Genetic correlation and LD Score Regression

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

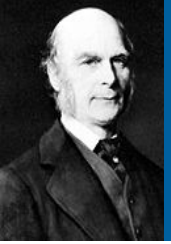
BROAD INSTITUTE

MGH 1811

CENTER FOR GENOMIC MEDICINE

STANLEY CENTER FOR PSYCHIATRIC RESEARCH

VE RI TAS



Francis Galton
Twin and family studies


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

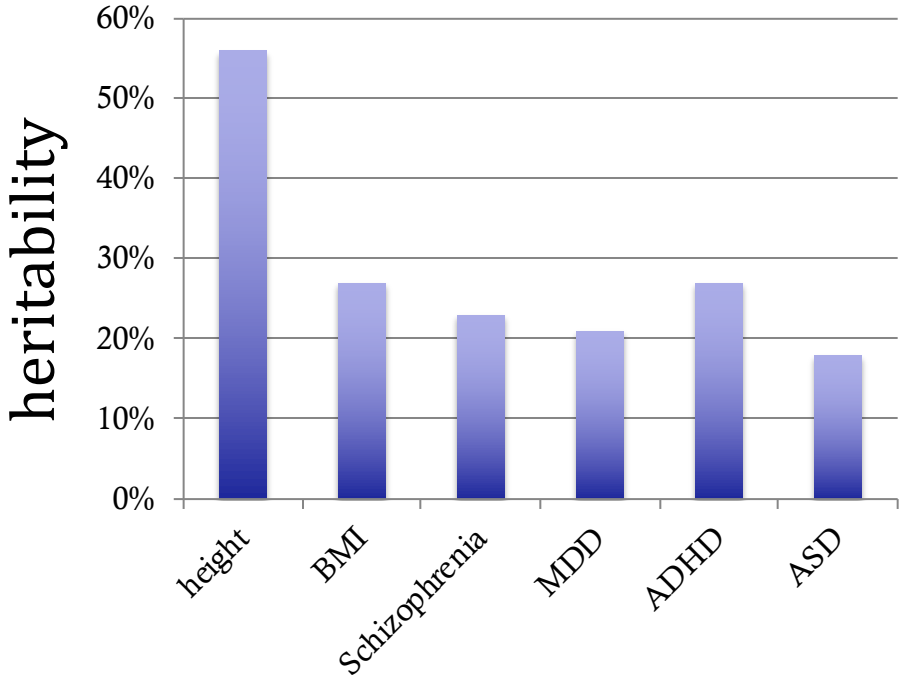
- Relatives are more similar

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

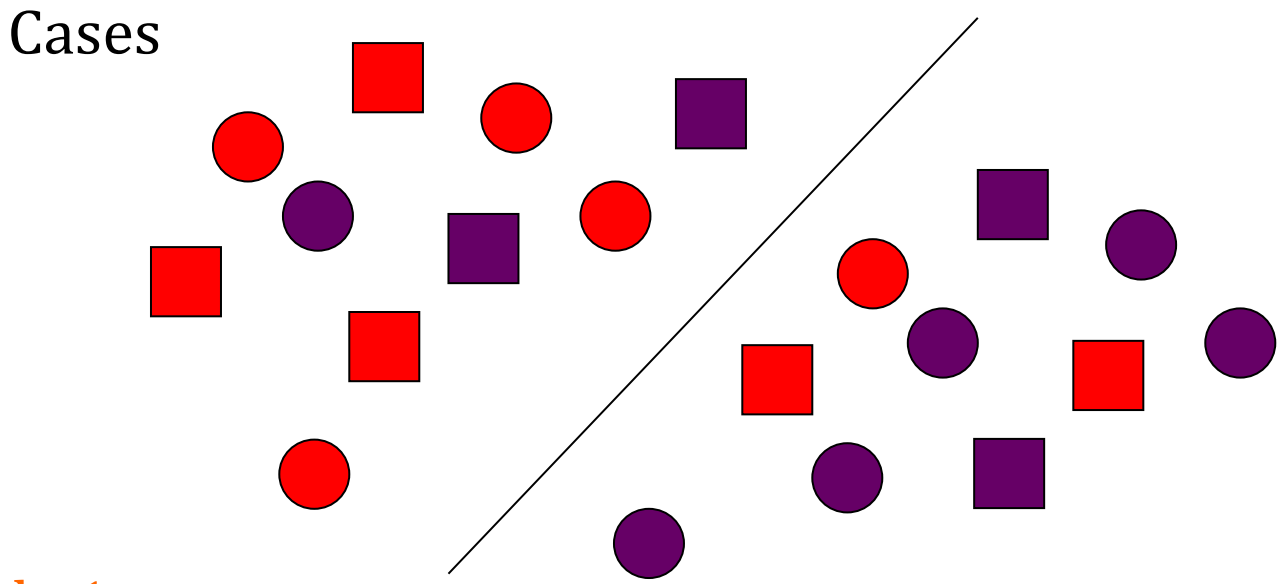
Jian Yang[1], Beben Benyamin[1], Brian P McEvoy[1], Scott Gordon[1], Anjali K Henders[1], Dale R Nyholt[1], Pamela A Madden[2], Andrew C Heath[2], Nicholas G Martin[1], Grant W Montgomery[1], Michael E Goddard[3] & Peter M Visscher[1]

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,*]

Genetic association study
Test for main effect of SNP



allele 1

allele 2

Test allele 1 frequency

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?

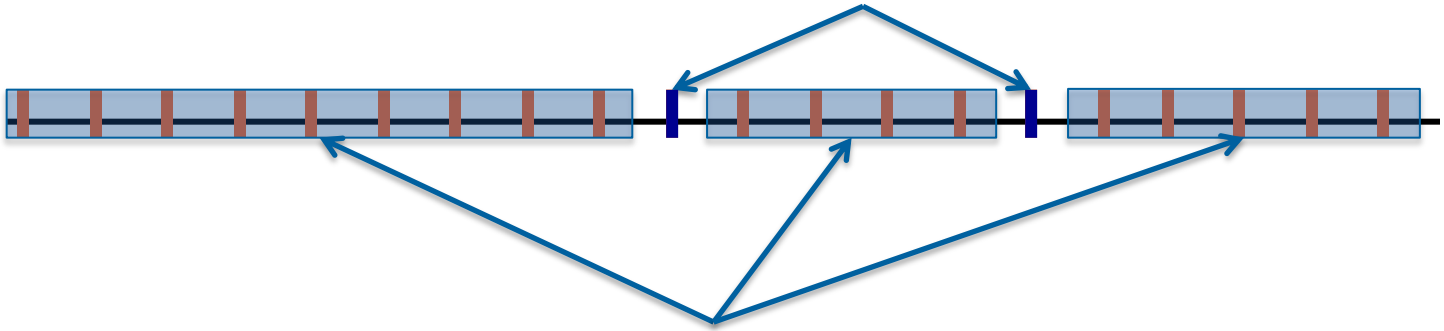
LD Score regression distinguishes confounding from polygenicity in genome-wide association studies

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?

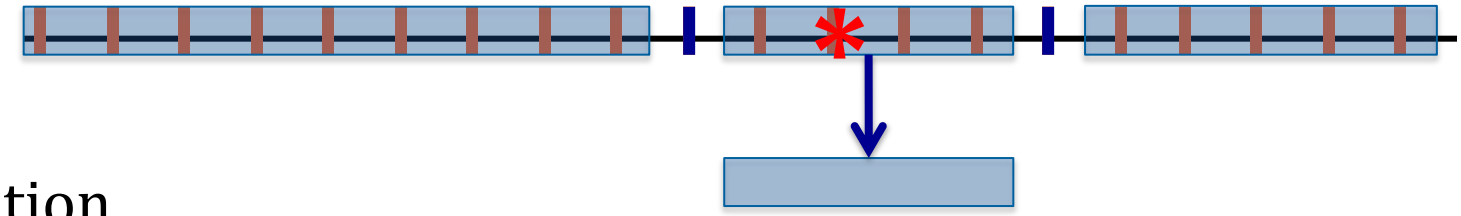
- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

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?

