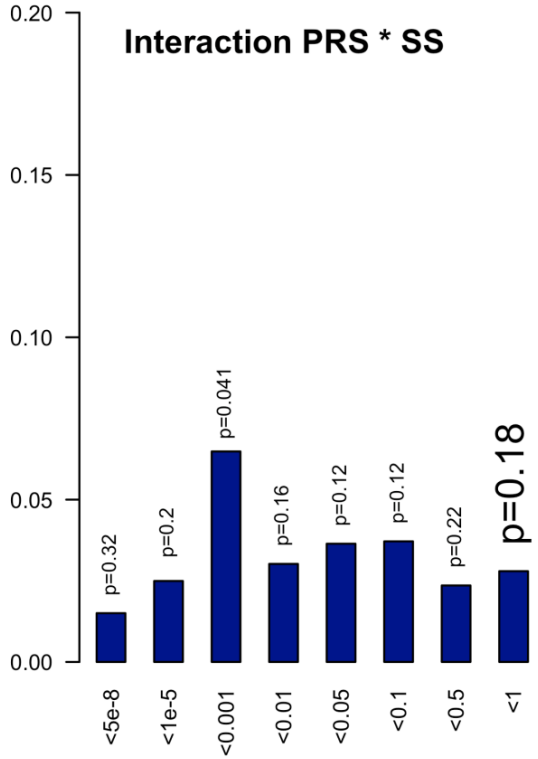
TEST OF INTERACTION



Significant interaction



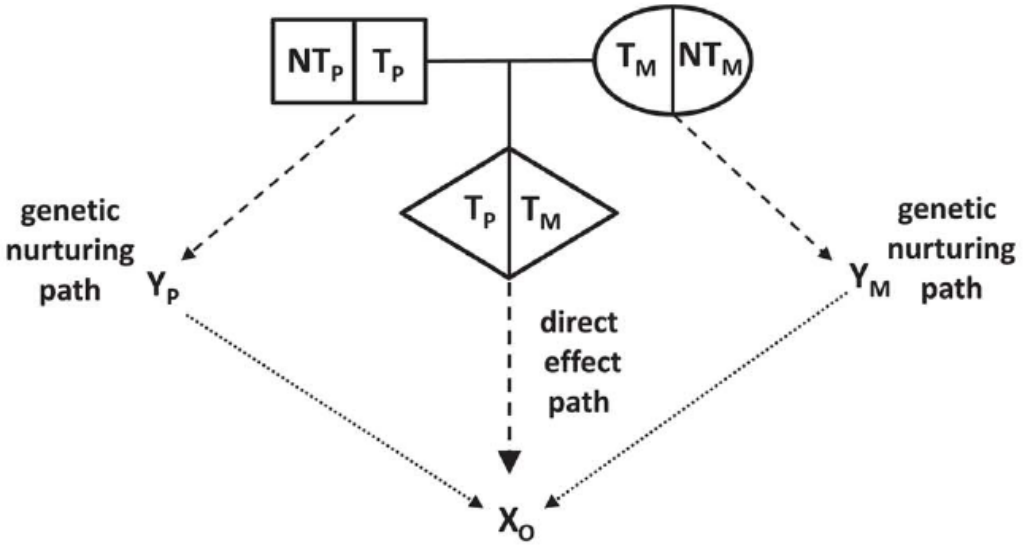
No interaction



No interaction

The nature of nurture: Effects of parental genotypes

Augustine Kong …….Kari Stefansson



Nontransmitted alleles can affect a child through their impacts on the parents and other relatives, a phenomenon we call “genetic nurture.” Using results from a meta-analysis of educational attainment, we find that the polygenic score computed for the nontransmitted alleles of 21,637 probands with at least one parent genotyped has an estimated effect on the educational attainment of the proband that is 29.9% ($P = 1.6 \times 10^{-14}$) of that of the transmitted polygenic score.

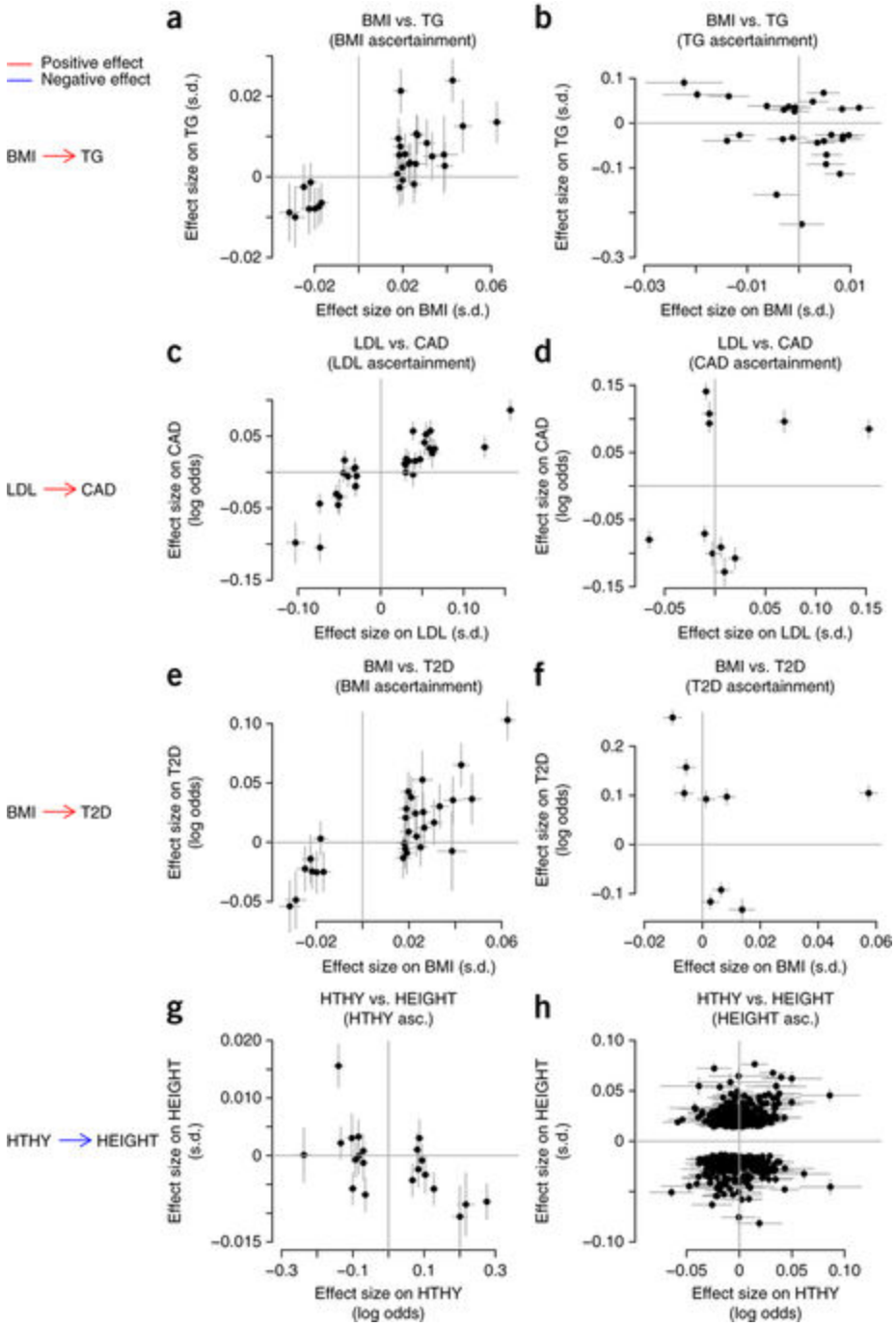
Kong *et al.*, *Science* **359**, 424–428 (2018) 26 January 2018

Detection and interpretation of shared genetic influences on 42 human traits

Joseph K Pickrell, Tomaz Berisa, Jimmy Z Liu, Laure Ségurel, Joyce Y Tung & David A Hinds.

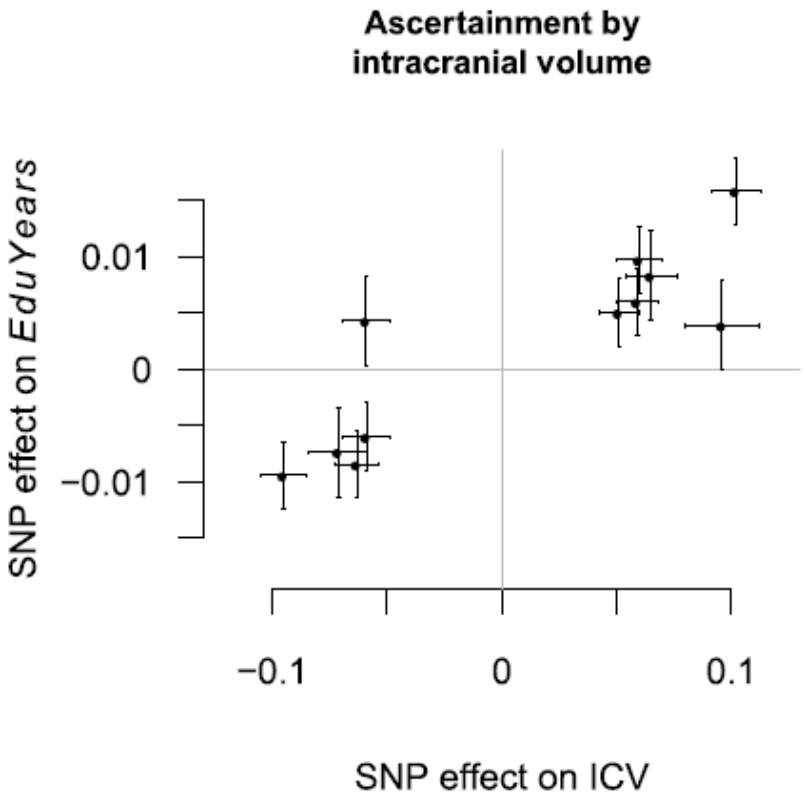
Nature Genetics 48; 709–717, 2016

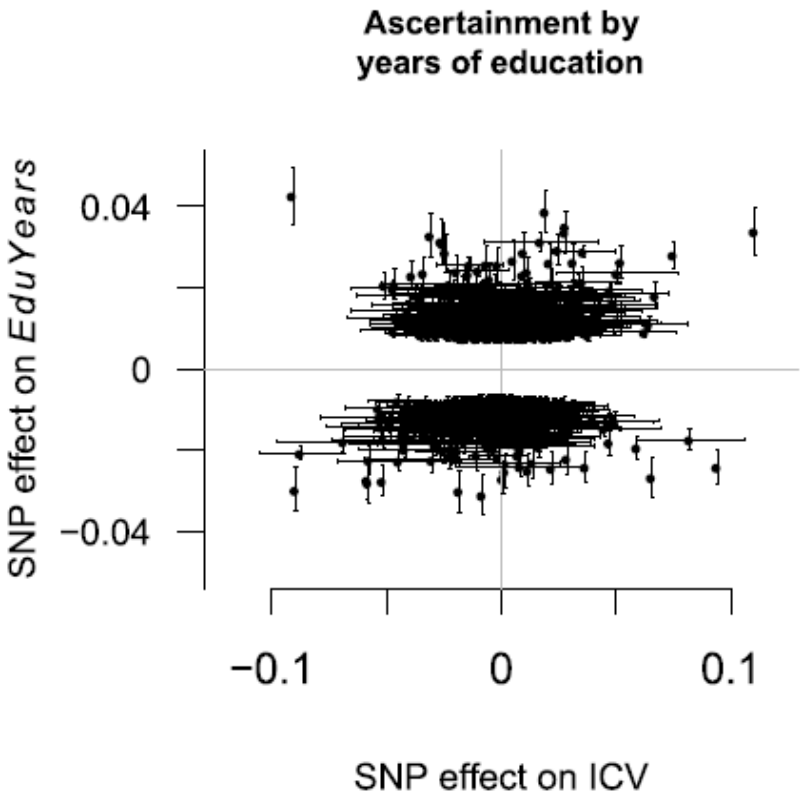
Powerful GWAS for traits A and B can help determine direction of causation



The causal influence of brain size on human intelligence: Evidence from within-family phenotypic associations and GWAS modeling

