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

Tinca J C Polderman¹,¹⁰, Beben Benyamin²,¹⁰, Christiaan A de Leeuw¹,³, Patrick F Sullivan⁴–⁶, Arjen van Bochoven⁷, Peter M Visscher²,⁸,¹¹ & Danielle Posthuma¹,⁹,¹¹

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, S. Hong Lee, Michael E. Goddard, and Peter M. Visscher

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

Estimating Missing Heritability for Disease from Genome-wide Association Studies
Sang Hong Lee, Naomi R. Wray, Michael E. Goddard, and Peter M. Visscher
LD Score regression

With thanks

Brendan Bulik-Sullivan  Hilary Finucane  Po-Ru Loh  Mark Daly  Alkes Price
How does LD shape association?
How does LD shape association?

Lonely SNPs [no LD]

LD blocks
How does LD shape association?

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

Association

All markers correlated with a causal variant show association
How does LD shape association?

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

Lonely SNPs only show association if they are causal
What happens under polygenicity?

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

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

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
Lambda = 1.48
Intercept = 1.06
Slope $p$-value < $10^{-300}$

Overwhelming majority of inflation is consistent with polygenic architecture
Draw polygenic effects from $N(0, n/m^2)$, var =

What is the $E[\chi^2]$ for variant $j$?

$$E[\chi^2_j] = 1 + Na + \frac{h^2_g N}{M} l_j$$

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
Questions for the audience

- What are the model assumptions?
- What are ways we can relax some of those assumptions?
Analysis of UK Biobank
GWAS of UK Biobank

Download and decryption

Software development

Phenotype wrangling

QC and GWAS

Heritability analysis

Sam Bryant

Cotton Seed

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

Rob Damien

Steven Gazal

Jackie Goldstein

Mary Haas

Joel Hirschhorn

Eric Jones

Sekar Kathiresan

Dan King

Ruchi Munshi

Tim Poterba

Manuel Rivas

Sailaja Vedantam
• 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

<table>
<thead>
<tr>
<th>Participants</th>
<th>498,619</th>
</tr>
</thead>
<tbody>
<tr>
<td>Item count</td>
<td>535,025</td>
</tr>
<tr>
<td>Stability</td>
<td>Complete</td>
</tr>
<tr>
<td>Value Type</td>
<td>Integer, hours/day</td>
</tr>
<tr>
<td>Item Type</td>
<td>Data</td>
</tr>
<tr>
<td>Strata</td>
<td>Primary</td>
</tr>
<tr>
<td>Sexed</td>
<td>Both sexes</td>
</tr>
<tr>
<td>Instances</td>
<td>Defined (3)</td>
</tr>
<tr>
<td>Array</td>
<td>No</td>
</tr>
</tbody>
</table>

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.

- 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.
What’s on the array?

UK Biobank Axiom array content
~825K markers total

- Markers relevant to specific phenotypes
  - Alzheimer’s disease
  - Autoimmune/inflammatory phenotypes
  - Blood phenotypes
  - Cancer (common and rare variants)
  - Neurological disease
  - Pharmacogenetics (ADME)
  - Cardiac disease
  - Cardiometabolic phenotypes
  - Lung function phenotypes
  ~45K

- Markers within genomic regions of interest
  - Known GWAS loci (NHGRI GWAS catalog)
  - Expression quantitative trait loci (eQTLs)
  - Mitochondria
  - Y chromosome
  - Human leukocyte antigen (HLA) system
  - Killer-cell immunoglobulin-like receptor (KIR)
  - Apolipoprotein E (APOE) gene
  - Neanderthal ancestry markers
  ~47K

- Genome-wide coverage for improved performance of array-based imputation
  - Common variants
    (MAF ≥ 5% in a European sample)
  - Low frequency variants
    (1% < MAF < 5% in a European sample)
  ~630K

- Coding variation
  - Protein truncating variants
  - Other rare coding variants
  - Rare, possibly disease causing, mutations
  ~125K

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

Nealelab.is/blog
GWASbot!

Miserableness
N. cases=151752; N. controls=203430

Trait info: http://www.ukbiobank.ac.uk/data-showcase/
All things UK Biobank GWAS: http://www.nealelab.is/uk-biobank/
Heritability at scale!

• Description: http://www.nealelab.is/blog/2017/9/15/heritability-of-2000-trait-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

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

Estimating heritability from GWAS summary statistics
How do round 2 ldsc results compare?

- Intercept less significant
- $h^2$ more significant with stable estimates

![Graphs showing comparisons](image_url)
Contrasting raw phenotypes to inverse rank normalize transformed
Let’s look at heritability

Lymphocyte count
Reticulocyte count
Reticulocyte %
High light scatter reticulocyte %
What about sex-specific effects?

