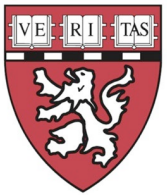
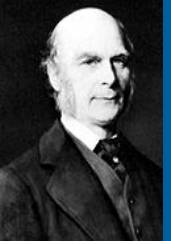
LD Score: theory

Benjamin Neale, Ph.D.
Analytic and Translational Genetics Unit, MGH
Stanley Center for Psychiatric Research & Program in Medical and Population Genetics, Broad Institute





Francis Galton
Twin and family studies

RATE OF REGRESSION IN HEREDITARY STATURE.
Fig. (a.)

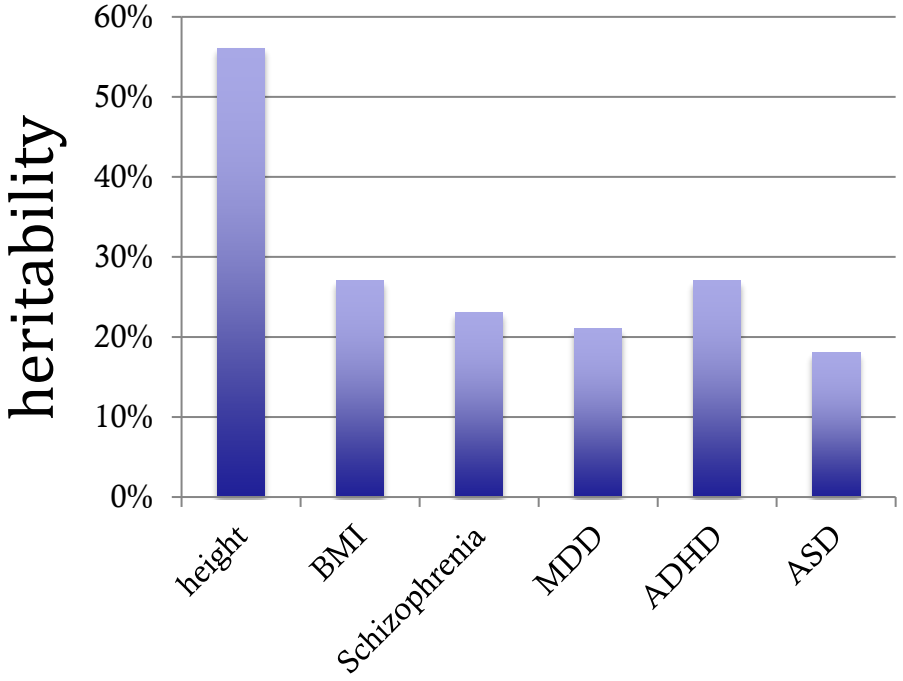
- Relatives are more similar

Meta-analysis of the heritability of human traits based on fifty years of twin studies

Tinca J C Polderman[1,10], Beben Benyamin[2,10], Christiaan A de Leeuw[1,3], Patrick F Sullivan[4–6], Arjen van Bochoven[7], Peter M Visscher[2,8,11] & Danielle Posthuma[1,9,11]

Average estimate of heritability 49%
69% of twin studies support a purely additive genetic model

GREML/GCTA



- Use estimated genetic similarity

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]

ANALYSIS

nature genetics

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]

ARTICLE

Estimating Missing Heritability for Disease from Genome-wide Association Studies

Sang Hong Lee,[1] Naomi R. Wray,[1] Michael E. Goddard,[2,3] and Peter M. Visscher[1,*]

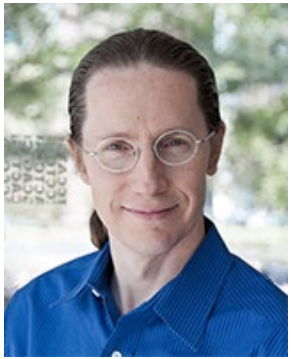
LD Score regression

With thanks

Brendan Bulik-Sullivan

Hilary Finucane

Po-Ru Loh



Mark Daly

Alkes Price

How does LD shape association?

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

Affiliations | Contributions | Corresponding author

Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

How does LD shape association?

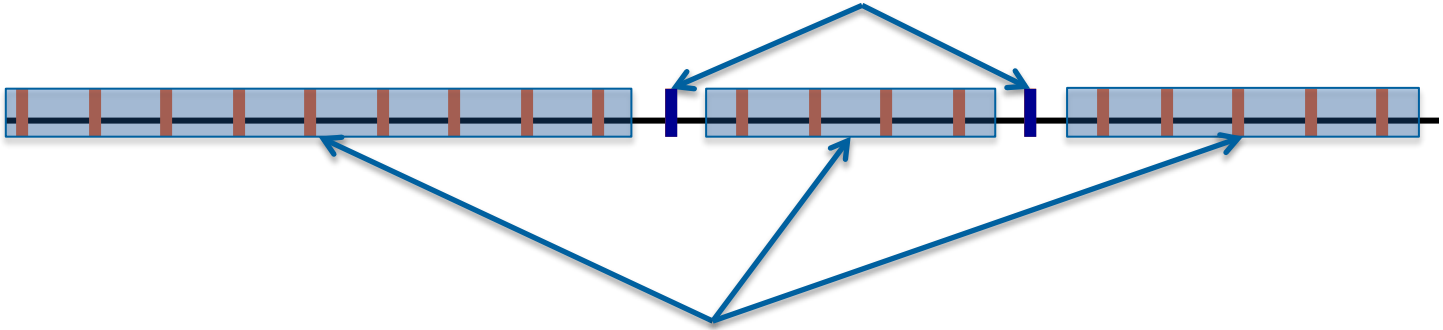
LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