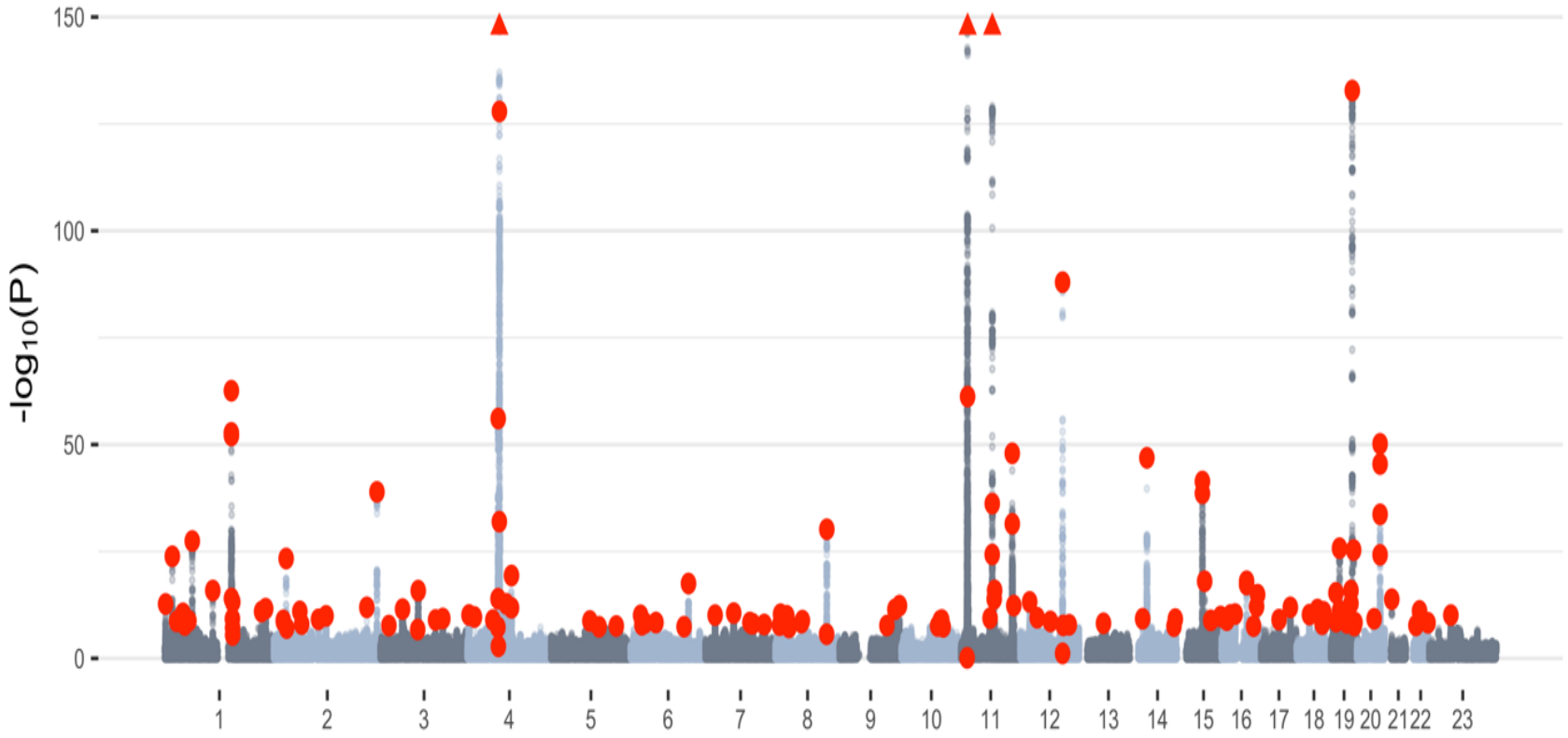
James J. Lee[a,*], Matt McGue[a], William G. Iacono[a], Andrew M. Michael[b,c], Christopher F. Chabris[b]





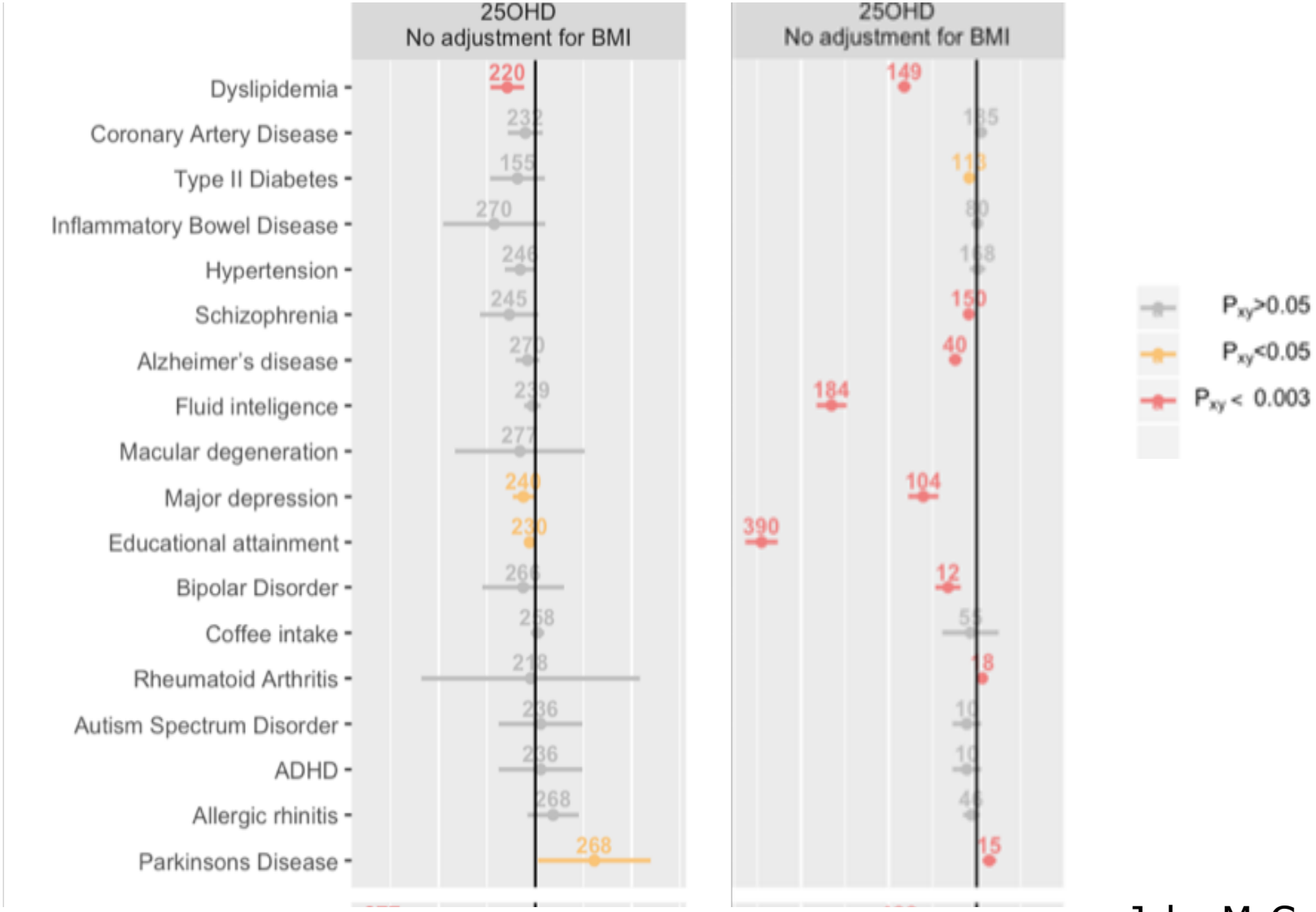
Intelligence 75 (2019) 48–58

Manhattan plot of the 25OHD (vitamin D) GWAS in the UK Biobank: n=417,580, 143 loci



John McGrath

Bidirectional Generalized Summary data level Mendelian Randomization (GSMR) between 25 hydroxyvitamin D concentrations and selected phenotypes



John McGrath

Inference on heritability and selection from WGS data

Heritability estimates from different studies:

Source	Height	BMI	Year	EA	SCZ
Pedigree studies	**~ 0.8**	**~ 0.4 - 0.6**	< 2010	**~0.4**	**~0.7**
Genotyped SNPs	~ 0.45 (0.03)	~ 0.16 (0.03)	2010+	**~0.15**	**~0.25**
Imputed SNPs	~ 0.56 (0.02)	~ 0.27 (0.02)	2015+	**~0.2**	**~0.3**
Whole Genome	???			**???**	**???**

Can we recover the 'still missing' heritability using WGS data?

Investigation using TOPMed WGS

TOPMed overview

NIH project, whole genome sequencing of entire cohorts in the US (~30 cohorts)

~40% Europeans

~30% African americans

~16% latinos

~9% Asians

~4% unkown

Median sequencing depth 30x

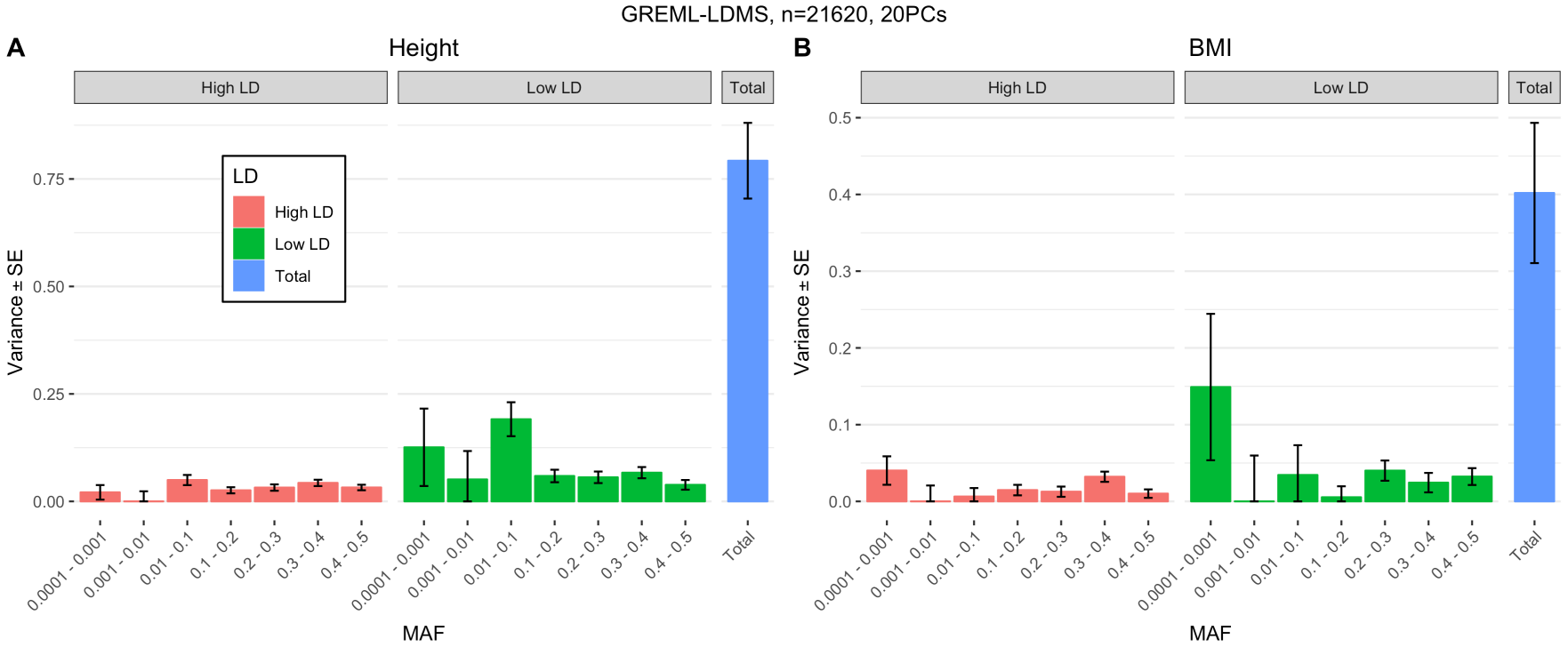
Multiple data freezes:

- 54k samples for freeze 5
- 140k samples for freeze 8

Peter Visscher

Pierrick Wainschtein

GREML-LDMS using WGS data: recovery of pedigree heritability?