- Sex-specific GWAS allow us to scan for:
  - Differences in female vs. male $h^2$
    - E.g. could indicate differences in variance of environmental effects, measurement differences
  - Female vs. male $r_g < 1$
    - E.g. relative effects of different SNPs differ by sex

- Can also test for SNP-level differences
  - Slower and labor intensive, so $h^2$, $r_g$ can help prioritize

- To start: look at 448 phenotypes with Neff > 10000 in both sexes and z-score of $h^2 > 4$ is at least 1 sex
Strong $h^2$ observed in both sexes

- >70% of traits at least nominally heritable in each sex
  - $P < .05$

- Mean $h^2 \sim .09$

- Consistent with joint analysis of both sexes
Is $h^2$ equal across sexes?

$h^2$ strongly correlated across sex

~10% of traits have nominally different $h^2$ between sexes

<table>
<thead>
<tr>
<th>Description</th>
<th>Fem. $h^2$</th>
<th>Male $h^2$</th>
<th>P diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average weekly beer plus cider intake</td>
<td>0.0416</td>
<td>0.1152</td>
<td>3.11E-10</td>
</tr>
<tr>
<td>Diastolic blood pressure, automated</td>
<td>0.1799</td>
<td>0.1160</td>
<td>1.13E-06</td>
</tr>
<tr>
<td>Systolic blood pressure, automated</td>
<td>0.1768</td>
<td>0.1208</td>
<td>1.03E-05</td>
</tr>
<tr>
<td>Number of operations, self-reported</td>
<td>0.0845</td>
<td>0.0491</td>
<td>2.53E-05</td>
</tr>
<tr>
<td>Duration of vigorous activity</td>
<td>0.0037</td>
<td>0.0555</td>
<td>3.91E-05</td>
</tr>
</tbody>
</table>
Functional partitioning

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

DHS
Coding

LD Score:
9  1  4  1  5

Finucane et al. 2015 Nat Gen

\[ l_j = \sum_{k \leq c} r_{jk}^2 \]
Functional partitioning

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

DHS
Coding

LD Score  9  1  4  1  5
DHS Score  5  0  0  0  0
Coding Score  0  0  1  1  3

Finucane et al. 2015 Nat Gen

\[ l_j = \sum_{k \in c} r_{jk}^2 \]
## Annotations

<table>
<thead>
<tr>
<th>Mark</th>
<th>Source/reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coding, 3’ UTR, 5’ UTR, Promoter, Intron</td>
<td>UCSC; Gusev et al., in press AJHG</td>
</tr>
<tr>
<td>Digital Genomic Footprint, TFBS</td>
<td>ENCODE; Gusev et al., in press AJHG</td>
</tr>
<tr>
<td>CTCF binding site, Promoter Flanking, Repressed, Transcribed, TSS, Enhancer, Weak Enhancer</td>
<td>ENCODE; Hoffman et al., 2012 Nucleic Acids Research</td>
</tr>
<tr>
<td>DHS, fetal DHS, H3K4me1, H3K4me3, H3K9ac</td>
<td>Trynka et al., 2013 Nature Genetics.*</td>
</tr>
<tr>
<td>Conserved</td>
<td>Lindblad-Toh et al., 2011 Nature</td>
</tr>
<tr>
<td>FANTOM5 Enhancer</td>
<td>Andersson et al., 2014 Nature</td>
</tr>
<tr>
<td>lincRNAs</td>
<td>Cabili et al., 2011 Genes Dev</td>
</tr>
<tr>
<td>DHS and DHS promoter</td>
<td>Maurano et al., 2012 Science</td>
</tr>
<tr>
<td>H3K27ac</td>
<td>Roadmap; PGC2 2014 Nature</td>
</tr>
</tbody>
</table>

*Post-processed from ENCODE and Roadmap data by S. Raychaudhuri and X. Liu labs*
### Datasets for GWAS