Affiliations | Contributions | Corresponding author

Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

Lonely SNPs [no LD]



LD blocks

How does LD shape association?

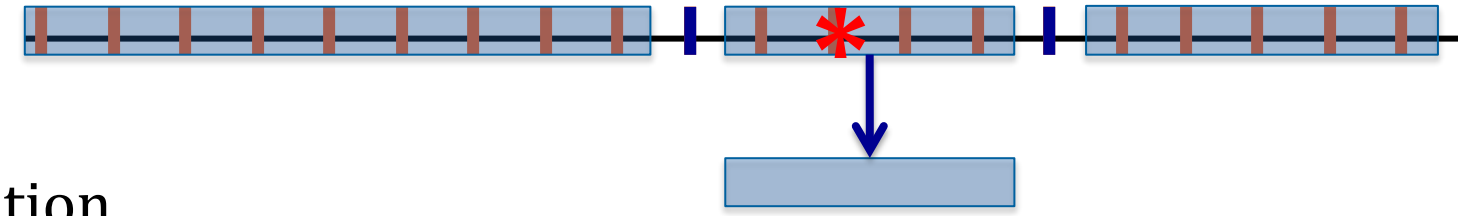
- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

Affiliations | Contributions | Corresponding author

Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015



Association

All markers correlated with a causal variant show association

How does LD shape association?

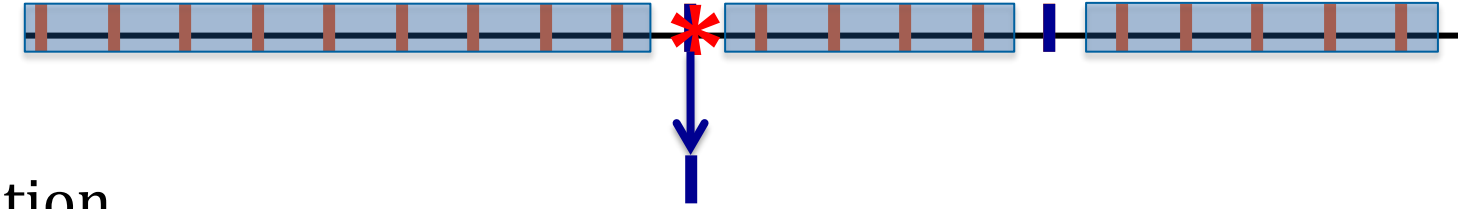
- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

Affiliations | Contributions | Corresponding author

Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015



Association

Lonely SNPs only show association if they are causal

What happens under polygenicity?

- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

Affiliations | Contributions | Corresponding author

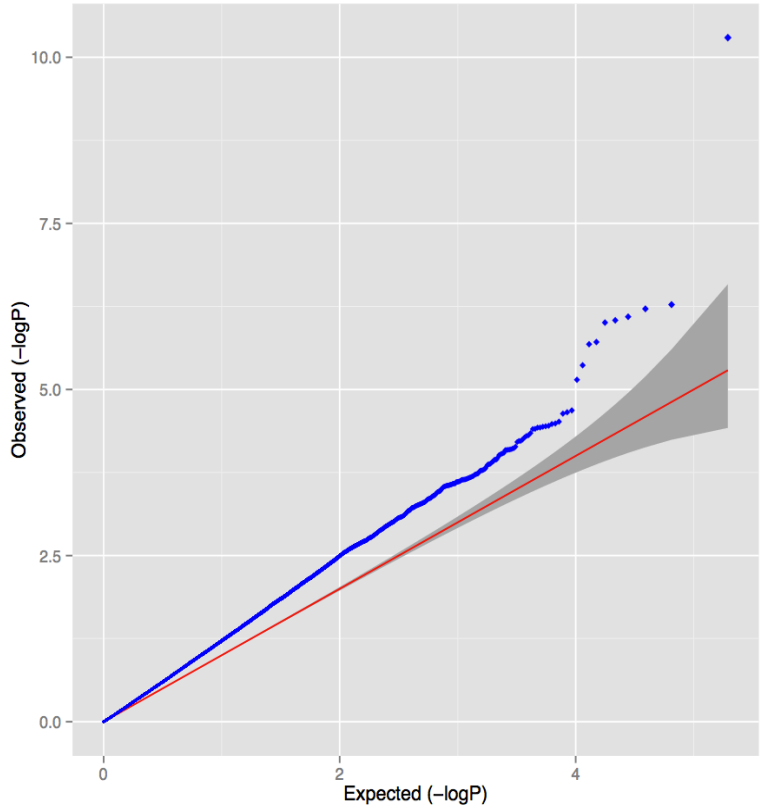
Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