Estimates using 20PCs as fixed effects:

- Height:

$$h^2_{WGS} = 0.79\ (0.09)$$

- BMI:

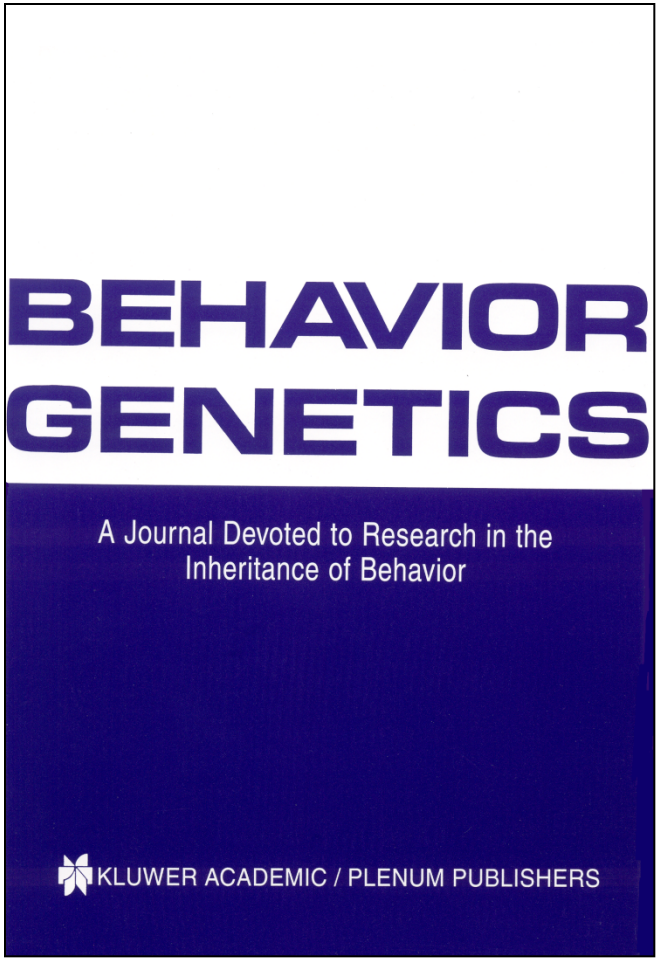
$$h^2_{WGS} = 0.40\ (0.09)$$

Estimates close to pedigree estimates

Large role for low LD and low MAF variants

Wainschtein, bioRxiv 2019

We also run two journals (1)



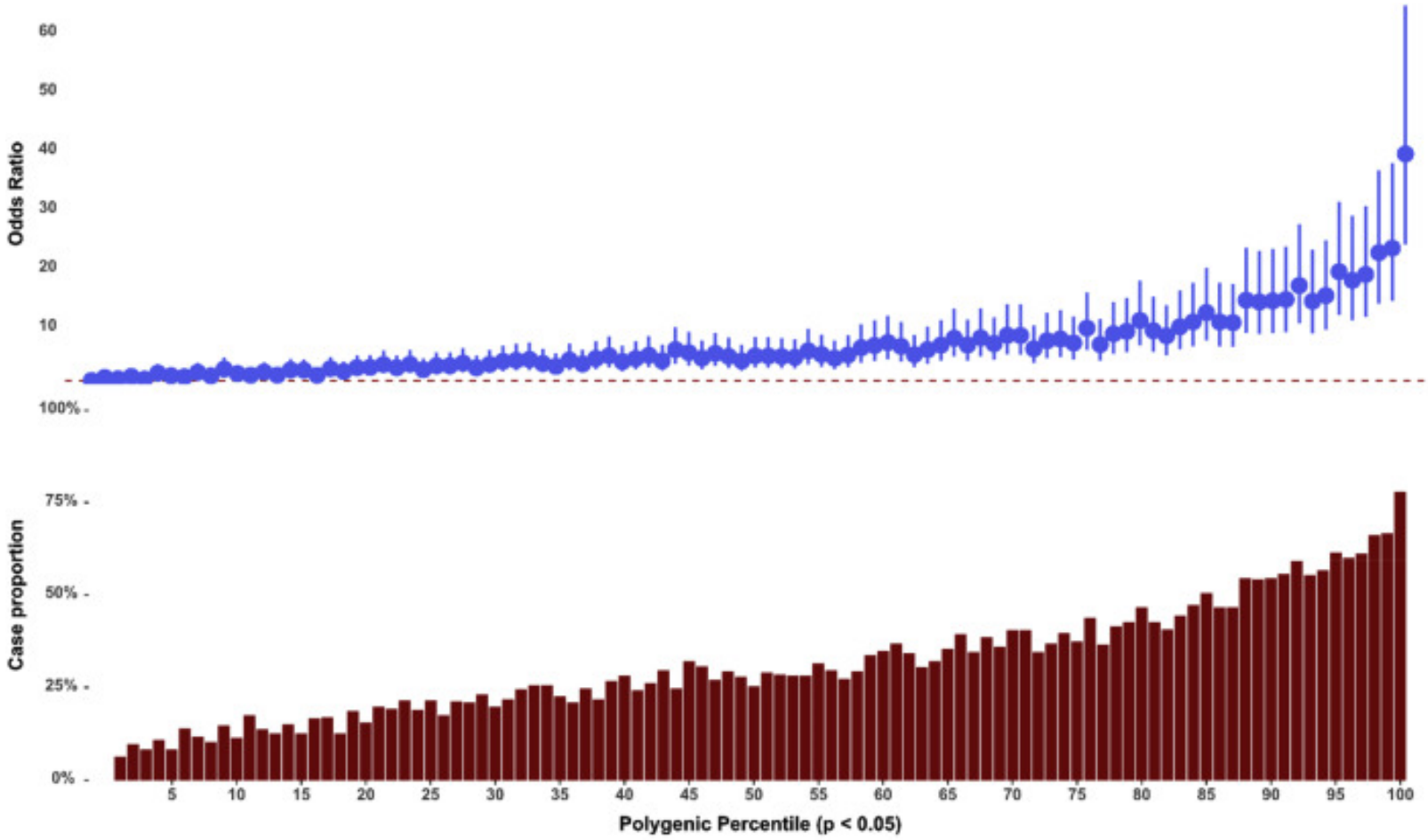
- Editor: John Hewitt
- Editorial assistant Christina Hewitt
- Publisher: Kluwer /Plenum
- Fully online
- http://www.bga.org

- Editor: Nick Martin
- Publisher: Cambridge University Press
- Fully online
- Fast turnaround
- **First submission free to workshop participants!!!!!☺**

Schizophrenia GWAS: polygenic centile analysis

Independent discovery sample (29,415 cases and 40,101 controls) used to predict to CLOZUK target sample (11,260 cases and 24,542 controls)

Risk of schizophrenia in top 1% is ~40-fold higher than in bottom 1%

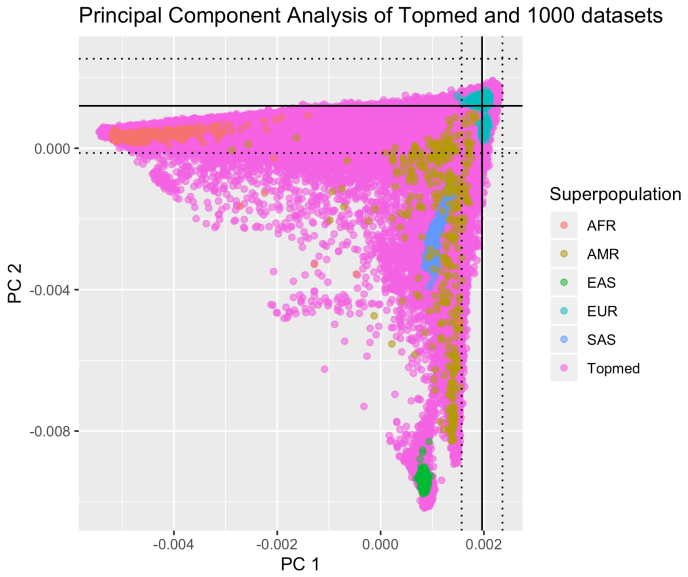


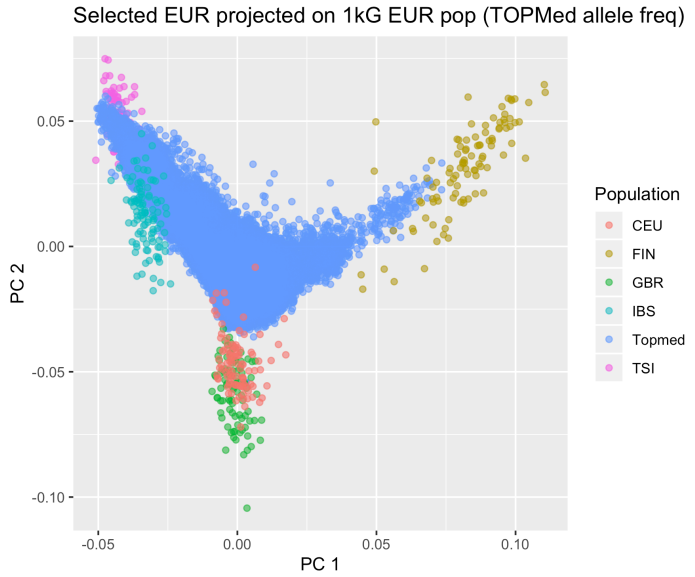
Pardinas et al. *Nature Genetics* 50: 381–389 (2018)

European ancestry selection

Selection of European ancestry samples

1. Using HM3 SNPs (1.2M)
2. From 1kG population references
3. Selected samples within 6sd of EUR pop





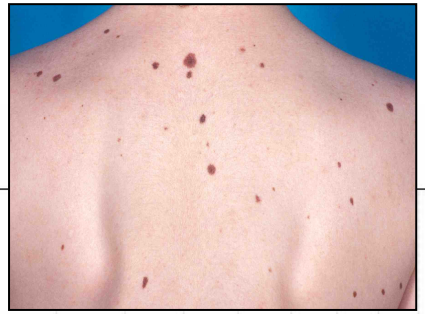
Dataset

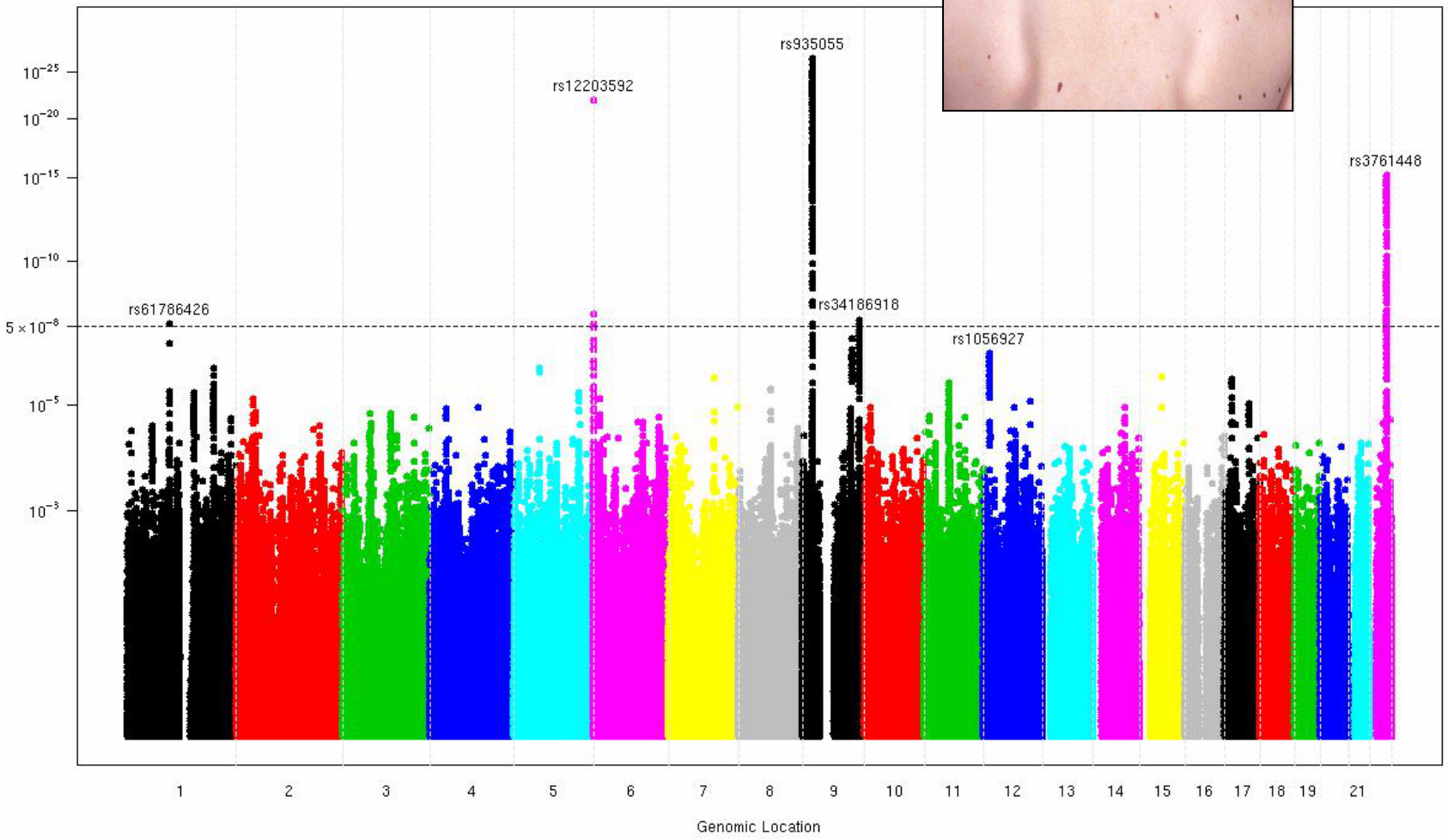
Analysis parameters:

- 21,620 unrelated individuals from European ancestry
- Genomic Relationship Matrices (GRMs) fitted:
  - 1 GRM from HM3 SNPs
  - 7 GRMs from all variants split by MAF
  - 14 GRM split by MAF and LD (7 AF bins * 2 LD bins)
- MAF threshold of 0.0001 (each variant seen > 2x)
- 47.1M variants
- Phenotypes standardized N(0, 1) within each cohort by sex

RINT = rank based inverse normal transformation

GWAS for moliness(11 studies, 23,000 subjects, 5 hits)





David Duffy, Gu Zhu

NATURE GENETICS ADVANCE ONLINE PUBLICATION

The support of human genetic evidence for approved drug indications

Matthew R Nelson[1], Hannah Tipney[2], Jeffery L Painter[1], Judong Shen[1], Paola Nicoletti[3], Yufeng Shen[3,4], Aris Floratos[3,4], Pak Chung Sham[5,6], Mulin Jun Li[6,7], Junwen Wang[6,7], Lon R Cardon[8], John C Whittaker[2] & Philippe Sanseau[2]

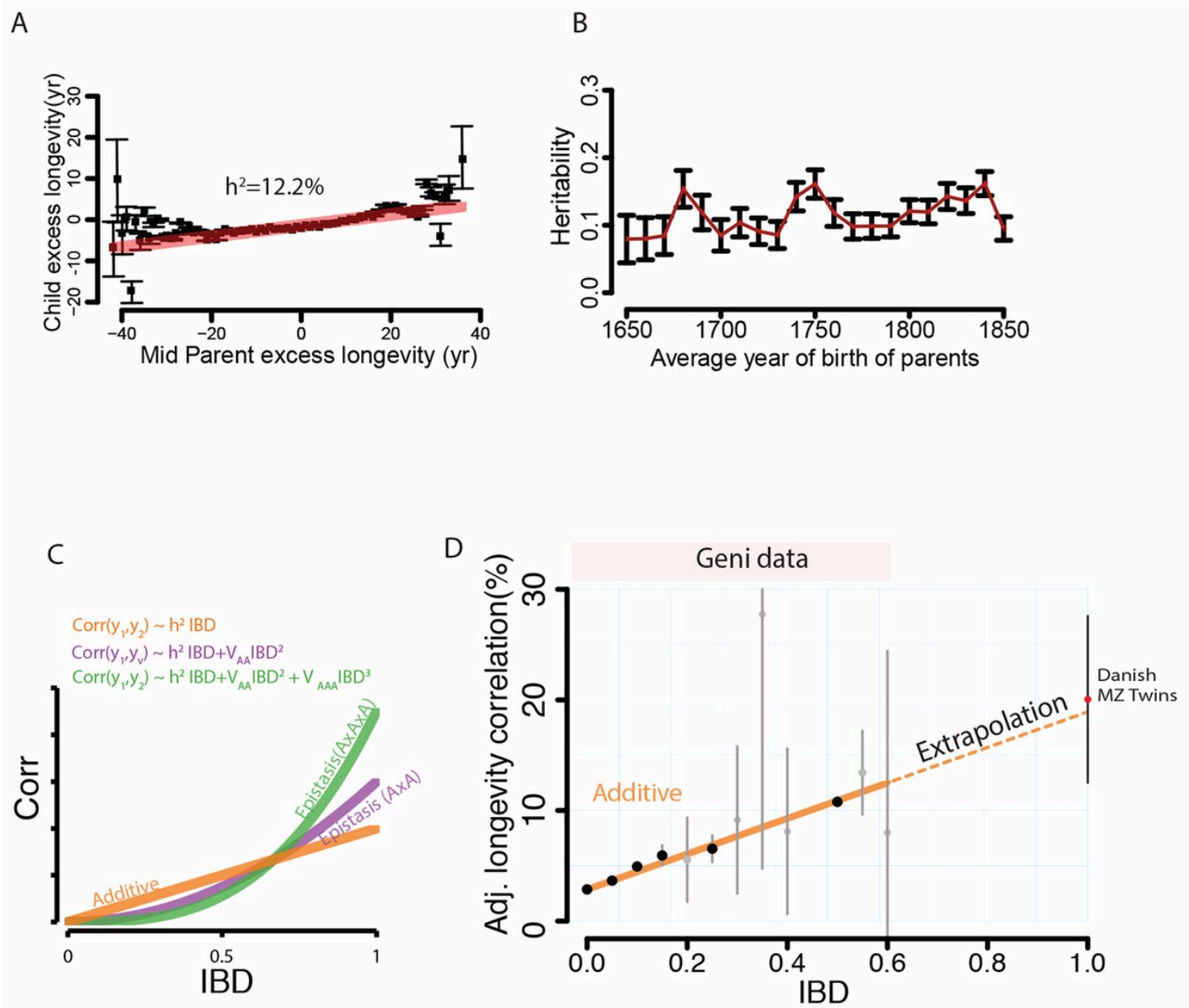
Table 1 The relative value of genetic support for the probability that a target-indication pair progresses along the drug development pipeline, based on historical drug trial information

	p(progress|genetic support)/(progress|no genetic support)		
Progression	GWASdb and OMIM	GWASdb	OMIM
Phase I to phase II	1.2 (1.1–1.3)	1.2 (1.1–1.3)	1.2 (1.1–1.3)
Phase II to phase III	1.5 (1.3–1.7)	1.4 (1.2–1.7)	1.6 (1.3–1.9)
Phase III to approval	1.1 (1.0–1.2)	1.0 (0.8–1.2)	1.1 (0.9–1.3)
Phase I to phase III	1.8 (1.5–2.1)	1.8 (1.4–2.1)	1.9 (1.5–2.3)
Phase I to approval	2.0 (1.6–2.4)	1.8 (1.3–2.3)	2.2 (1.6–2.8)

"We estimate that selecting genetically supported targets could double the success rate in clinical development. Therefore, using the growing wealth of human genetic data to select the best targets and indications should have a measurable impact on the successful development of new drugs."

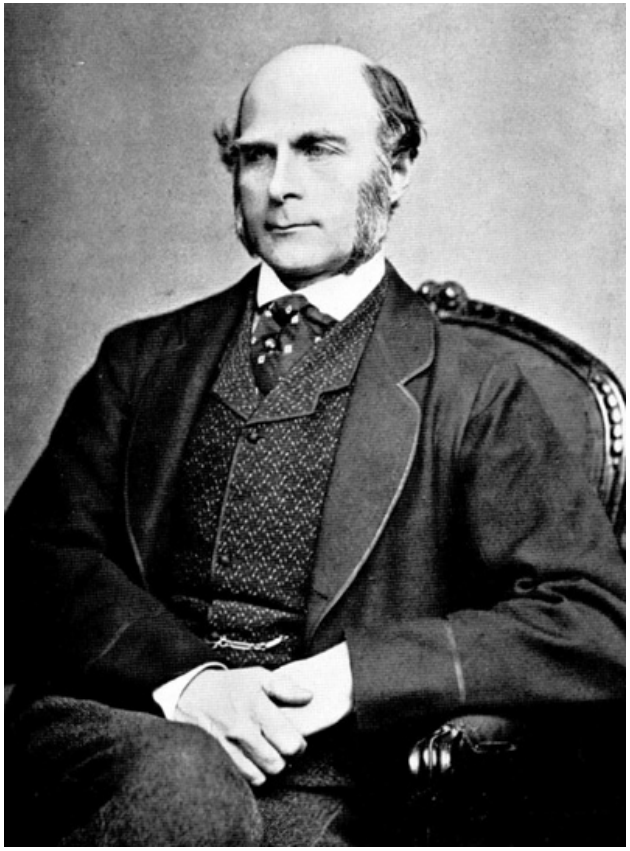
Quantitative analysis of population-scale family trees with millions of relatives

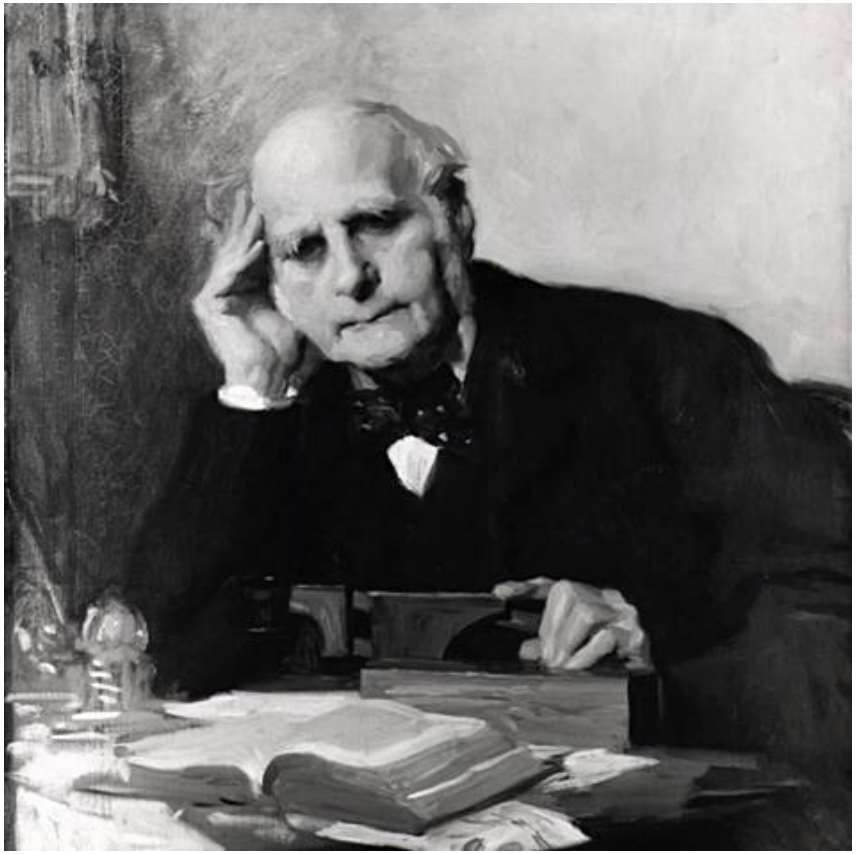
Joanna Kaplanis *et al.* *Science* 01 Mar 2018:



“HEREDITY”

Francis Galton (1822-1911)