**Selected for a Z>7 for \( h^2 \)**

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Reference</th>
<th>Phenotype</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Height</td>
<td>Lango Allen, 2010</td>
<td>Schizophrenia</td>
<td>PGC, 2014</td>
</tr>
<tr>
<td>BMI</td>
<td>Speliotes, 2010</td>
<td>Bipolar</td>
<td>Sklar, 2011</td>
</tr>
<tr>
<td>Age of menarche</td>
<td>Perry, 2014</td>
<td>Anorexia</td>
<td>Boraska, 2014</td>
</tr>
<tr>
<td>LDL</td>
<td>Teslovich, 2010</td>
<td>Education years</td>
<td>Rietveld, 2013</td>
</tr>
<tr>
<td>HDL</td>
<td>Teslovich, 2010</td>
<td>Ever smoked</td>
<td>TAG, 2010</td>
</tr>
<tr>
<td>Triglycerides</td>
<td>Teslovich, 2010</td>
<td>Rheumatoid Arth</td>
<td>Okada, 2014</td>
</tr>
<tr>
<td>CAD</td>
<td>Schunkert, 2011</td>
<td>Crohn’s Disease</td>
<td>Jostins, 2012</td>
</tr>
<tr>
<td>T2D</td>
<td>Morris, 2012</td>
<td>Ulcerative Colitis</td>
<td>Jostins, 2012</td>
</tr>
<tr>
<td>Fasting Glucose</td>
<td>Manning, 2012</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Average enrichments per class
Collapsed results across 17 traits

<table>
<thead>
<tr>
<th>Category</th>
<th>% SNPs</th>
<th>% $h^2$</th>
<th>Enrichment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Conserved</td>
<td>2.6</td>
<td>34.7</td>
<td>13.4x</td>
</tr>
<tr>
<td>Coding</td>
<td>1.5</td>
<td>10.4</td>
<td>7.0x</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>13.3</td>
<td>34.4</td>
<td>2.6x</td>
</tr>
<tr>
<td>H3K4me3 (peaks)</td>
<td>4.2</td>
<td>15.8</td>
<td>3.8x</td>
</tr>
</tbody>
</table>
Specific trait enrichments

- Fantom5 Enhancers massively enriched for Immune traits
  - both significantly enriched
Cell type enrichments

Warning
P-value scale changes
Use the lines as guides

77 from H3K4me1
81 from H3K4me3
27 from H3K9ac
35 from H3K27ac

Hierarchical clustering into sets
Genetic Correlation Method in:

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

Slope estimates heritability
LD Score regression
Genetic correlation

We can a second trait and obtain two heritability estimates.
Z*Z = \chi^2

So we can estimate genetic covariance from the product of the Z-scores
$Z^*Z = \chi^2$

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

$R_G = 0.5$
Here $R_G = 0$

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

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

It’s all heritable!

- 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
### Brainstorm – within psychiatry

<table>
<thead>
<tr>
<th>Condition</th>
<th>ADHD</th>
<th>Anorexia Nervosa</th>
<th>Anxiety Disorder</th>
<th>ASD</th>
<th>Bipolar Disorder</th>
<th>MDD</th>
<th>OCD</th>
<th>Schizophrenia</th>
<th>Tourette Syndrome</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anorexia Nervosa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Anxiety Disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ASD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bipolar Disorder</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>OCD</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Schizophrenia</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tourette Syndrome</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Correlation and P-value**

- Correlation values range from -1 to 1, indicating the strength and direction of the relationship.
- P-values indicate statistical significance:
  - * indicates p < 0.05
  - ** indicates p < 0.001
Brainstorm within neurology
Brainstorm – across neurology and psychiatry
# Brainstorm – take it further?

<table>
<thead>
<tr>
<th></th>
<th>ADHD</th>
<th>Anorexia nervosa</th>
<th>Anxiety disorders</th>
<th>ASD</th>
<th>Bipolar disorder</th>
<th>MDD</th>
<th>OCD</th>
<th>Schizophrenia</th>
<th>Tourette Syndrome</th>
<th>Alzheimer's disease</th>
<th>Epilepsy</th>
<th>Focal epilepsy</th>
<th>Generalized epilepsy</th>
<th>ICH</th>
<th>Ischemic stroke</th>
<th>Early-onset stroke</th>
<th>Migraine with aura</th>
<th>Migraine without aura</th>
<th>Multiple sclerosis</th>
<th>Parkinson's disease</th>
</tr>
</thead>
<tbody>
<tr>
<td>College attainment</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Years of education</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Cognitive performance</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Extraversion</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Depressive symptoms</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Subjective well-being</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Never/ever smoked</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
<tr>
<td>Cigarettes per day</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>
Generalizations of genetic correlation

Genetic sharing across men and women
Female (1) vs male (0) GWAS

\[ h^2 (ldsc) = 0.012 (0.002) \]
Differential ascertainment bias
Male/Female genetic correlation

- Next step is to look at genetic correlation between female and male results for each trait
  - Again using LD score regression

- Focus on 448 traits with significant $h^2$ in at least one sex
  - After Bonferroni correction for 865 traits
Genetic correlation estimate between females and males

Cereal type: Bran cereal

Disease of urinary system

Hernia
Phenotypes with male/female \( r_g \) significantly < 1 \( (p < 1e^{-5}) \)