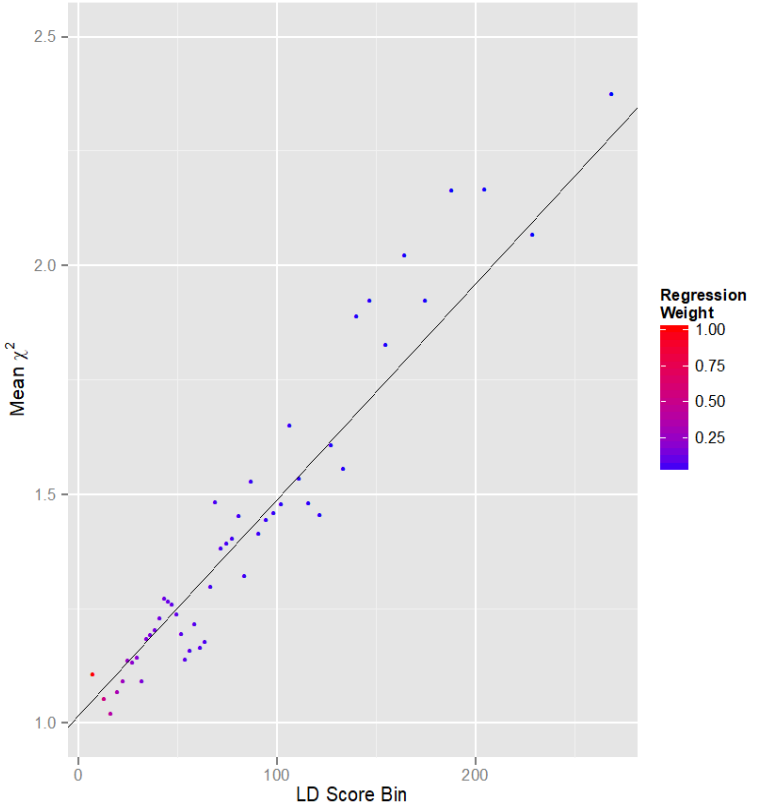
Assuming a uniform prior, we see SNPs with more LD friends showing more association

The more you tag, the more likely you are to tag a causal variant

Simulated polygenic architecture

Lambda = 1.30 LD score intercept = 1.02





What happens under stratification?

- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

Brendan K Bulik-Sullivan, Po-Ru Loh, Hilary K Finucane, Stephan Ripke, Jian Yang, Schizophrenia Working Group of the Psychiatric Genomics Consortium, Nick Patterson, Mark J Daly, Alkes L Price & Benjamin M Neale

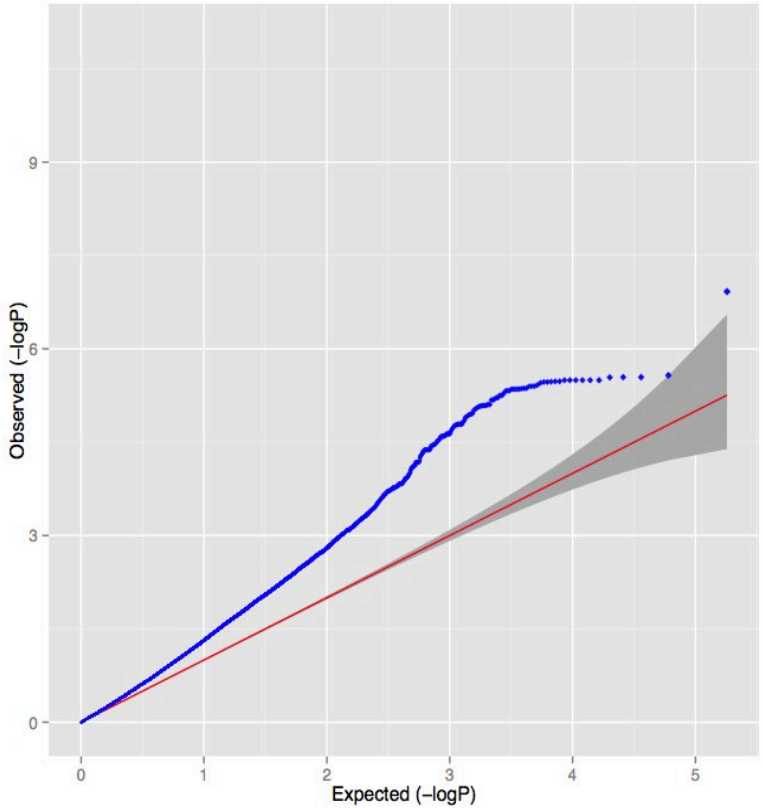
Affiliations | Contributions | Corresponding author