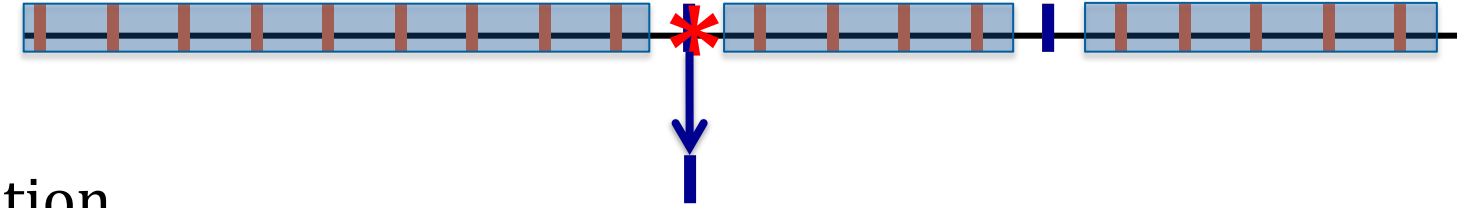
- Lonely SNPs [no LD]
- LD blocks
- * Causal variants

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

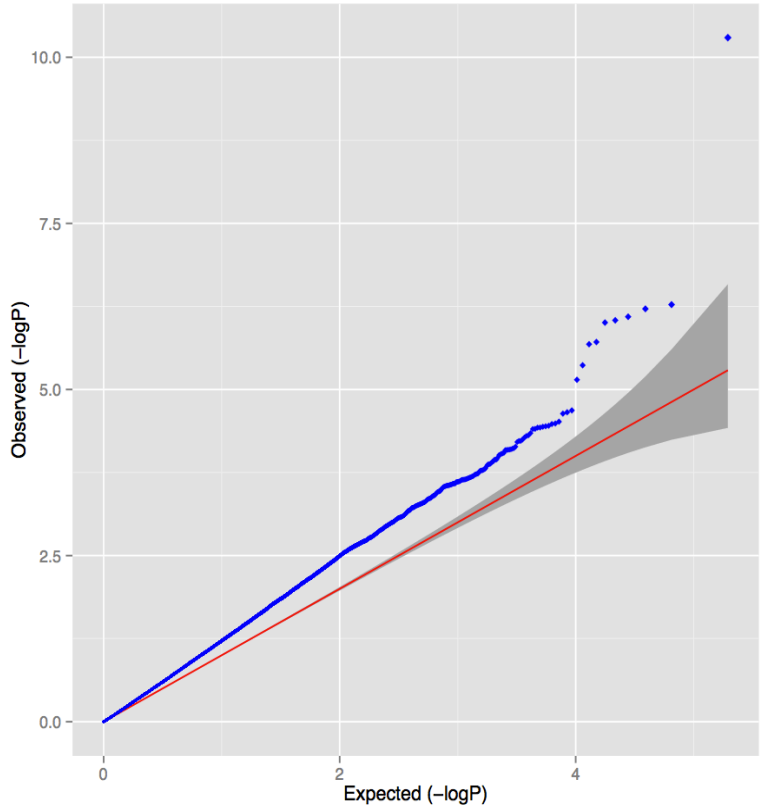
Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

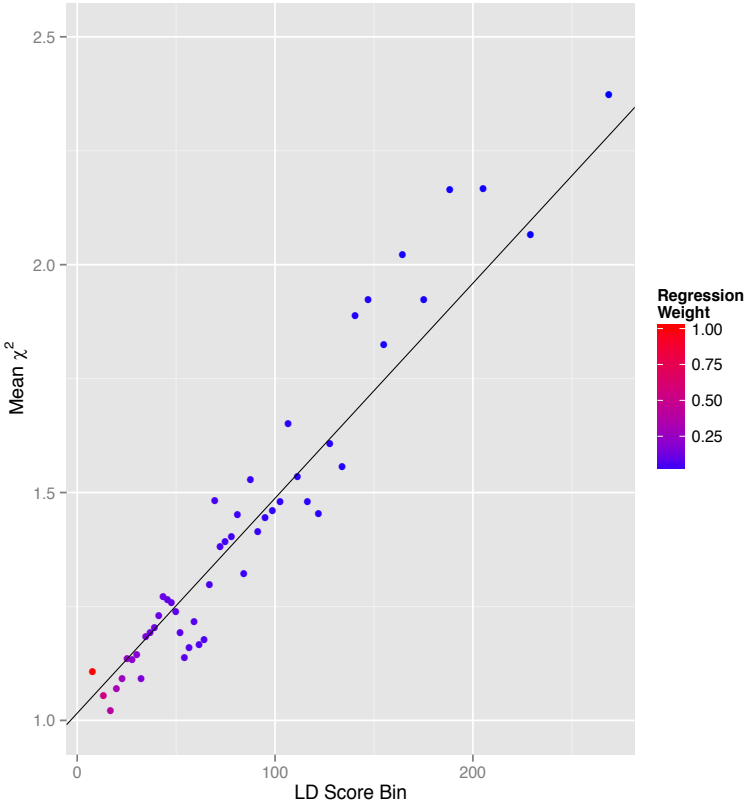
Assuming a uniform prior, we see SNPs with more LD friends showing more association

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

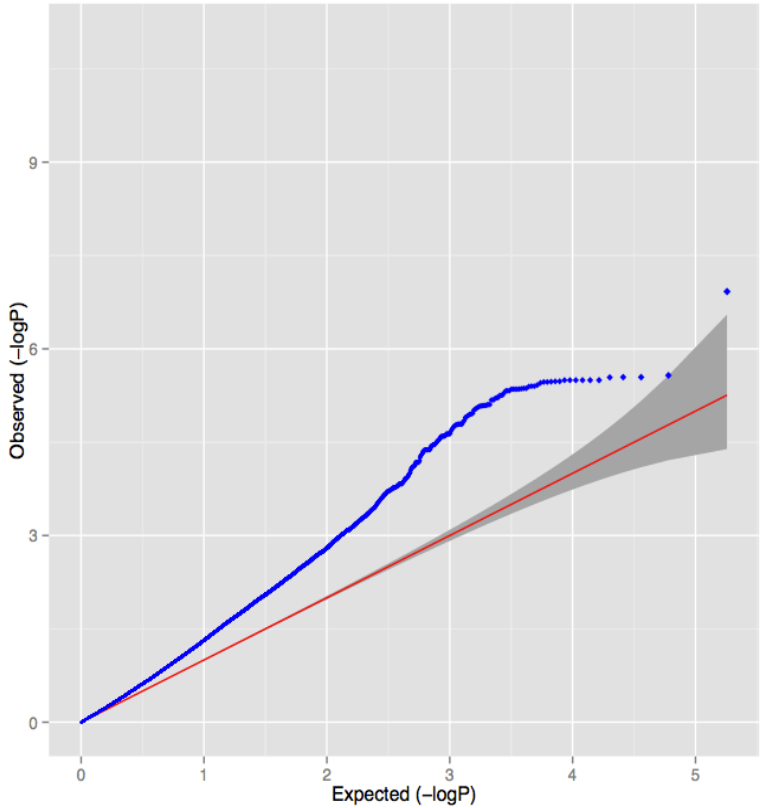
Affiliations | Contributions | Corresponding author

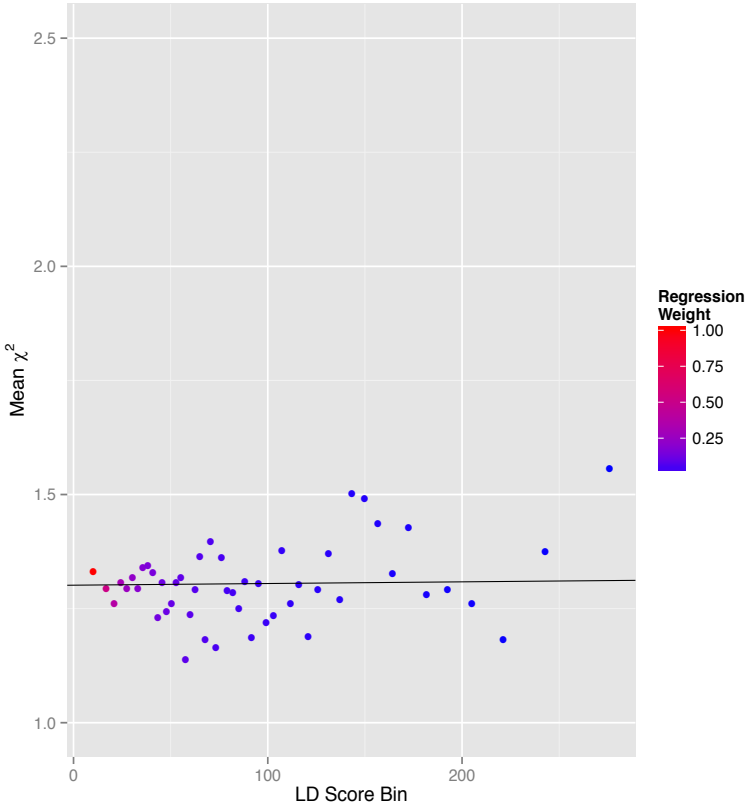
Nature Genetics **47**, 291–295 (2015) | doi:10.1038/ng.3211
Received 07 March 2014 | Accepted 07 January 2015 | Published online 02 February 2015