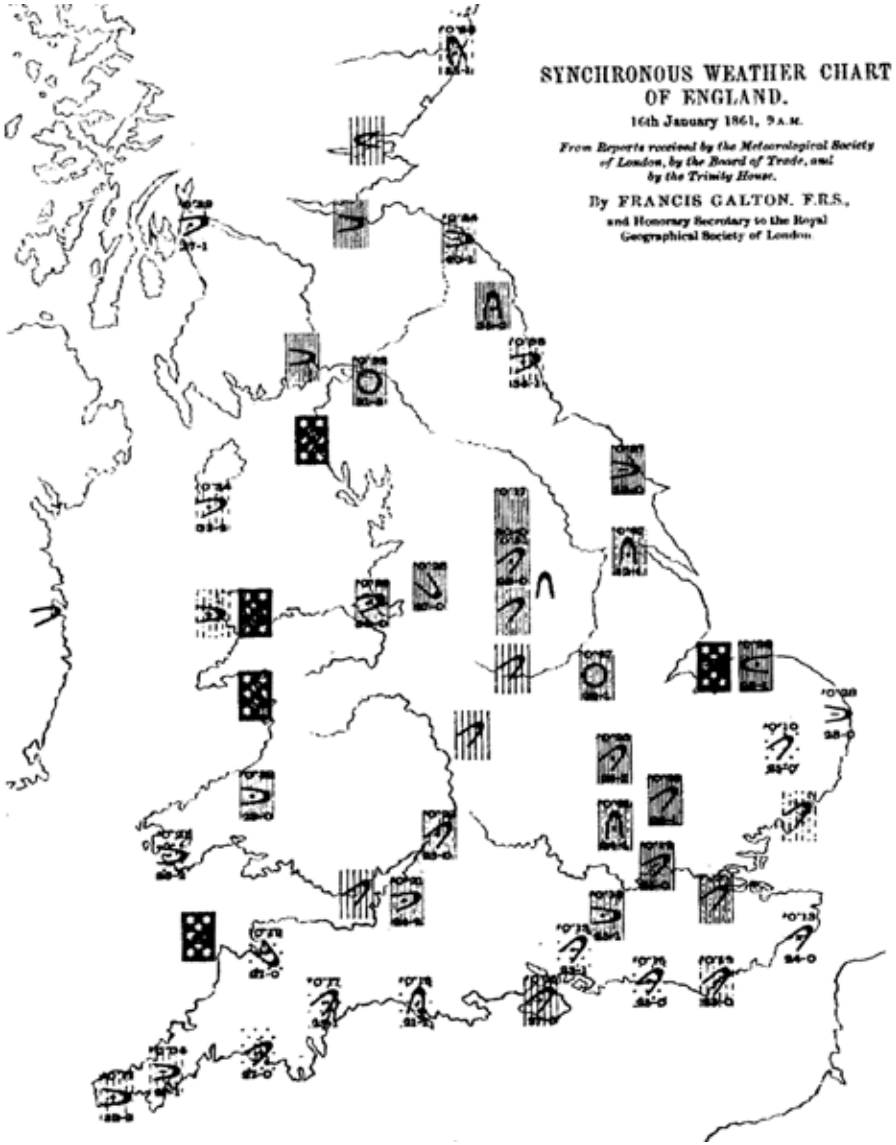


1869: Hereditary Genius

1883: Inquiries into Human Faculty and its Development

1884-5: Anthropometic Laboratory at “National Health Exhibition”

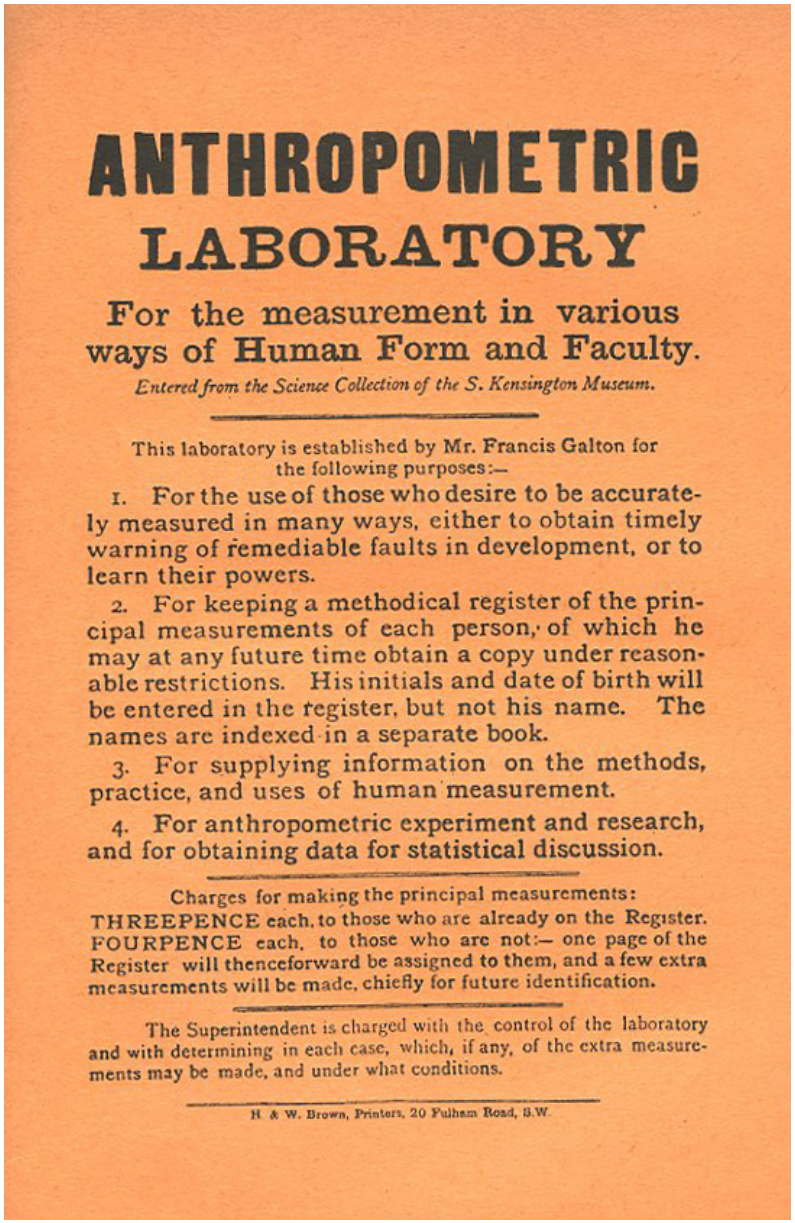
Galton's Other Work e.g. Meteorology



Hereditary Genius (1869, p 317)

	Judges, p. 61.	Statesmen, p. 109.	Commanders, p. 148.	Literary, p. 171.	Scientific, p. 195.	Poets, p. 227.	Artists, pp. 238 and 249.	Divines, p. 275.	Illustrious and Eminent Men of all Classes.		
	B.	B.	B.	B.	B.	B.	B.	B.	B.	C.	D.
Father	26	33	47	48	26	20	32	28	31	100	31
Brother	35	39	50	42	47	40	50	36	41	150	27
Son	36	49	31	51	60	45	89	40	48	100	48
Grandfather	15	28	16	24	14	5	7	20	17	200	8
Uncle	18	18	8	24	16	5	14	40	18	400	5
Nephew	19	18	35	24	23	50	18	4	22	400	5
Grandson	19	10	12	9	14	5	18	16	14	200	7
Great-grandfather	2	8	8	3	0	0	0	4	3	400	1
Great-uncle	4	5	8	6	5	5	7	4	5	800	1
First cousin	11	21	20	18	16	0	1	8	13	800	2
Great-nephew	17	5	8	6	16	10	0	0	10	800	1
Great-grandson	6	0	0	3	7	0	0	0	3	400	1
All more remote	14	37	44	15	23	5	18	16	31	?	...

Galton's Anthropometric Laboratory (1884-1885)



ANTHROPOMETRIC LABORATORY

For the measurement in various ways of Human Form and Faculty.

Entered from the Science Collection of the S. Kensington Museum.

This laboratory is established by Mr. Francis Galton for the following purposes:—

1. For the use of those who desire to be accurately measured in many ways, either to obtain timely warning of remediable faults in development, or to learn their powers.
2. For keeping a methodical register of the principal measurements of each person, of which he may at any future time obtain a copy under reasonable restrictions. His initials and date of birth will be entered in the register, but not his name. The names are indexed in a separate book.
3. For supplying information on the methods, practice, and uses of human measurement.
4. For anthropometric experiment and research, and for obtaining data for statistical discussion.

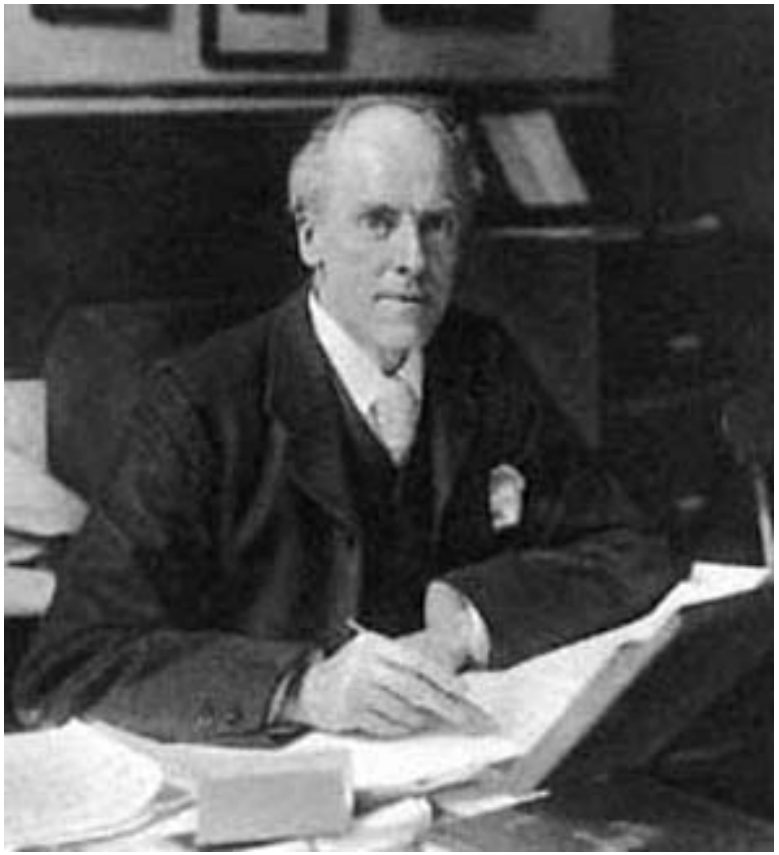
Charges for making the principal measurements:
THREEPENCE each, to those who are already on the Register.
FOURPENCE each, to those who are not:— one page of the Register will thenceforward be assigned to them, and a few extra measurements will be made, chiefly for future identification.

The Superintendent is charged with the control of the laboratory and with determining in each case, which, if any, of the extra measurements may be made, and under what conditions.

H. & W. Brown, Printers, 20 Fulham Road, S.W.

Francis Galton's First Anthropometric Laboratory at the International Health Exhibition, South Kensington, 1884–5.

Karl Pearson (1857-1936)



1903: On the Laws of Inheritance in Man: I Physical Characteristics (with Alice Lee)

1904: II Mental and Moral Characteristics

1914: The Life, Letters and Labours of Francis Galton

FAMILY MEASUREMENTS.

Professor KARL PEARSON, of University College, London, would esteem it a great favour if any persons in a position to do so, would assist him by making one set (or if possible several sets) of anthropometric measurements on their own family, or on families with whom they are acquainted. The measurements are to be made use of for testing theories of heredity, no names, except that of the recorder, are required, but the Professor trusts to the *bona fides* of each recorder to send only correct results.

Each family should consist of a father, mother, and at least one son or daughter, not necessarily the eldest. The sons or daughters are to be at least 18 years of age, and measurements are to be made on not more than two sons and two daughters of the same family. If more than two sons or two daughters are easily accessible, then not the tallest but the eldest of those accessible should be selected.

To be of real service the whole series ought to contain 1000—2000 families, and therefore the Professor will be only too grateful if anyone will undertake several families for him.

Copies of this paper, together with cards for recording data, may be obtained from

.....

or from the above-named Professor.