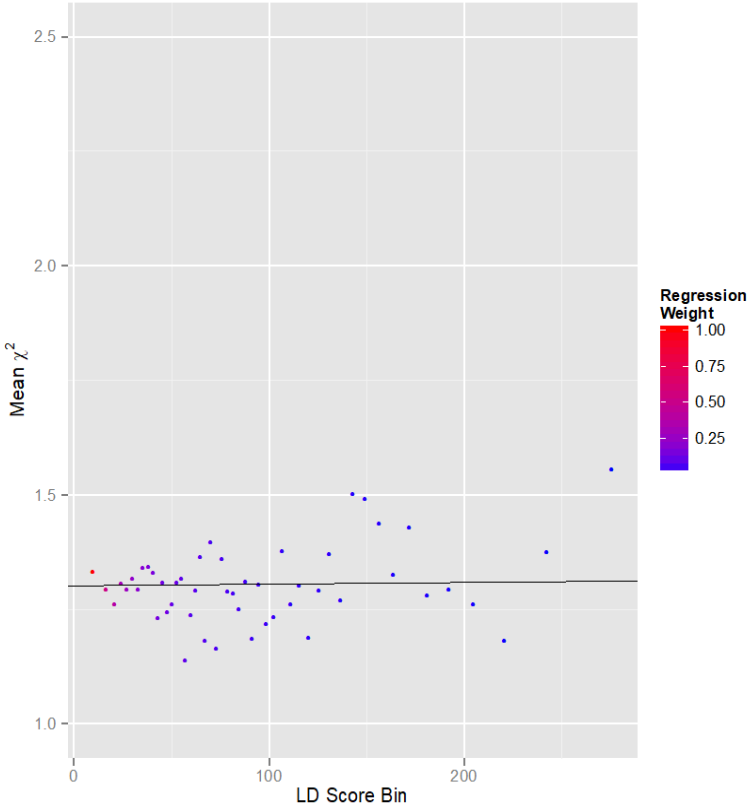
Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

Under pure drift we expect LD to have no relationship to differences in allele frequencies between populations

UK controls versus Sweden controls

Lambda = 1.30 LD score intercept = 1.32





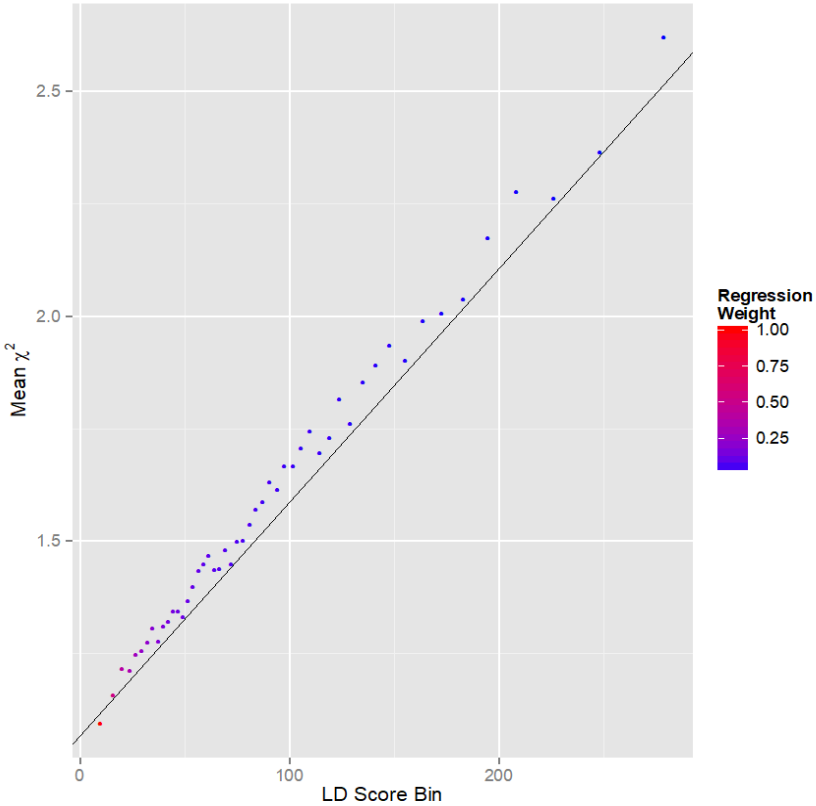
PGC Schizophrenia

Lambda = 1.48

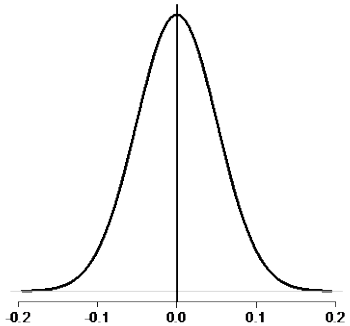
Intercept = 1.06

Slope p-value < 10^{-300}

Overwhelming majority of inflation is consistent with polygenic architecture



LD Score regression



Draw polygenic effects from
N (0, n/m^2), var =

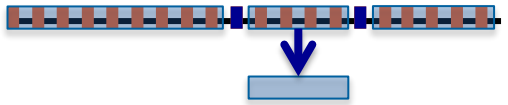
What is the E[χ^2] for variant *j*?