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





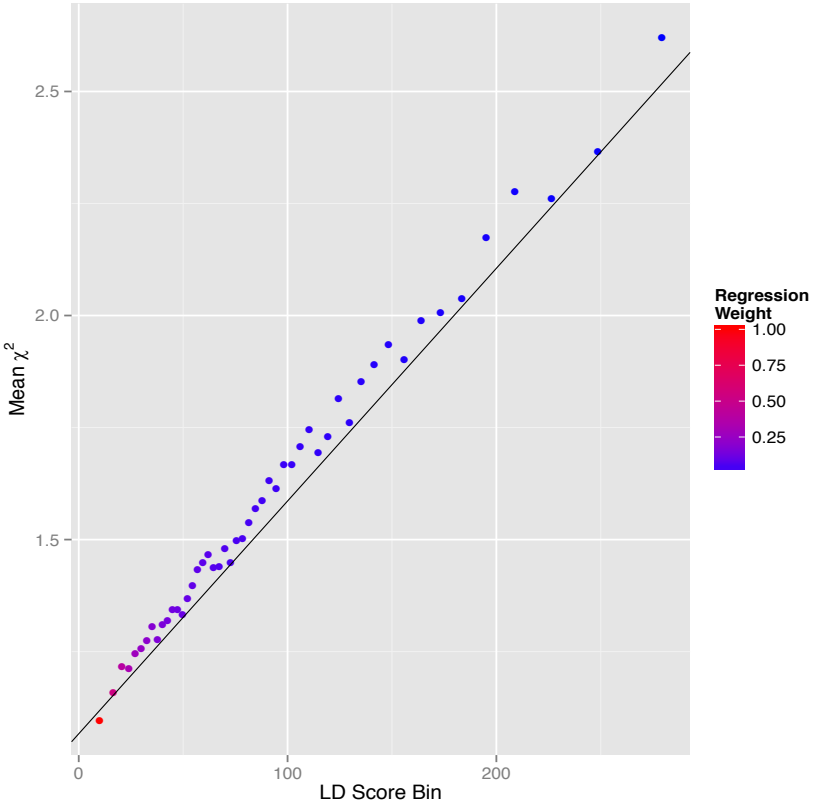
PGC Schizophrenia

Lambda = 1.48

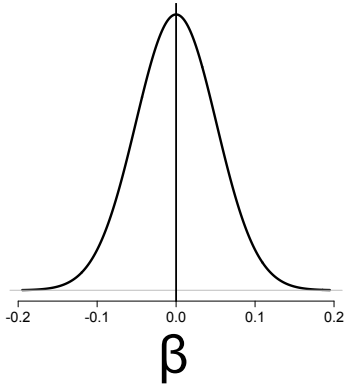
Intercept = 1.06

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*?

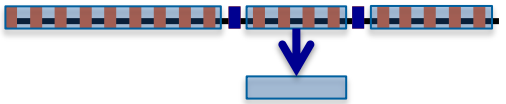
$$E[\chi_j^2] = 1 + Na + \frac{h_g^2 N}{M} l_j$$

New estimator of heritability

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

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

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

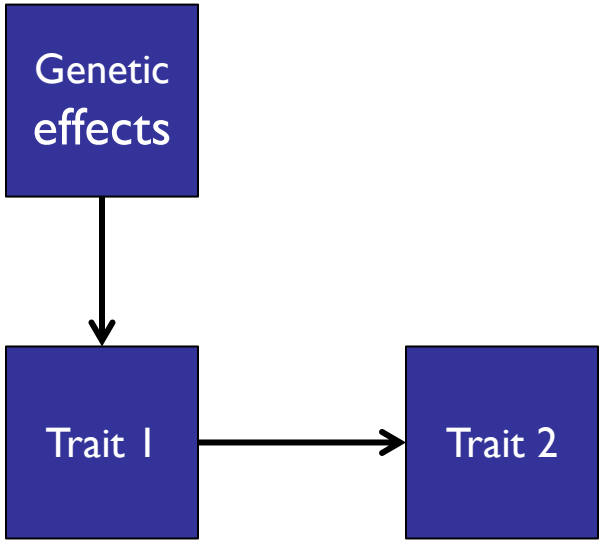


Genetic Correlation Method in:

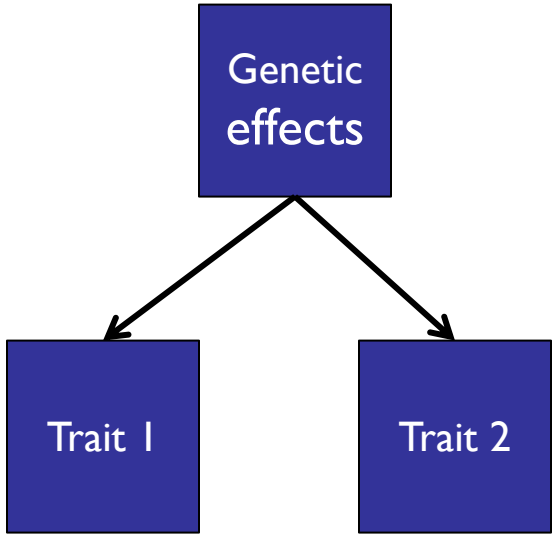
nature genetics

An atlas of genetic correlations across human diseases and traits

Potential sources of genetic correlation



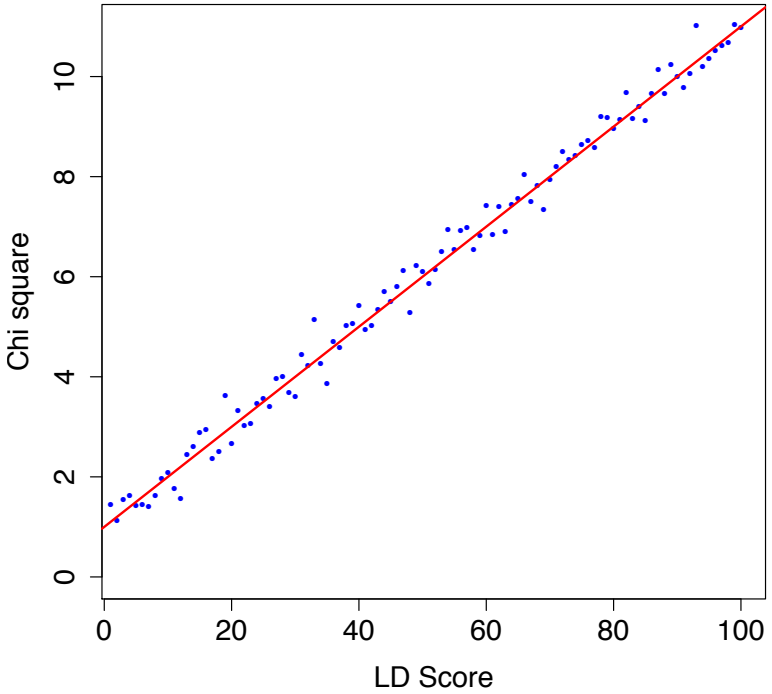
Trait 1 exerts causal effect on Trait 2



Genetic effects influence Trait 1 and Trait 2

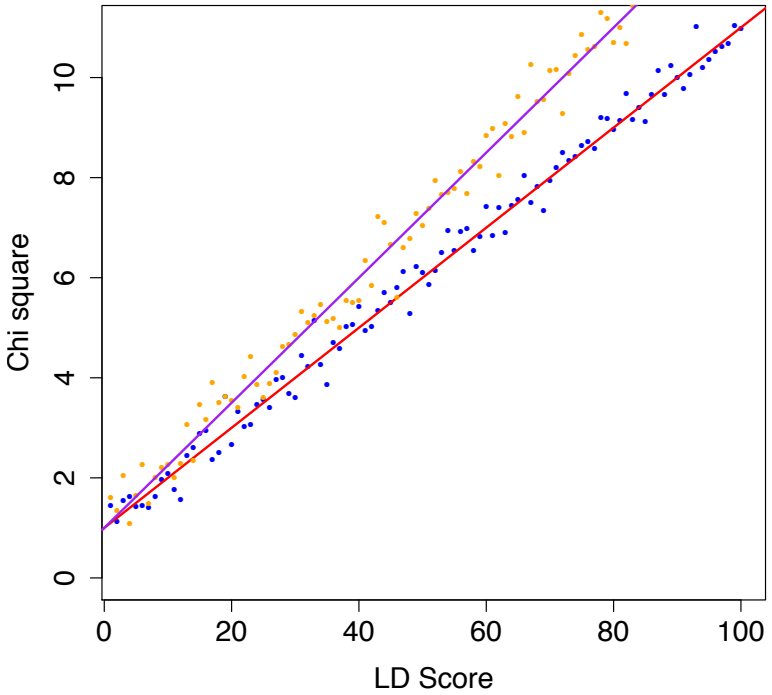
LD Score regression
Genetic correlation

Trait 1



Slope estimates heritability

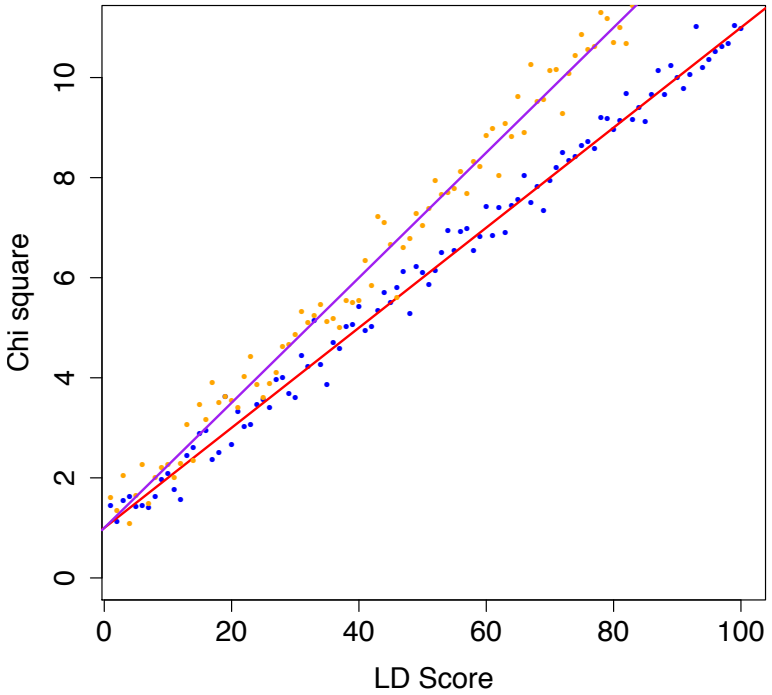
LD Score regression
Genetic correlation



Trait 1
Trait 2

We can a second trait and obtain two heritability estimates

LD Score regression
Genetic correlation

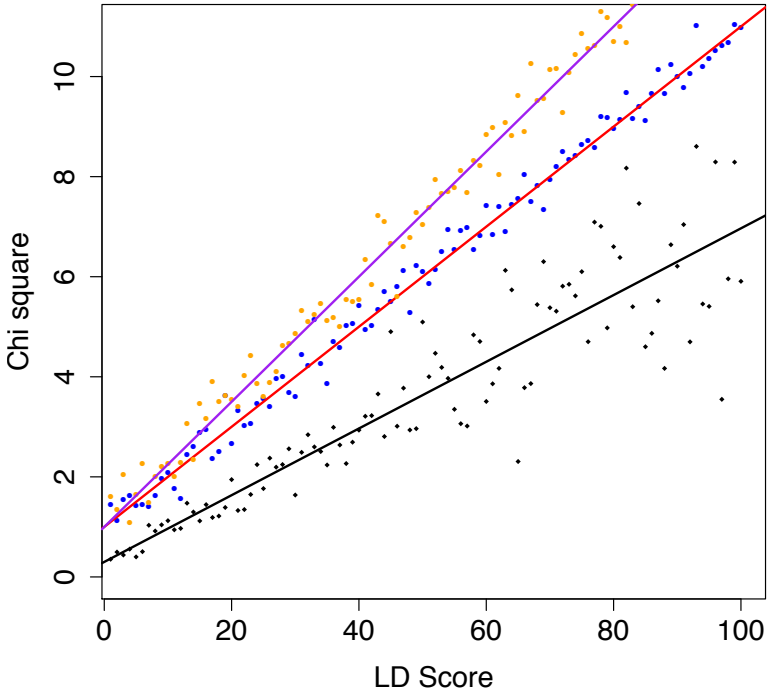


Trait 1
Trait 2

$Z^*Z = \chi^2$

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

LD Score regression
Genetic correlation



Trait 1

Trait 2

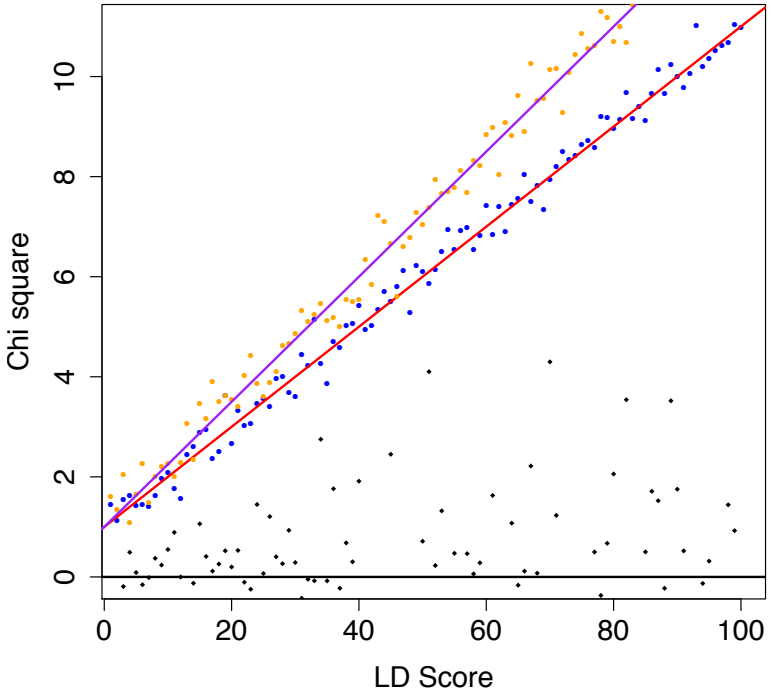
R_G

$Z^*Z = \chi^2$

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

You can do it yourself
ldsc.broadinstitute.org

LD Hub Home About Update log Software

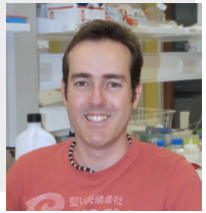
BROAD INSTITUTE · University of BRISTOL · MRC Integrative Epidemiology Unit

LD Hub is a centralised database of summary-level GWAS results and a web interface for LD score regression.

Get Started with LD Hub

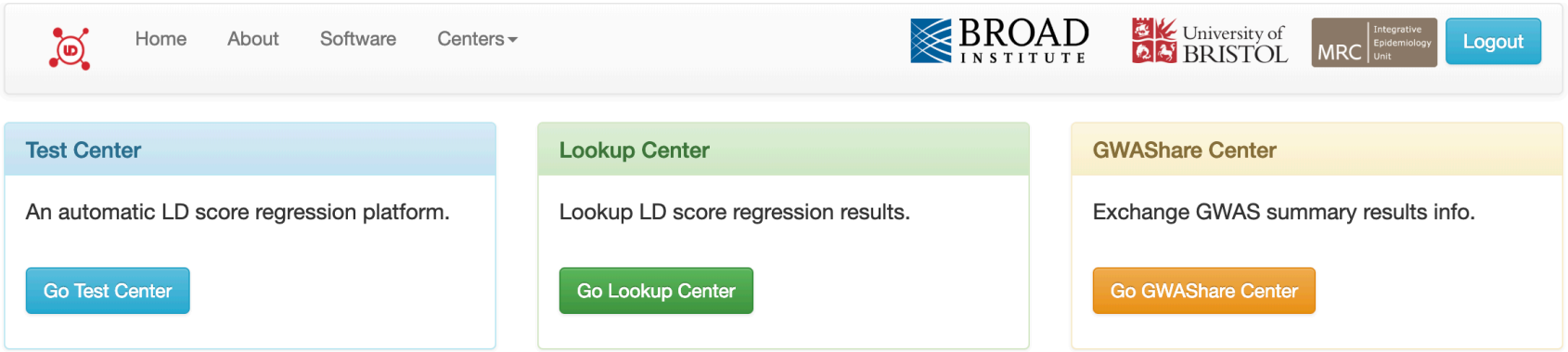
Currently v1.0.1

Jie Zheng

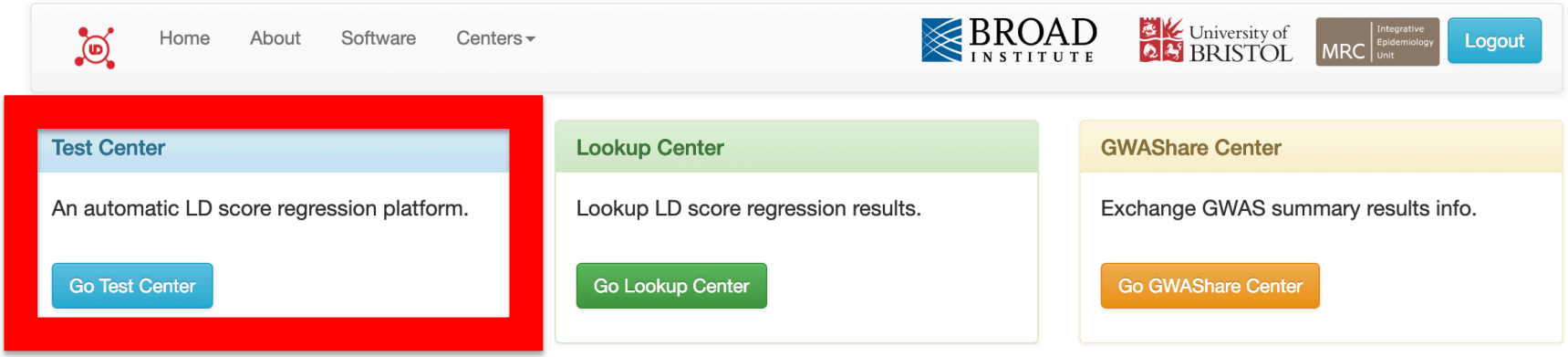


David Evans

LD Hub practical



Test center



Running your results through LD-score genetic correlation

Test center

Home About Software Centers BROAD INSTITUTE University of BRISTOL MRC Integrative Epidemiology Unit Logout

Test Center

Please follow the steps to Upload file and Select data.