Pearson and Lee's diagram for measurement of "span" (finger-tip to finger-tip distance)

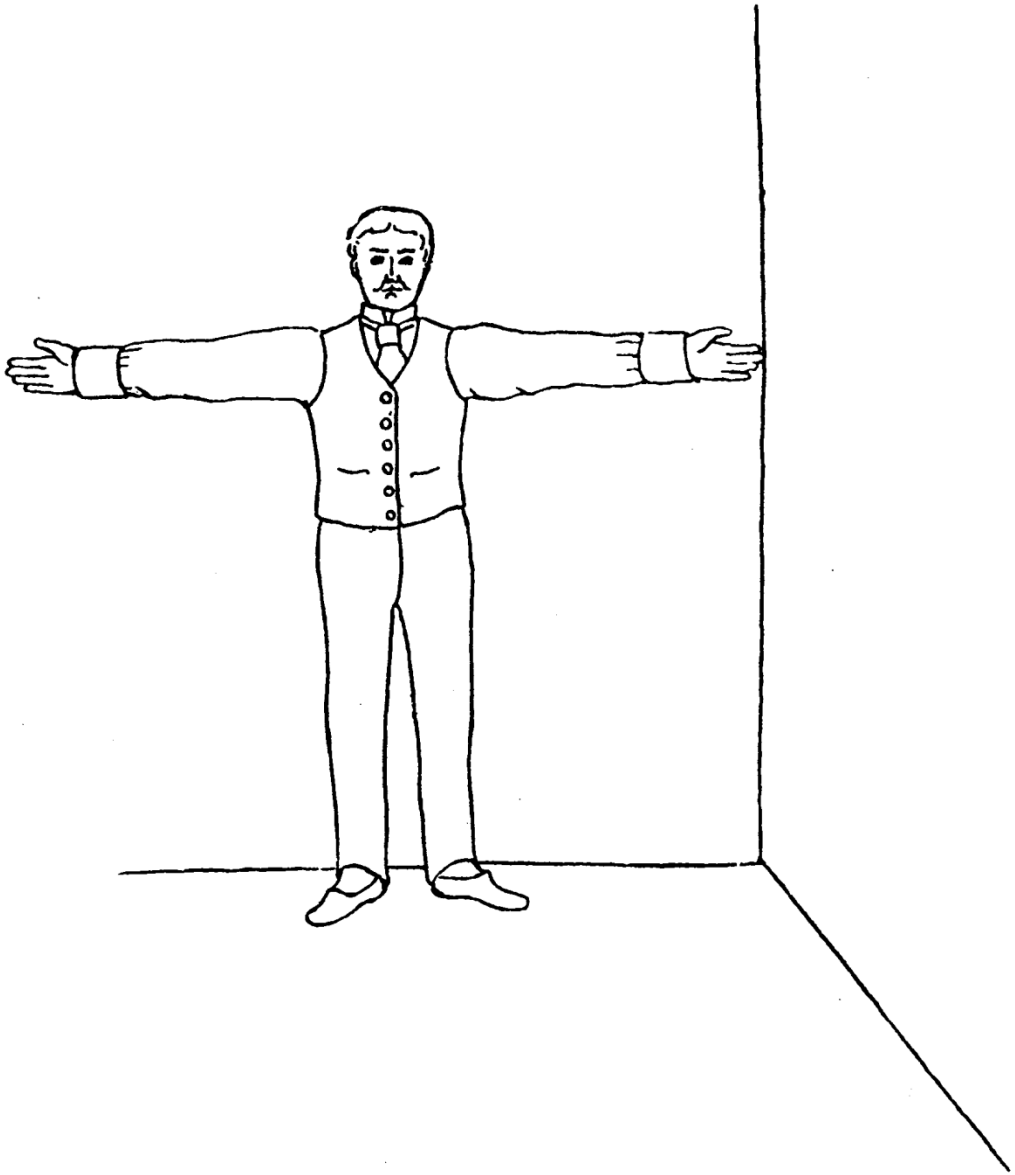


TABLE IV.

Coefficients of Heredity. Parents and Offspring.

Character	Father and		Mother and	
	Son	Daughter	Son	Daughter
Stature	·514 ± ·015	·510 ± ·013	·494 ± ·016	·507 ± ·014
Span	·454 ± ·016	·454 ± ·014	·457 ± ·016	·452 ± ·015
Forearm	·421 ± ·017	·422 ± ·015	·406 ± ·017	·421 ± ·015

From Pearson and Lee (1903) p.378

Correlation Coefficients for Direct Fraternal Heredity.

Character	Brother and Brother	Sister and Sister	Brother and Sister	Mean
Stature	·511 ± ·028	·537 ± ·022	·553 ± ·013	·534
Span	·549 ± ·026	·555 ± ·021	·525 ± ·013	·543
Forearm	·491 ± ·029	·507 ± ·023	·440 ± ·015	·479
Mean	·517	·533	·506	·519
Eye Colour*	·517 ± ·020	·446 ± ·023	·462 ± ·022	·475
Total mean	·517	·511	·495	·508

From Pearson and Lee (1903) p.387

Assortative Mating. Based on 1000 *to* 1050 *Cases of Husband and Wife.*

	Husband's Character	Wife's Character	Correlation and Probable Error	Symbol
Direct	Stature	Stature	$\cdot 2804 \pm \cdot 0189$	r_{12}
	Span	Span	$\cdot 1989 \pm \cdot 0204$	r_{34}
	Forearm	Forearm	$\cdot 1977 \pm \cdot 0205$	r_{56}

From Pearson and Lee (1903) p. 373

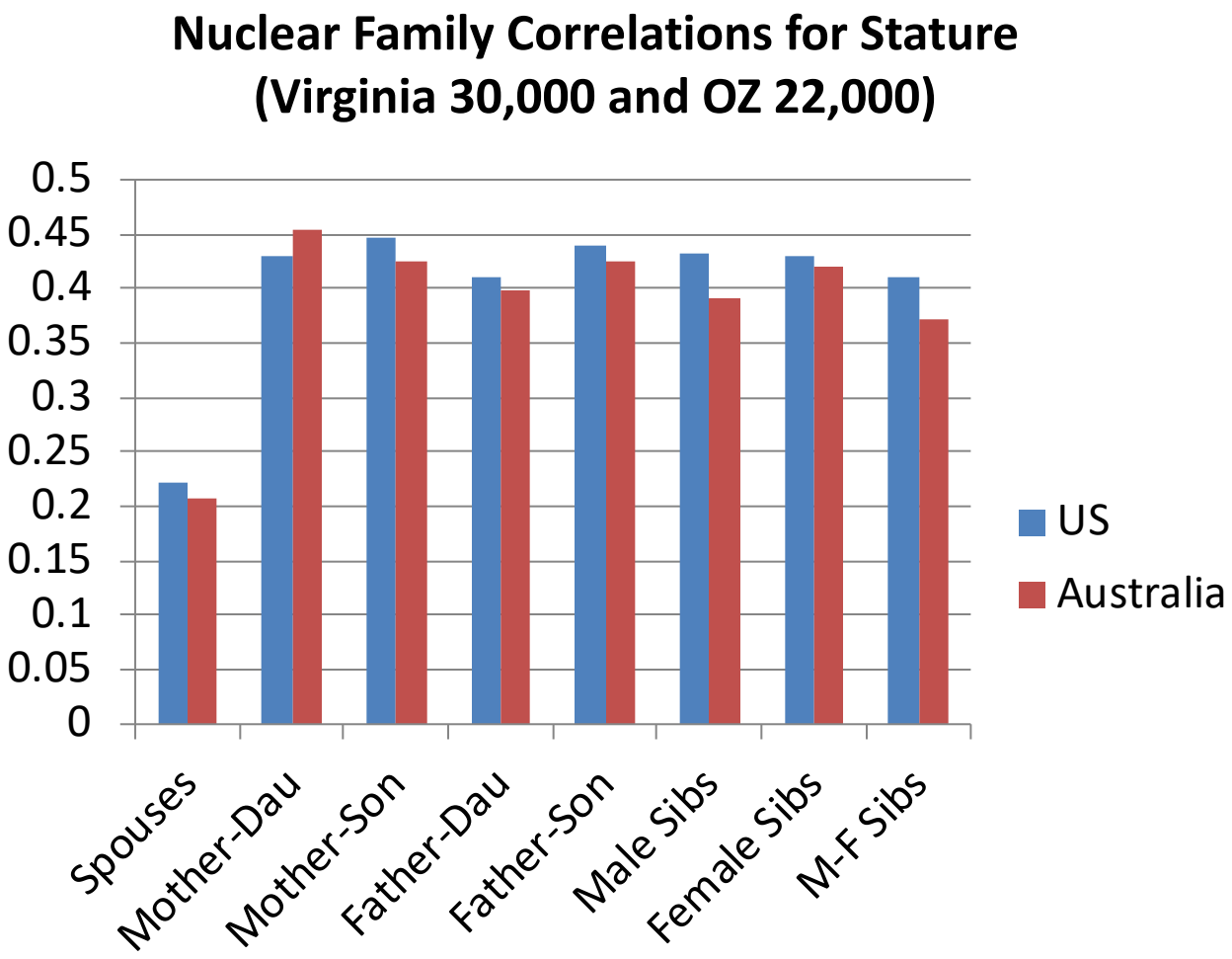
Modern Data

The Virginia 30,000 (N=29691)

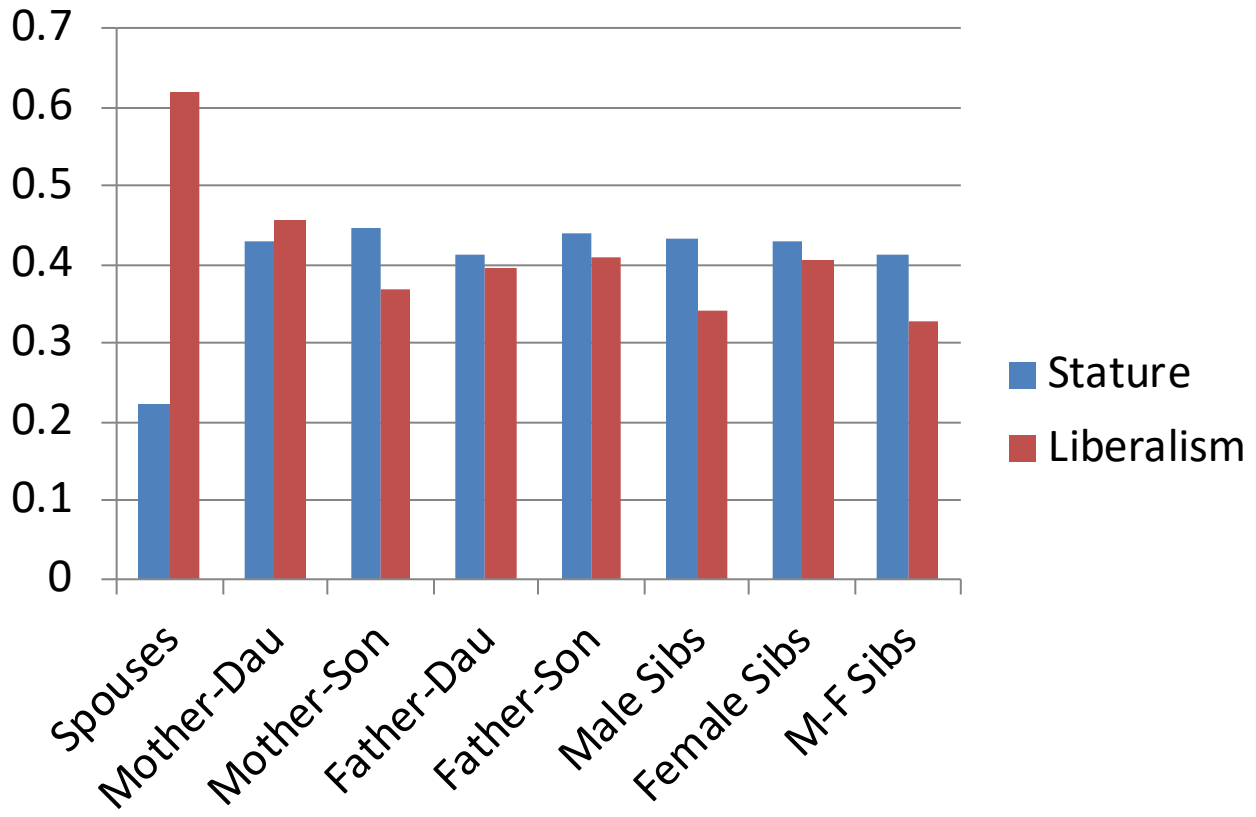
The Australia 22,000 (N=20480)