$$E[\chi_j^2] = 1 + Na + \frac{h_g^2 N}{M} l_j$$ New estimator of heritability

where N=sample size, M=# of SNPs, a=inflation due to confounding, h^2g is heritability (total obs.) and l_j is the *LD Score*

Bulik-Sullivan et al. Nature Genetics 2015
Yang et al. EJHG 2011

$$l_j = \sum_{k \neq j} r_{jk}^2$$



Brainstorm Project

Analysis of shared heritability in common disorders of the brain

Verneri Anttila, Brendan Bulik-Sullivan, Hilary Kiyo Finucane, Jose Bras, Laramie Duncan, Valentina Escott-Price, Guido Falcone, Padhraig Gormley, Rainer Malik, Nikolaos Patsopoulos, Stephan Ripke, Raymond Walters, Zhi Wei, Dongmei Yu, Phil Lee, IGAP consortium, IHGC consortium, ILAE Consortium on Complex Epilepsies, IMSGC consortium, IPDGC consortium, METASTROKE and ICH Studies of the ISGC, ADHD Working Group of the PGC, Anorexia Nervosa Working Group of the PGC, ASD Working Group of the PGC, Bipolar Disorders Working Group of the PGC, Major Depressive Disorder Working Group of the PGC, OCD and TS Working Group of the PGC, Schizophrenia Working Group of the PGC, Gerome Breen, Cynthia Bulik, Mark Daly, Martin Dichgans, Stephen Faraone, Rita Guerreiro, Peter Holmans, Kenneth Kendler, Bobby Koeleman, Carol Mathews, Jeremiah Scharf, Pamela Sklar, Julie Williams, Nick Wood, Chris Cotsapas, Aarno Palotie, Jordan Smoller, Patrick Sullivan, Jonathan Rosand, Aiden Corvin, Benjamin Neale

doi: https://doi.org/10.1101/048991

Verneri Anttila

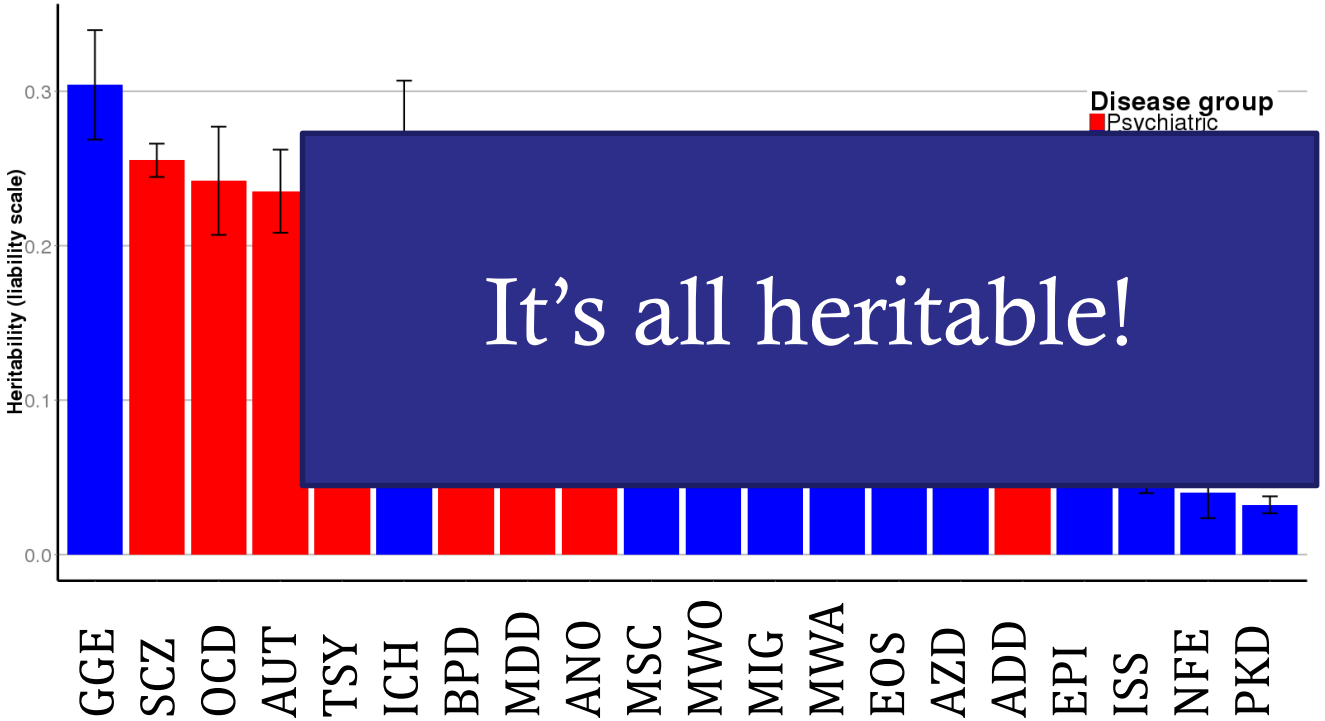
Aiden Corvin

Brendan Bulik-Sullivan
Hilary Finucane
Jonathan Rosand
Aarno Palotie
Mark Daly
Patrick Sullivan
Bobby Koeleman
Nick Wood
Julie Williams