Step 0: Existing results | Step 1: File upload | Step 2: Data selection

- We selected traits for inclusion via the following procedure:
    1. Begin with all publicly available **non-sex-stratified** and **predominantly European** summary statistics.
    2. Remove studies that do not provide signed summary statistics.
    3. Remove studies not imputed to at least **HapMap 2**.
    4. Remove studies that adjust for heritable covariates
    5. Remove studies that with **number of SNPs smaller than 450,000**
    6. Remove studies that with **number of individuals smaller than 5,000**
    7. Remove all traits with **heritability z-score below 2**. (Genetic correlation estimates for traits with heritability z-score below 2 are generally too noisy to report.) We recommand traits with heritability z-score larger than 4.
    8. Remove SNPs with **extermely large effect sizes (X^2 > 80)**, because outliers can unduly influence the regression.
    9. Remove all variants on **chromosome 6 in the region 26MB to 34MB** (the MHC region).
- Precalculated LD score regression SNP heritability and genetic correlation analysis results can be found here.
- Information of the GWA studies included in LD Hub can be found here.

Uploading your own results

Step 0: Existing results | Step 1: File upload | Step 2: Data selection

Input format

The input format is: Show/Hide. Headers are needed for the input file. More details are explained here.

LD Hub can handle both space and tab delimited files. By default, please prepare your file using tab as delimiter.

LD Hub can handle but Z scores and betas. By default, please use Z scores in your file.

Important notes for your uploaded file:

1. To save the uploading time, LD Hub only accepts **zipped** files as input (e.g. mydata.zip).

2. Please check that there is **ONLY ONE** plain **TXT** file (e.g. mydata.txt) in your zipped file.

3. Please make sure you do **NOT** zip any folder together with the plain txt file (e.g. /myfolder/mydata.txt), otherwise you will get an error: [Errno 2] No such file or directory

4. Please do **NOT** zip multiple files (e.g. zip mydata.zip file1.txt file2.txt ..) or zip a file with in a folder (e.g. zip mydata.zip /path/to/my/file/mydata.txt).

5. Please keep the file name of your plain txt file **short (less than 50 characters)**, otherwise you may get an error: [Errno 2] No such file or directory

6. Please zip your plain txt file using following command (ONE file at a time):

For Windows system: 1) Locate the file that you want to compress. 2) Right-click the file, point to Send to, and then click Compressed (zipped) folder.

For Linux and Mac OS system: zip mydata.zip mydata.txt

Reminder: for Mac OS system, please do **NOT** zip you file by right click mouse and click "Compress" to zip your file, this will automatically create a folder called "__MACOS". You will get an error: [Errno 2] No such file or directory.

Pick your traits to compare

Data selection

Please select the traits you are interested in from our database (click trait name to show / hide sub catalog for each catalog). More details of the traits can be found here.

We have removed variants in MHC region (chromosome 6 in the region 26MB to 34MB) for all traits in LD Hub. For the Eczema GWAS, we further removed all variants +/-500KB from the top variant (rs61813875) in the filaggrin region.

- Select All / Unselect All
- Autoimmune diseases (new)
- Smoking behaviour
- Neurological diseases
- Personality traits
- Reproductive traits
- Haemotological traits
- Sleeping
- Cognitive
- [NEW] 597 UK Biobank traits (from Ben Neale's group)
- Anthropometric traits
- Blood lipids
- Education
- Uric acid
- Brain Volume (ENIGMA)
- Cancer
- Metal
- Other
- Metabolites (Kettunen et al)
- Glycemic traits
- Bone mineral density
- Psychiatric diseases
- Kidney diseases / traits
- Cardiometabolic traits (new)
- Hormone
- Aging
- Lung function (new)

Reminder:

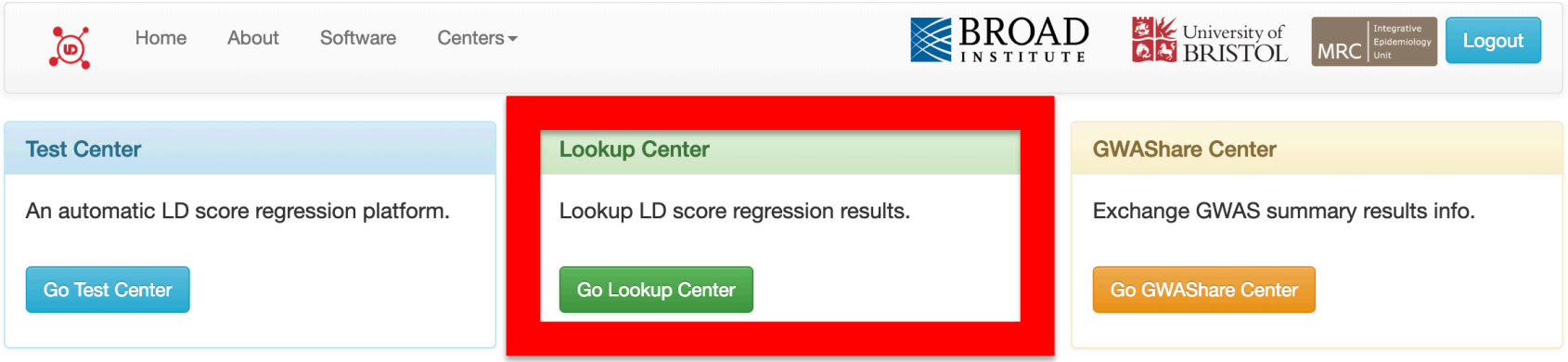
1) Please make sure you select at least one of the above traits, otherwise an error page will appear.

2) Each test may take about 20 seconds. An analysis of all traits may take up to five hours.

3) Your uploaded file will be removed directly from the server after the analysis. If you are willing to share your GWAS results with us. Please visit GWAShare center

Submit your request | Reset

Lookup center



Browse previously generated results

Heritability

Lookup Center

Lookup existing LD score regression analysis results

SNP Heritability results | Genetic correlation results

To download the existing SNP heritability results of 219 traits, please click here

The existing SNP heritability for **229** traits can be found here (the SNP heritability results are on the observed scale):

Trait name	H2	SE_H2	Z_H2	Lambda GC	Chi2	Intercept
Adiponectin	0.1369	0.0242	5.65702	1.068	1.09	1.0133
Age of smoking initiation	0.0665	0.0185	3.59459	1.0345	1.0295	0.9981
Child birth length	0.1697	0.0229	7.41048	1.0588	1.0672	0.9926
Child birth weight	0.1124	0.0179	6.27933	1.0466	1.0618	1.0043
Body mass index	0.1855	0.0089	20.8427	1.3675	1.4681	1.0188
Body fat	0.104	0.0076	13.6842	1.0315	1.0578	0.9083
Coronary artery disease	0.0728	0.0054	14.463	1.2386	1.3288	1.0475

Genetic correlation

Lookup Center

Lookup existing LD score regression analysis results

SNP Heritability results | Genetic correlation results

1. To download the existing genetic correlation results for 49 traits from Bulik Sullivan et al. (2015), please click here

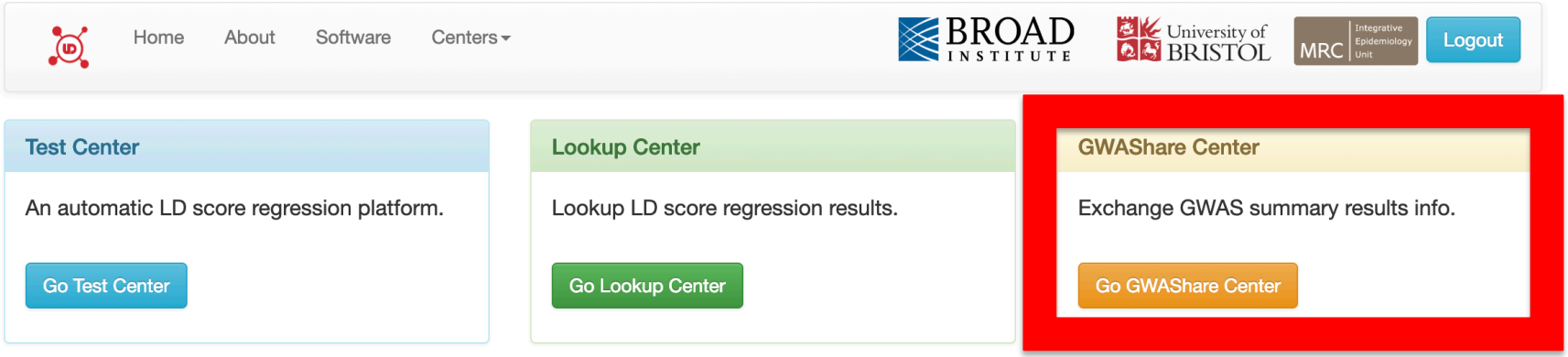
2. To download the existing genetic correlation results for 221 traits (without 7 traits from ENIGMA) using data from LD Hub, please click here