Overall sample sizes

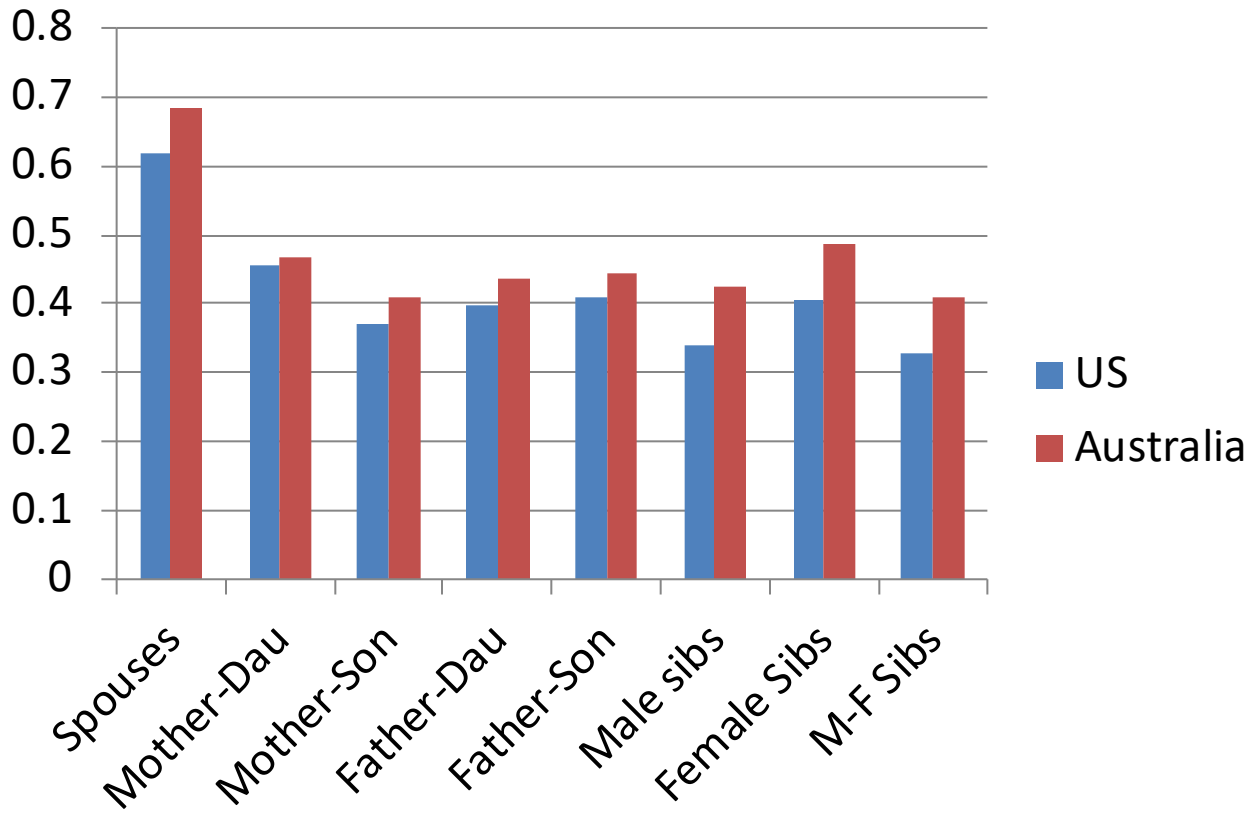
Relationship	# of pairs
Parent-offspring	25018
Siblings	18697
Spouses	8287
DZ Twins	5120
MZ Twins	4623


Nuclear Family Correlations for Stature
(Virginia 30,000 and OZ 22,000)
0.5
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
Spouses
Mother-Dau
Mother-Son
Father-Dau
Father-Son
Male Sibs
Female Sibs
M-F Sibs
US
Australia

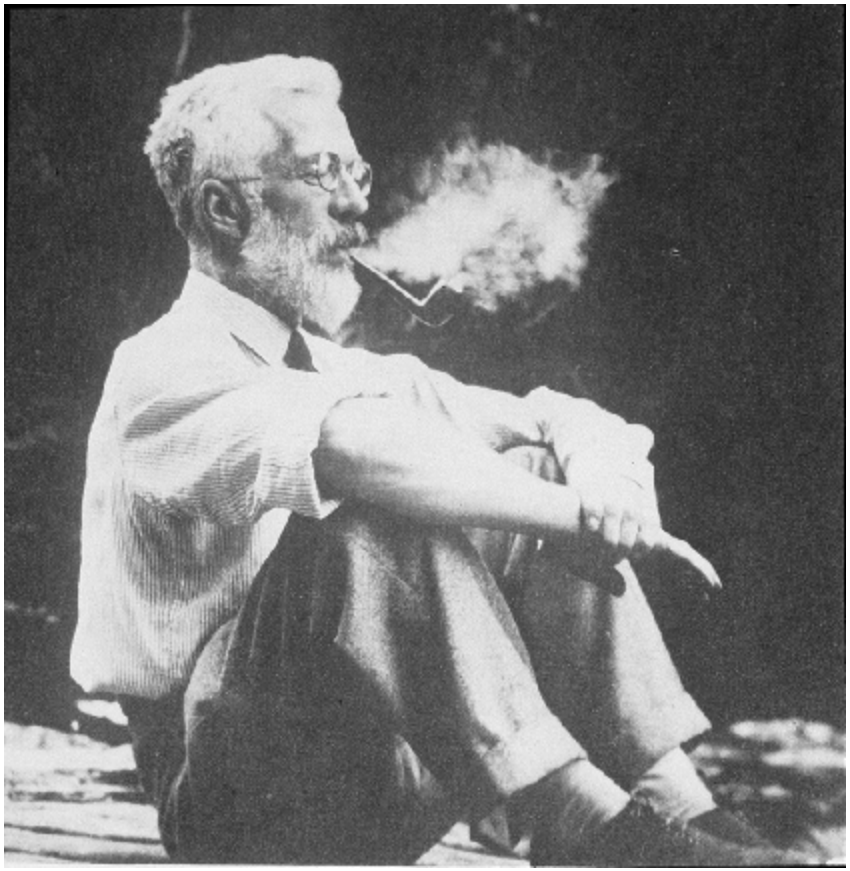
Nuclear Family Correlations for Stature and Liberalism/Conservatism (Virginia 30,000)



Nuclear Family Correlations for Liberalism/Conservatism (Virginia 30,000 and Australia 22,000)

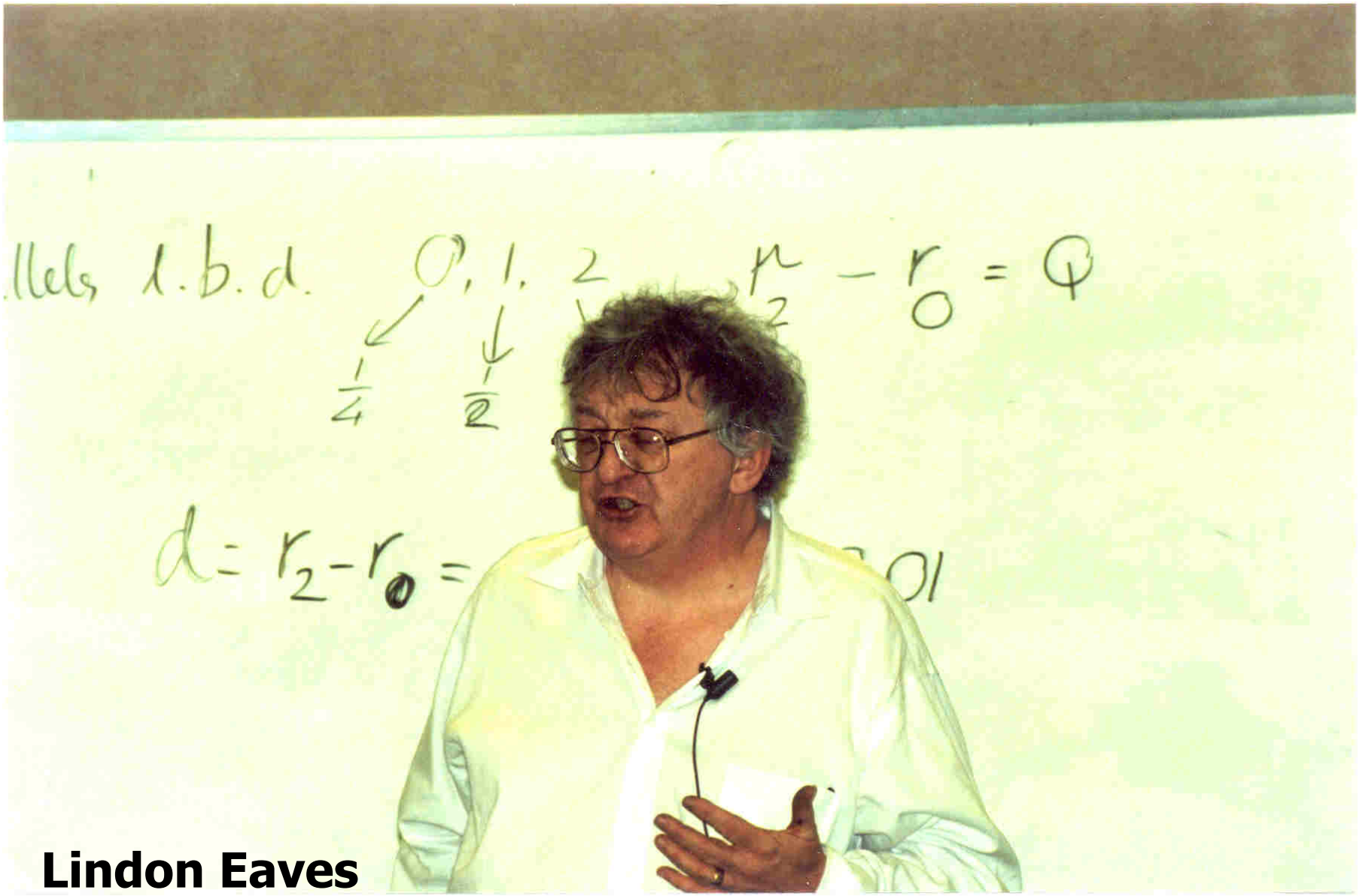


Ronald Fisher (1890-1962)

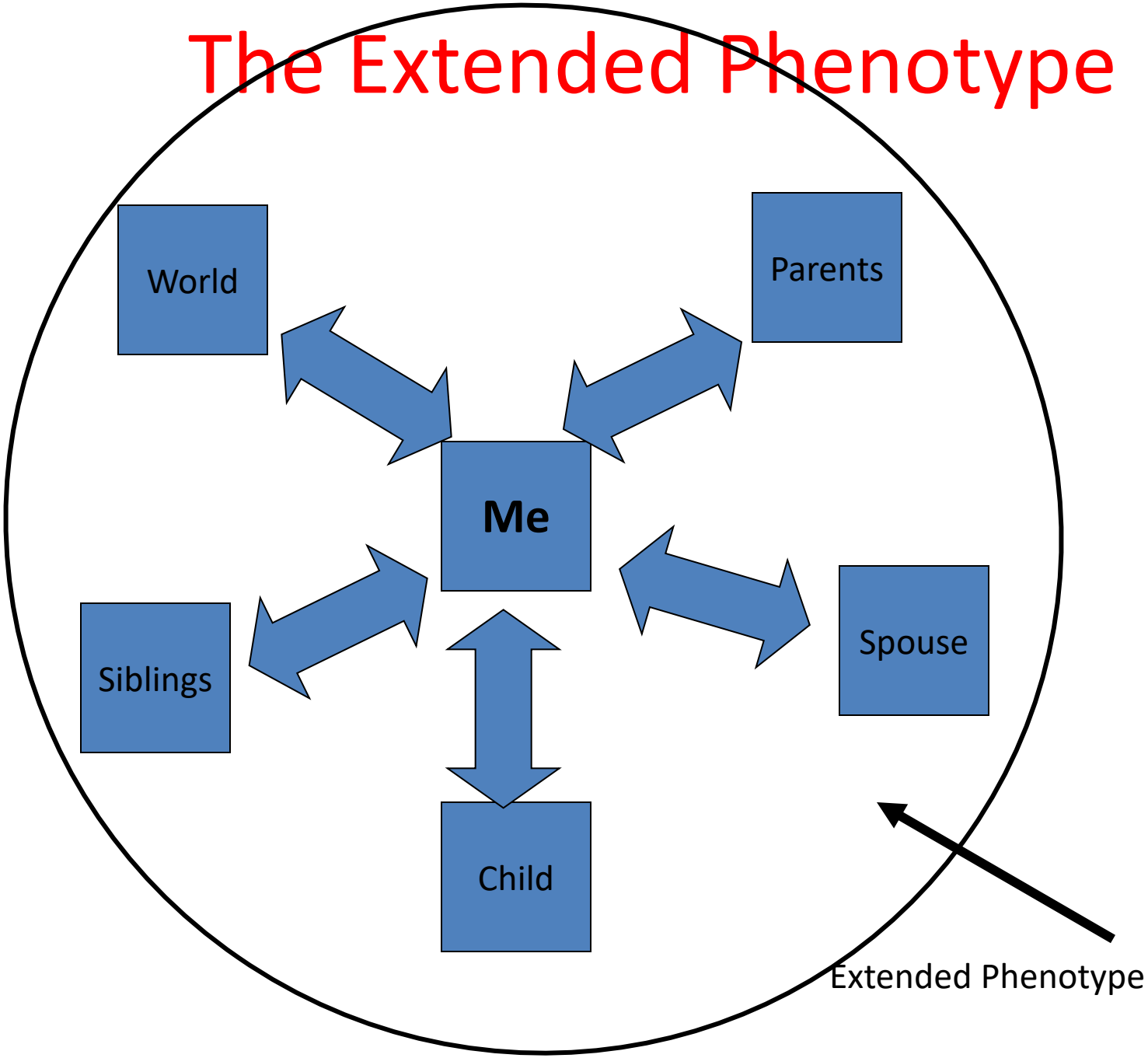


1918: On the Correlation Between Relatives on the Supposition of Mendelian Inheritance
1921: Introduced concept of “likelihood”
1930: The Genetical Theory of Natural Selection
1935: The Design of Experiments