Alessandro Biffi
Jeremiah Scharf
Kenneth Kendler
Stephan Ripke
Alkes Price
Chris Cotsapas
Padhraig Gormley
Zhi Wei
Rainer Malik

Hailiang Huang
Andrea Byrnes
Dongmei Yu
Laramie Duncan
Kai-How Farh
Namrata Gupta
Miriam Raffeld
…and many, many others
in their respective study groups

Univariate heritability from common variation



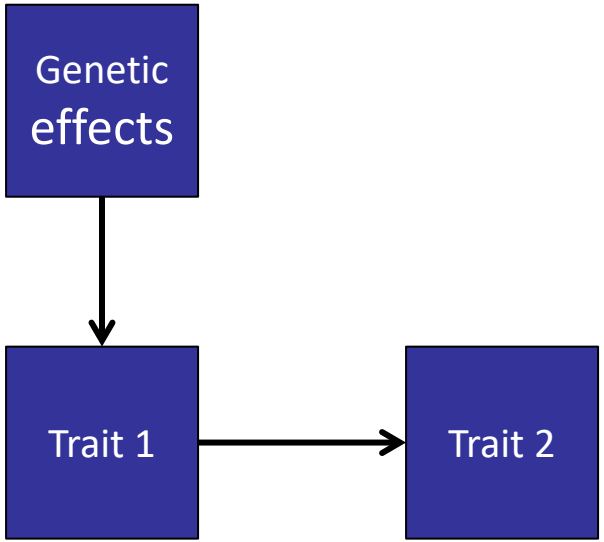
GGE = Generalized Epilepsy
SCZ = Schizophrenia
OCD = Obsessive Compulsive Disorder
AUT = Autism
TSY = Tourette's Syndrome
ICH = Intracerebral Hemorrhage
BPD = Bipolar Disorder
MDD = Major Depressive Disorder
ANO = Anorexia Nervosa
MSC = Multiple Sclerosis
MWO = Migraine without Aura
MIG = Migraine
MWA = Migraine with Aura
EOS = Early Onset Stroke
AZD = Alzheimer's Disease
ADD = Attention Deficit/Hyperactivity
EPI = Epilepsy (all)
ISS = Ischemic Stroke
NFE = Non-acquired focal epilepsy
PKD = Parkinson's Disease

Genetic Correlation Method in:

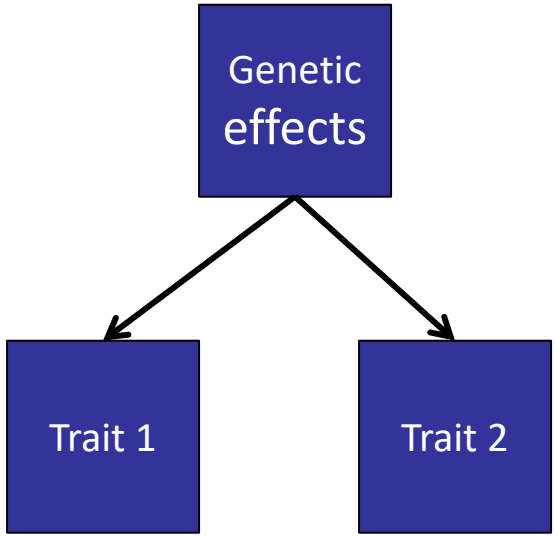
nature genetics

An atlas of genetic correlations across human diseases and traits

Potential sources of genetic correlation



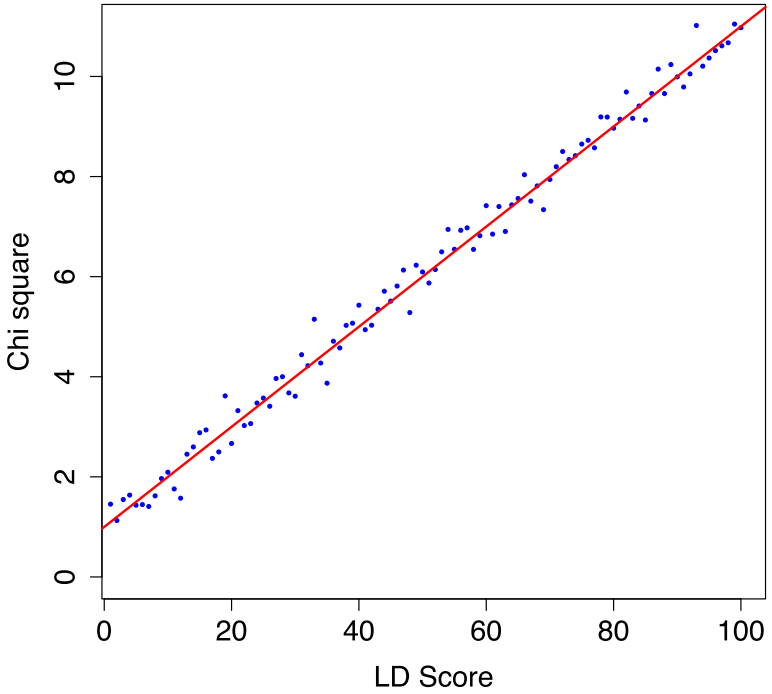
Trait 1 exerts causal effect on Trait 2



Genetic effects influence Trait 1 and Trait 2

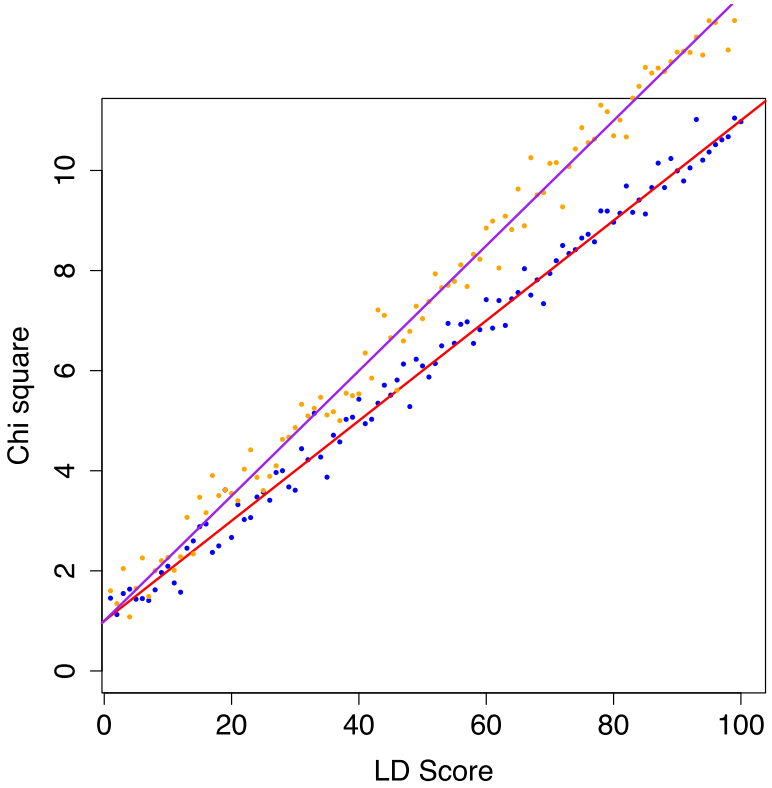
LD Score regression
Genetic correlation

Trait 1



Slope estimates heritability

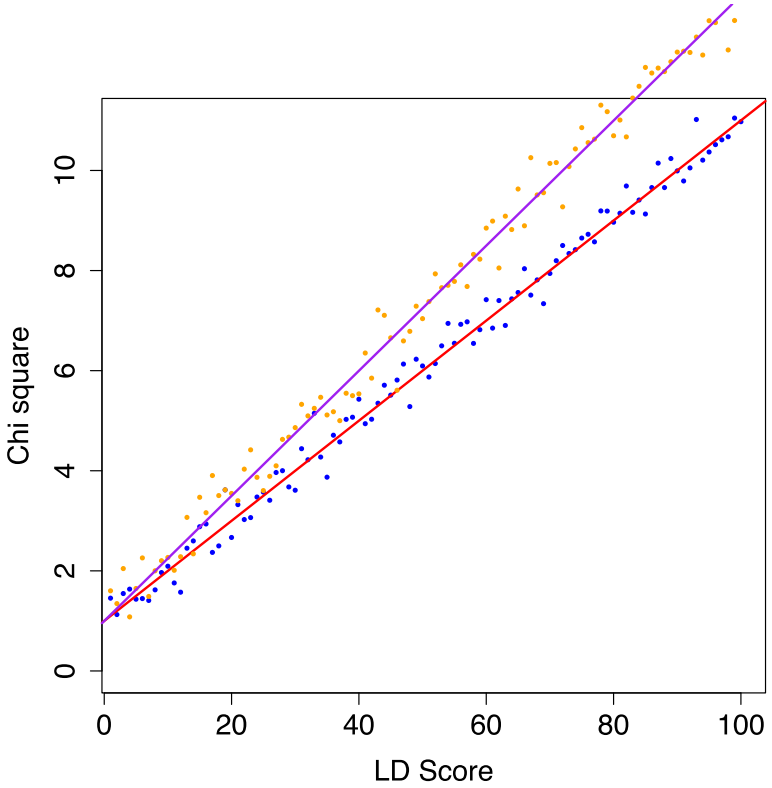
LD Score regression
Genetic correlation



Trait 1
Trait 2

We can a second trait and obtain two heritability estimates