Note: in the above genetic correlation results file, there are two sheets: 1) the 'rg' sheet contains the genetic correlation matrix of 196x196 traits. 2) The 'all-info' sheet contains all bivariate LD score regression results of 196x196 traits; each cell contains 8 values for a certain pair-wise correlation, the 8 values refer to 'rg se z p h2_obs h2_obs_se h2_int h2_int_se gcov_int gcov_int_se' respectively. For a certain cell, the 7th value 'gcov_int' is the phenotypic correlation between two traits, which take into account the influence of sample overlap between two GWA studies (e.g. if there is no sample overlap, the gcov_int will near zero; if two traits are measured in the same samples, gcov_int will be the phenotypic correlation between these two traits).

3. The existing genetic correlation for 49 traits from Bulik Sullivan et al. (2015) can be found here:

Trait1	Trait2	rg	se	z	p
ADHD	Age at Menarche	-0.153	0.08218	-1.858	0.063
ADHD	Age at Smoking	-0.036	0.2427	-0.147	0.883
ADHD	Alzheimer's	-0.055	0.2191	-0.249	0.803
ADHD	Anorexia	0.192	0.1162	1.649	0.099
ADHD	Autism Spectrum	-0.164	0.1438	-1.144	0.253
ADHD	BMI	0.287	0.08913	3.222	0.001

LD Hub practical



Sharing and exchanging GWAS results

Download results or share your own!

Browse existing GWAS resources | Share your GWAS data

We provided a list of exsiting GWAS resoruces here: (cloumns are filename, trait name, consortium/database, sample size, PMID, publish year and ethnicity)

To download the study information of the existing traits, please click here

File name	Trait name	Consortium/ first_author/ database	Sample size	PMID	Publish year	Ethnicity
adipogen.discovery.eur_.meta_.public.release.txt.noMHC.sumstats_deGC.gz	Adiponectin	ADIPOGen	39883	22479202	2012	Mixed
Age_of_smoking.sumstats.gz	Age of smoking initiation	TAG	47961	20418890	2010	European
Birthlength.sumstats.gz	Child birth length	EGG	28459	25281659	2015	European
Birthweight.sumstats.gz	Child birth weight	EGG	26836	23202124	2013	European
BMI_2010.sumstats_deGC.gz	Body mass index	GIANT	123912	20935630	2010	European
body_fat_percentage_GWAS_PLUS_MC_ALL_ancestry_se_Sex_combined_for_locus_zoom_plot.TBL.txt.tab.sumstats.gz	Body fat	Lu	100716	26833246	2016	Mixed

Analysis of UK Biobank

GWAS of UK Biobank

Download and decryption → Software development → Phenotype wrangling → QC and GWAS → Heritability analysis

Sam Bryant

Cotton Seed

hail

Andrea Ganna, Duncan Palmer, Caitlin Carey

Liam Abbott
Dan Howrigan

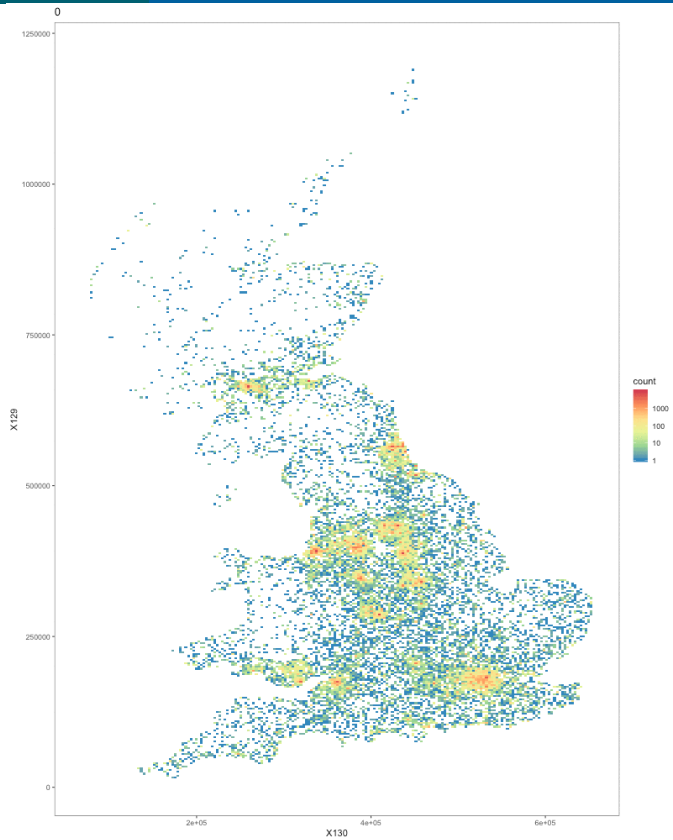
Raymond Walters

Also thanks to: Verneri Anttila, Krishna Aragam, Alex Baumann, Jon Bloom, Joanne Cole, Mark J. Daly, Mark J. Daly, Rob Damien, Steven Gazal, Jackie Goldstein, Mary Haas, Joel Hirschhorn, Eric Jones, Sekar Kathiresan, Dan King, Ruchi Munshi, Tim Poterba, Manuel Rivas, Sailaja Vedantam

biobank uk

Improving the health of future generations

- Follows health and well-being of 500,000 participants
- Genotyped using the Affymetrix Biobank Array
- Lots of phenotypes collected [needs harmonization]
- Lots of opportunity!



Example self-report

Data-Field 1080

Description: Time spent using computer

Category: Physical activity - Lifestyle and environment - Touchscreen - UK Biobank Assessment Centre

Participants	498,619
Item count	535,025
Stability	Complete

Value Type	Integer, hours/day
Item Type	Data
Strata	Primary

Sexed	Both sexes
Instances	Defined (3)
Array	No

Data | **3 Instances** | **Notes** | **4 Categories** | **0 Related Data-Fields** | **0 Tabulations** | **2 Resources**

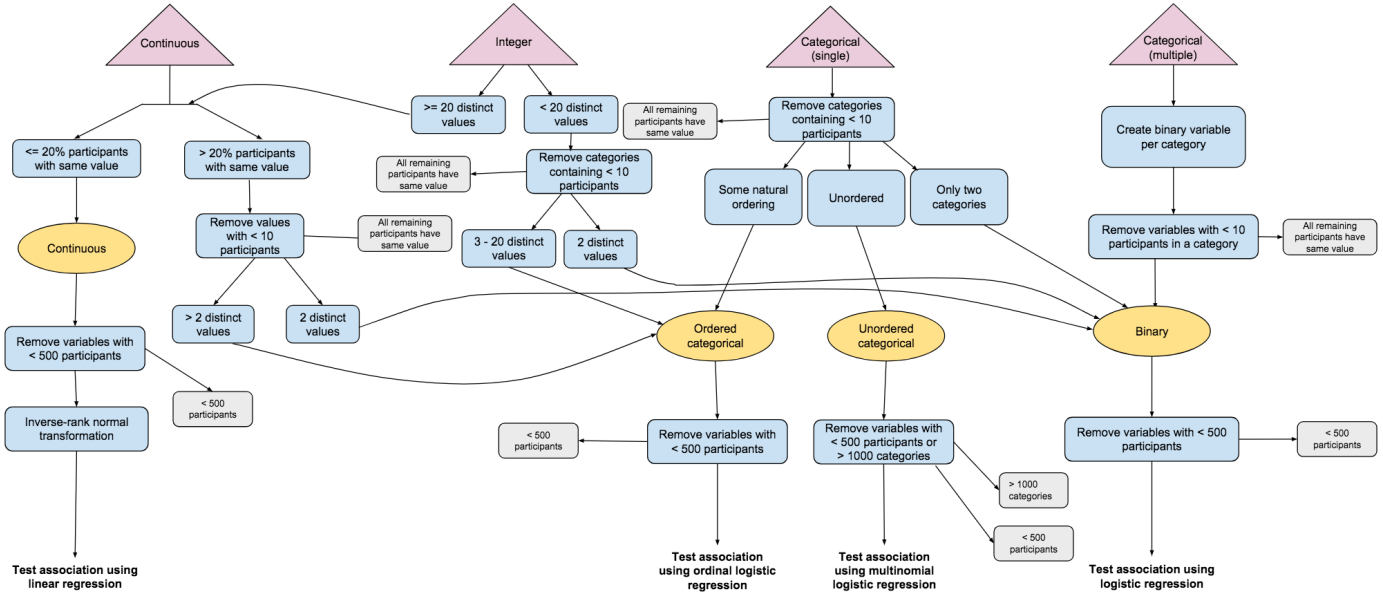
535,025 items of data are available, covering 498,619 participants.
Some values have special meanings defined by Data-Coding 100329.
Defined-instances run from 0 to 2, labelled using Instancing 2.
Units of measurement are hours/day.