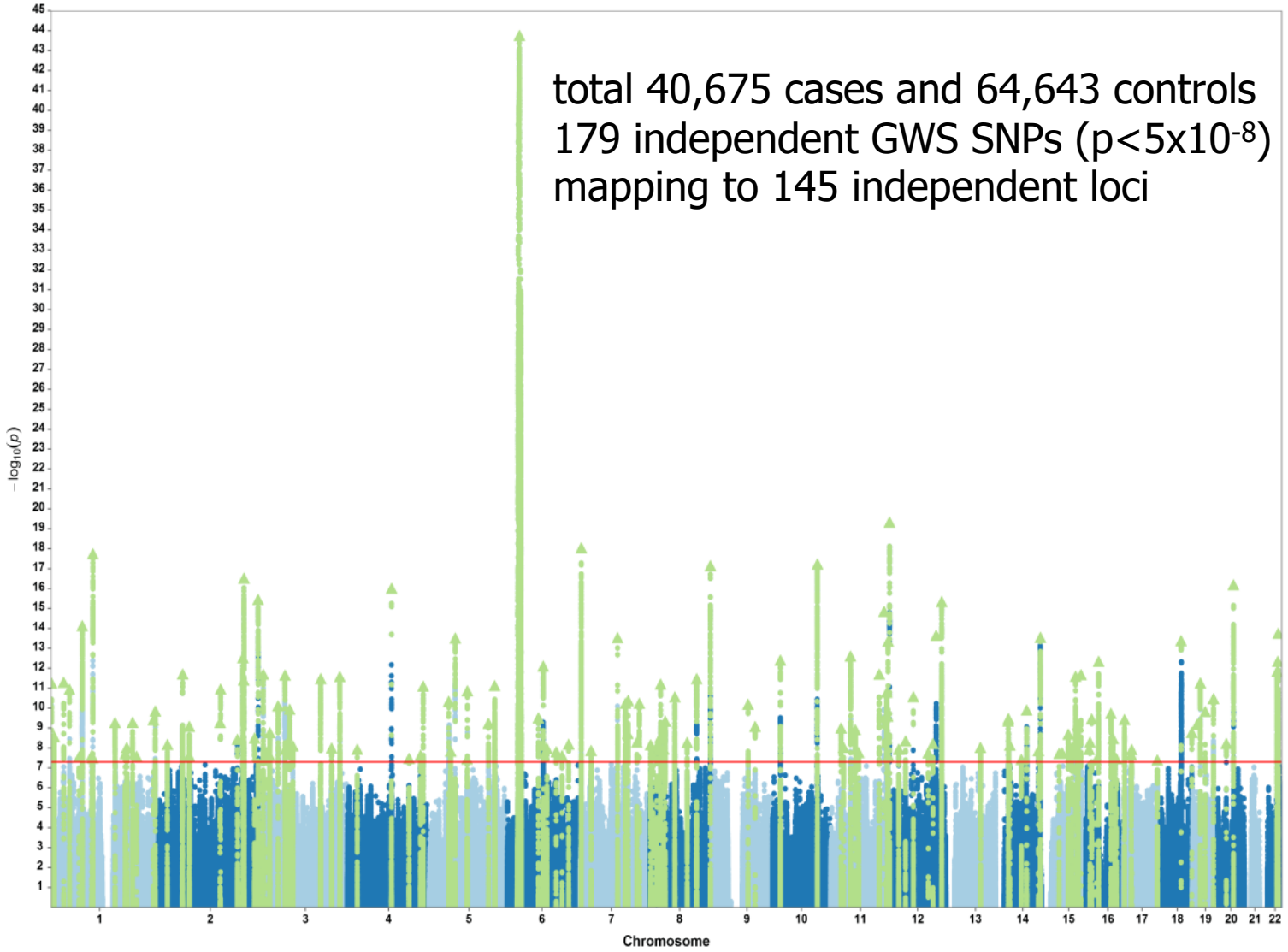
Fisher developed mathematical theory that reconciled Mendel's work with Galton and Pearson's correlations


Lindon Eaves

The Extended Phenotype

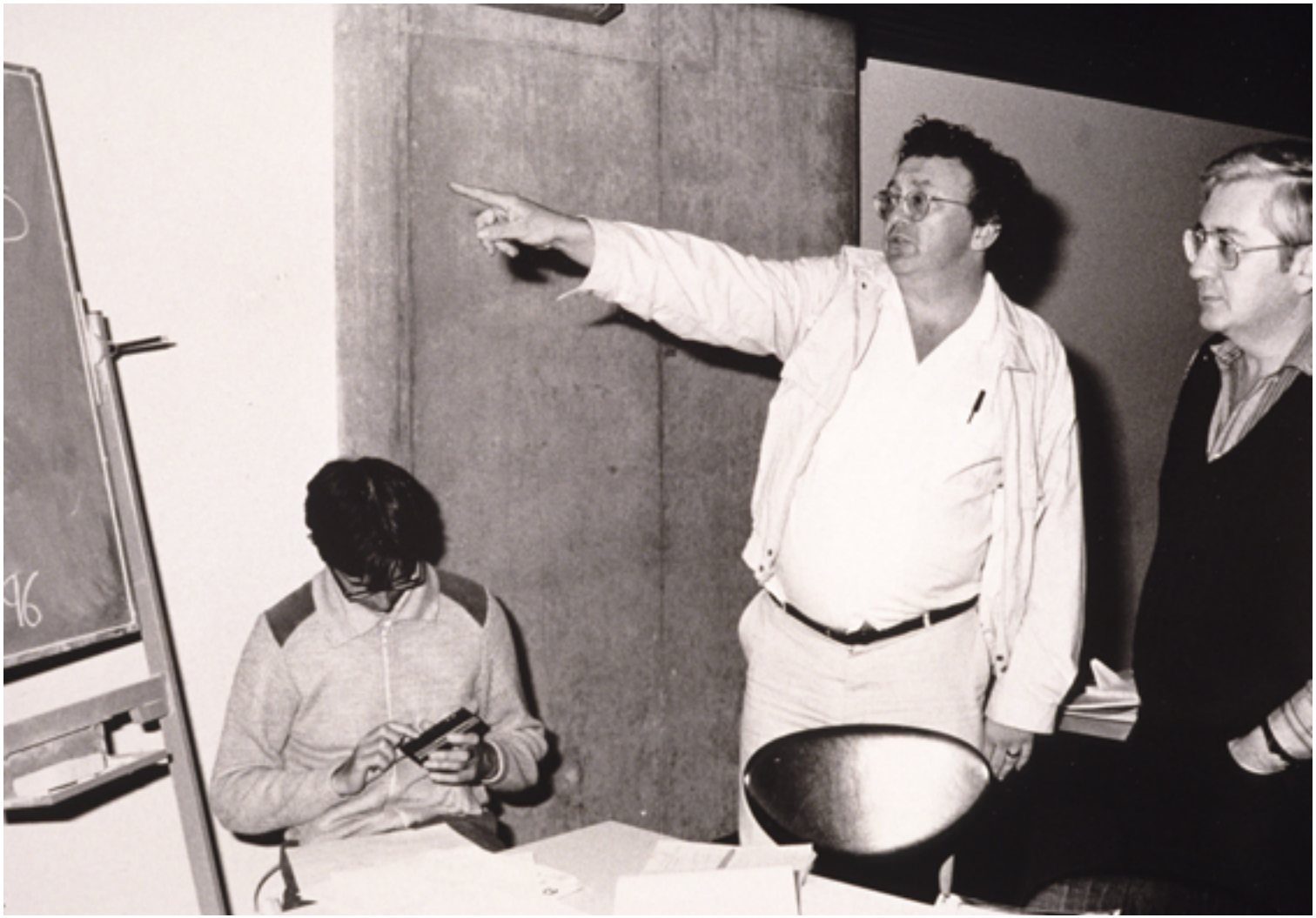


Schizophrenia GWAS associations from the meta-analysis of CLOZUK and PGC



Pardinas et al. *Nature Genetics* 50: 381–389 (2018)

Lindon Eaves



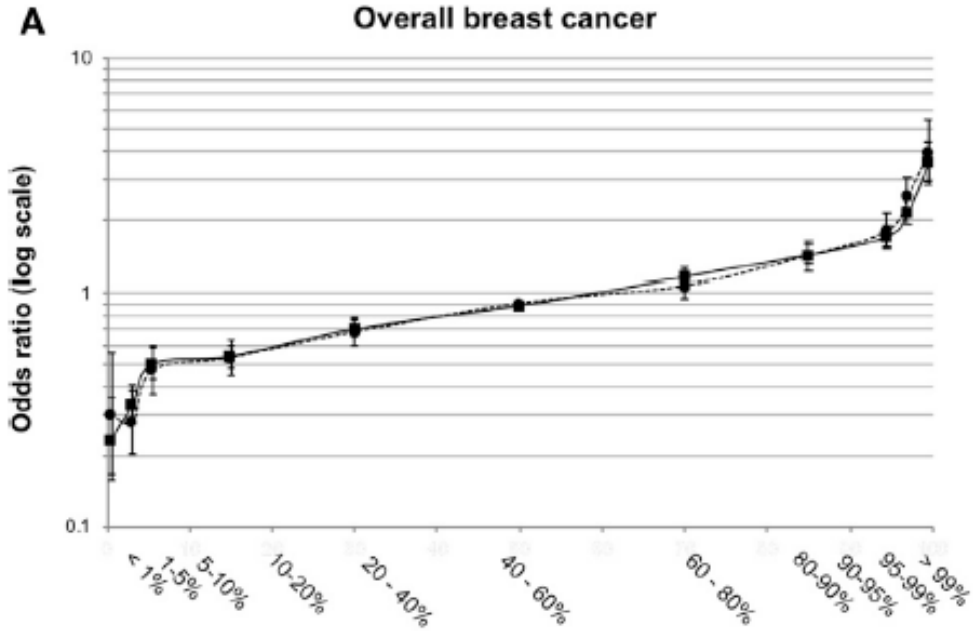
Andrew Heath

David Fulker

Polygenic Risk Scores for Prediction of Breast Cancer and Breast Cancer Subtypes

Nasim Mavaddat,[1,*] Kyriaki Michailidou,[1,2] Joe Dennis,[1] Michael Lush,[1] Laura Fachal,[3] Andrew Lee,[1]

Compared with women in the middle quintile, those in the highest 1% of risk had 4.37- and 2.78-fold risks, and those in the lowest 1% of risk had 0.16- and 0.27-fold risks, of developing ER-positive and ER-negative disease, respectively. This PRS is a powerful and reliable predictor of breast cancer risk that may improve breast cancer prevention programs.



American Journal of Human Genetics 104, 21–34, January 3, 2019