LD Score regression
Genetic correlation

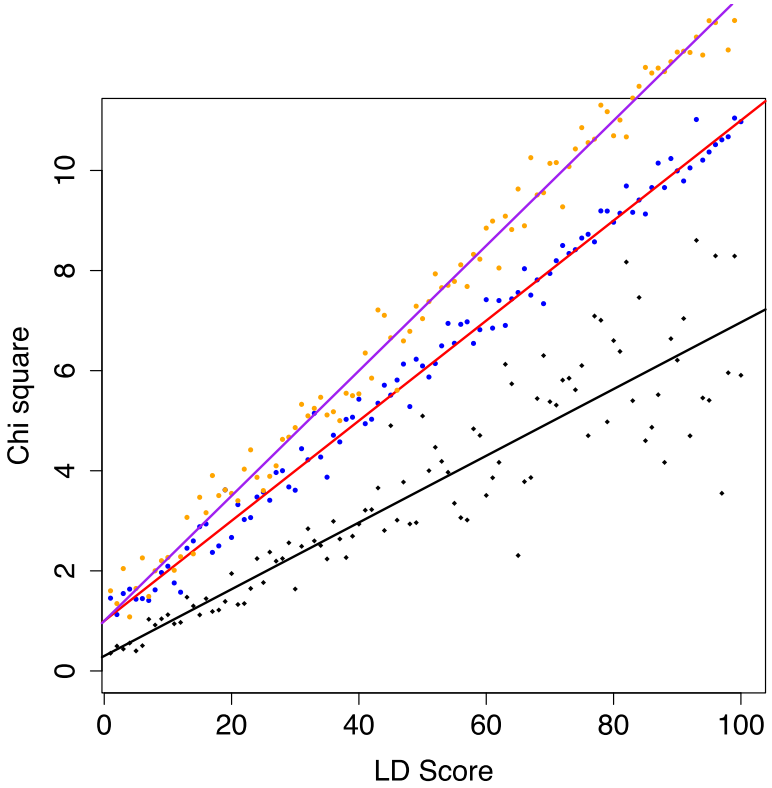


Trait 1
Trait 2

$Z^*Z = \chi^2$

So we can estimate genetic covariance from the product of the Z-scores

LD Score regression
Genetic correlation



Trait 1

Trait 2

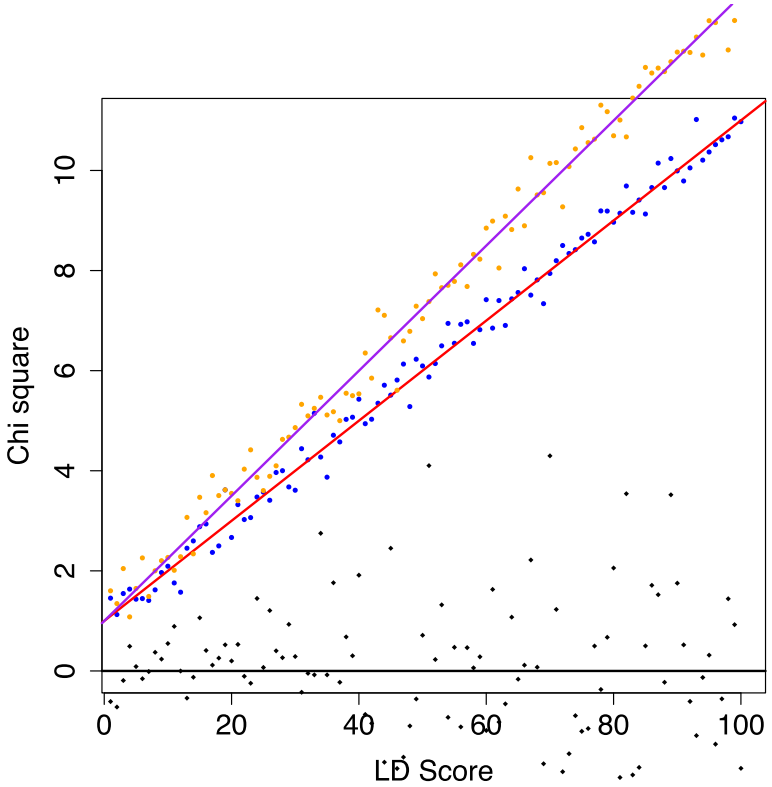
R_G

$Z^*Z = \chi^2$

So we can estimate genetic covariance from the product of the Z-scores for the two traits

$R_G = 0.5$

LD Score regression
Genetic correlation



Trait 1

Trait 2

R_G

Here $R_G = 0$

This approach is robust to sample overlap as all variants are equally inflated