Maximum	24
Decile 9	3
Decile 8	2
Decile 7	1
Decile 6	1
Median	1
Decile 4	1
Decile 3	0
Decile 2	0
Decile 1	0
Minimum	0

- There are 23 distinct values.
- Mean = 1.27211
- Std.dev = 1.52124
- 5230 items above graph maximum of 6
- 109750 items have value -10 (Less than an hour a day)
- 1598 items have value -3 (Prefer not to answer)
- 3240 items have value -1 (Do not know)

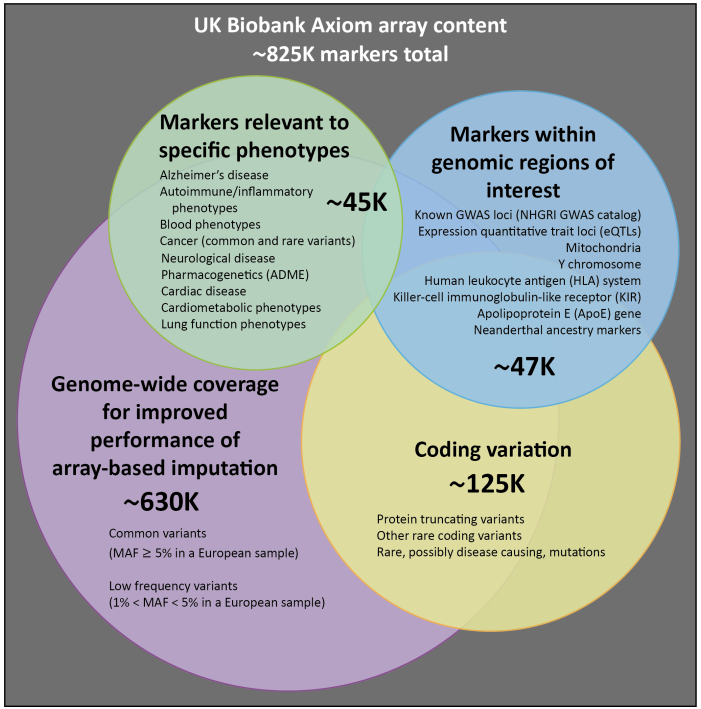
Counts of participants/items last updated 04 Feb 2017.

PHESANT!



Copious thanks to Millard LAC, Davies NM, Gaunt TR, Davey Smith G, Tilling K. PHESANT: a tool for performing automated phenome scans in UK Biobank. bioRxiv (2017)

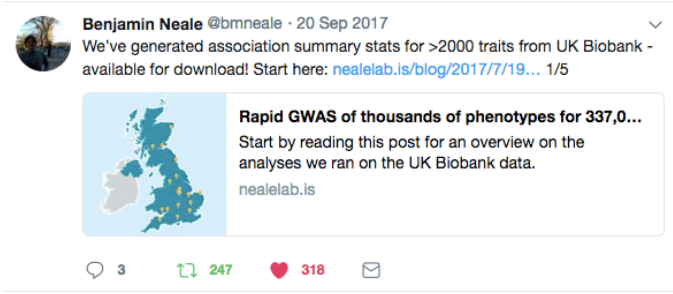
What's on the array?



Imputed to HRC + 1KG

Round 1 GWAS

- Fall 2017, the Neale lab…
  - GWASed 2,419 phenotypes
    - Blogged about it
    - Put them on dropbox
      - And people made browsers
  - Estimated h^2 for all of them
  - Made an h^2 browser
    - Blogged about that too



Show 10 entries Search: home area

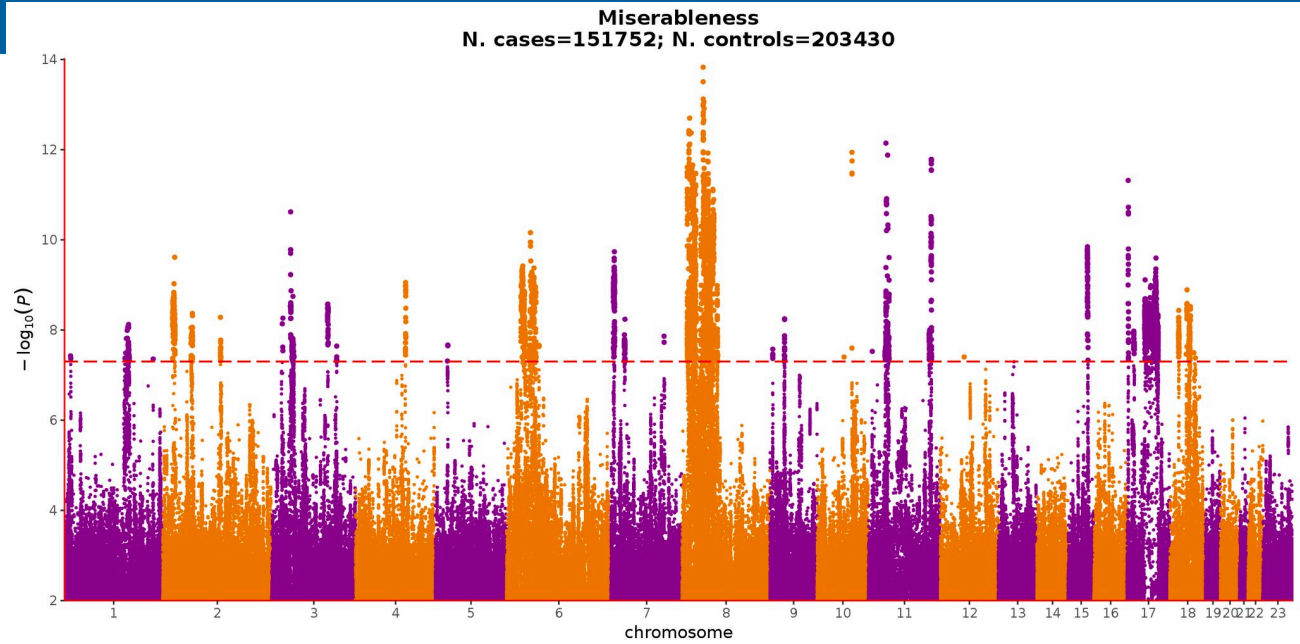
ID	Phenotype	N	Prev.	Int.	Int. p	h2	h2 p
20118_11	Home area population density - urban or rural: Scotland - Large Urban Area	333,997	0.056	2.103	0.00	0.0885	0.0000535
20118_12	Home area population density - urban or rural: Scotland - Other Urban Area	333,997	0.011	1.195	2.98e-59	0.0565	0.0994
20118_13	Home area population density - urban or rural: Scotland - Accessible Small Town	333,997	0.0031	1.079	1.88e-20	-0.117	0.932
20118_16	Home area population density - urban or rural: Scotland - Accessible Rural	333,997	0.0034	1.077	2.41e-18	-0.0363	0.686
20118_6	Home area population density - urban or rural: England/Wales - Town and Fringe - less sparse	333,997	0.073	1.031	0.0000822	0.00155	0.416
20118_7	Home area population density - urban or rural: England/Wales - Village - less sparse	333,997	0.052	1.013	0.0643	0.0219	0.0202
20118_8	Home area population density - urban or rural: England/Wales - Hamlet and Isolated Dwelling - less sparse	333,997	0.023	1.003	0.346	0.0139	0.199

Showing 1 to 7 of 7 entries (filtered from 2,304 total entries) PREVIOUS 1 NEXT

Nealelab.is/blog

GWASbot!

@SbotGWA



Trait info: http://www.ukbiobank.ac.uk/data-showcase/

All things UK Biobank GWAS: http://www.nealelab.is/uk-biobank/

Andrea Ganna

Heritability at scale!

- Description: http://www.nealelab.is/blog/2017/9/15/heritability-of-2000-traits-and-disorders-in-the-uk-biobank
- Browser: https://nealelab.github.io/UKBB_ldsc/

9,928 GWAS later… let's talk h^2 using LD score regression

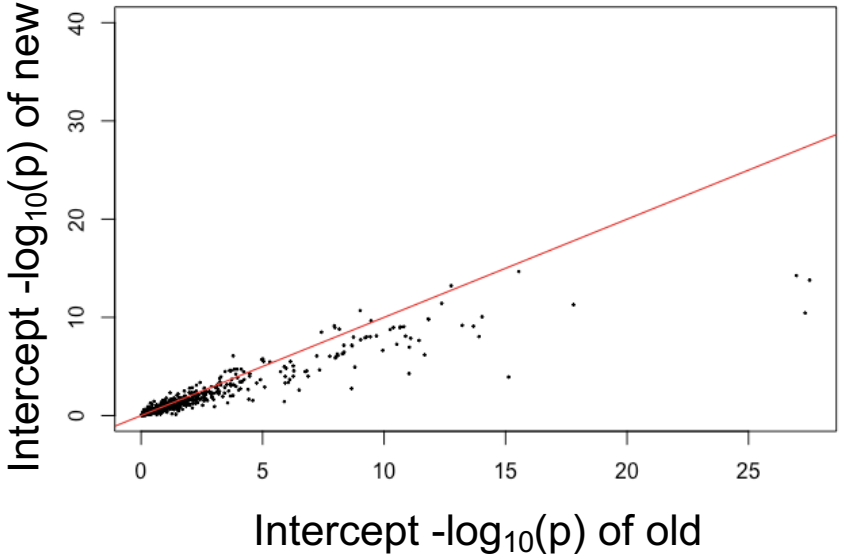
$$E[\chi_j^2] = 1 + Na + \frac{h_g^2 N}{M} l_j$$

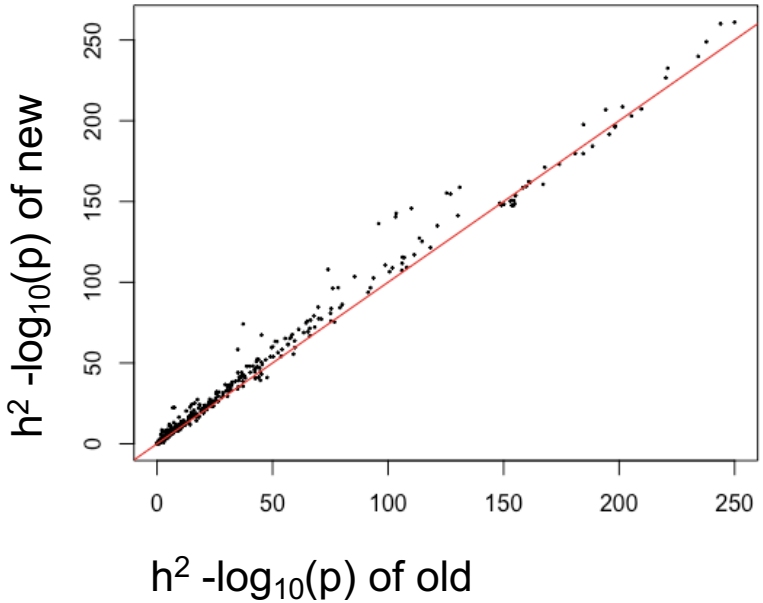
Estimating heritability from GWAS summary statistics

How do round 2 ldsc results compare?

Raymond Walters

- Intercept less significant
- h2 more significant with stable estimates

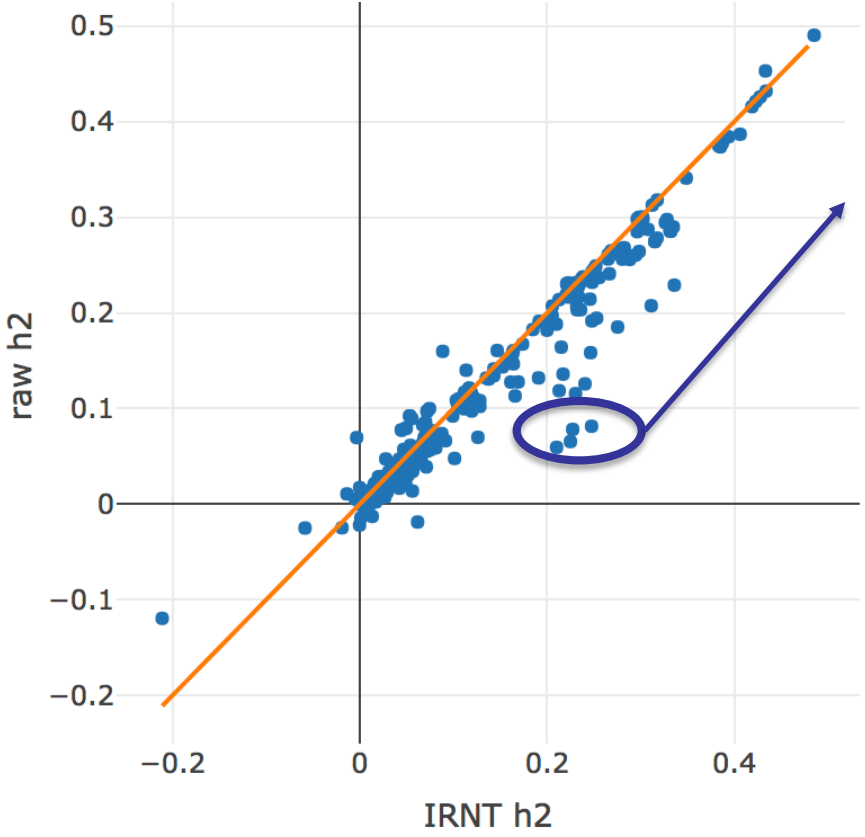




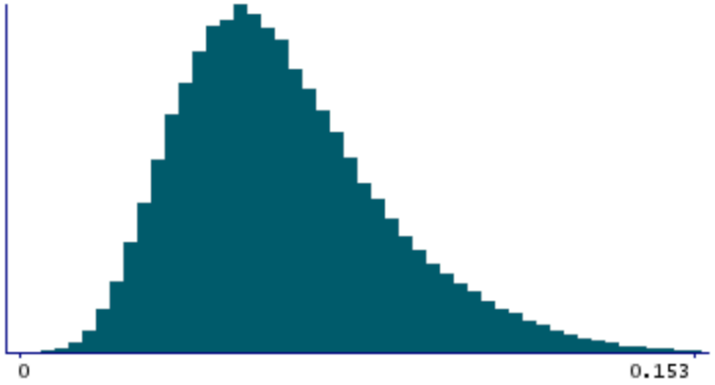
Contrasting raw phenotypes to inverse rank normalize transformed

Let's look at heritability

Raymond Walters

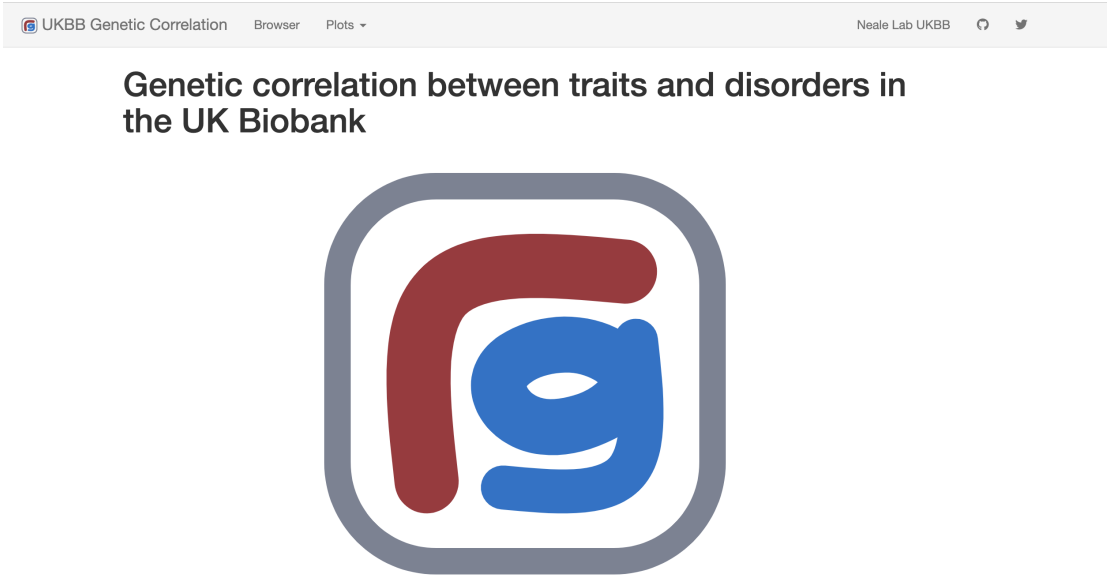


Lymphocyte count
Reticulocyte count
Reticulocyte %
High light scatter reticulocyte %



Reticulocyte count

Browser for UKBB genetic correlation



https://ukbb-rg.hail.is/