- Trunk fat percentage* \( (P=8e^{-28}) \)
- Age first had sexual intercourse \( (P=1e^{-22}) \)
- Hip circumference \( (P=3e^{-16}) \)
- Waist circumference \( (P=4e^{-15}) \)
- Facial ageing \( (P=1e^{-13}) \)
- Body mass index (BMI) \( (P=6e^{-12}) \)
- Lifetime number of sexual partners \( (P=6e^{-11}) \)
- Average weekly beer plus cider intake* \( (P=5e^{-10}) \)
- Hand grip strength (right)* \( (P=2e^{-09}) \)
- Weight \( (P=2e^{-09}) \)
- Pulse rate, automated reading \( (P=1e^{-08}) \)
- Basal metabolic rate \( (P=1e^{-08}) \)
- Alcohol intake frequency \( (P=3e^{-08}) \)
- Haemoglobin concentration \( (P=6e^{-08}) \)
- Haematocrit percentage \( (P=7e^{-08}) \)
- Sleeplessness / insomnia \( (P=1e^{-07}) \)
- Heel Broadband ultrasound attenuation \( (P=4e^{-07}) \)
- Ankle spacing width \( (P=6e^{-07}) \)
- Cereal type: Bran cereal \( (P=6e^{-07}) \)
- Phys. activity in last 4 weeks: Other exercises \( (P=9e^{-07}) \)
- Past tobacco smoking \( (P=1e^{-06}) \)
- Hernia \( (P=1e^{-06}) \)
- Pain in last month: Neck or shoulder \( (P=3e^{-06}) \)
- Leisure/social activities: club / gym \( (P=4e^{-06}) \)
- Ever smoked \( (P=5e^{-06}) \)
Male GWAS

Facial aging

DOCK8

PAX1

Female GWAS

IRF4
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
Test center

Running your results through LD-score genetic correlation
Test Center

Please follow the steps to Upload file and Select data.

- 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 recommend traits with heritability z-score larger than 4.
  8. Remove SNPs with extremely 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

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
- Haematological 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
Browse previously generated results
# Heritability

## Lookup Center

Lookup existing LD score regression analysis results

### SNP Heritability 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):

<table>
<thead>
<tr>
<th>Trait name</th>
<th>H2</th>
<th>SE_H2</th>
<th>Z_H2</th>
<th>Lambda GC</th>
<th>Chi2</th>
<th>Intercept</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adiponectin</td>
<td>0.1369</td>
<td>0.0242</td>
<td>5.65702</td>
<td>1.068</td>
<td>1.09</td>
<td>1.0133</td>
</tr>
<tr>
<td>Age of smoking initiation</td>
<td>0.0665</td>
<td>0.0185</td>
<td>3.59459</td>
<td>1.0345</td>
<td>1.0295</td>
<td>0.9981</td>
</tr>
<tr>
<td>Child birth length</td>
<td>0.1697</td>
<td>0.0229</td>
<td>7.41048</td>
<td>1.0588</td>
<td>1.0672</td>
<td>0.9926</td>
</tr>
<tr>
<td>Child birth weight</td>
<td>0.1124</td>
<td>0.0179</td>
<td>6.27933</td>
<td>1.0466</td>
<td>1.0618</td>
<td>1.0043</td>
</tr>
<tr>
<td>Body mass index</td>
<td>0.1855</td>
<td>0.0089</td>
<td>20.8427</td>
<td>1.3675</td>
<td>1.4681</td>
<td>1.0188</td>
</tr>
<tr>
<td>Body fat</td>
<td>0.104</td>
<td>0.0076</td>
<td>13.6842</td>
<td>1.0315</td>
<td>1.0578</td>
<td>0.9983</td>
</tr>
<tr>
<td>Coronary artery disease</td>
<td>0.0728</td>
<td>0.0054</td>
<td>14.463</td>
<td>1.2386</td>
<td>1.3288</td>
<td>1.0475</td>
</tr>
</tbody>
</table>
Genetic correlation

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.int h2.obs_h2.int 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:

<table>
<thead>
<tr>
<th>Trait1</th>
<th>Trait2</th>
<th>rg</th>
<th>se</th>
<th>z</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADHD</td>
<td>Age at Menarche</td>
<td>-0.153</td>
<td>0.08218</td>
<td>-1.858</td>
<td>0.063</td>
</tr>
<tr>
<td>ADHD</td>
<td>Age at Smoking</td>
<td>-0.036</td>
<td>0.2427</td>
<td>-0.147</td>
<td>0.883</td>
</tr>
<tr>
<td>ADHD</td>
<td>Alzheimer’s</td>
<td>-0.055</td>
<td>0.2191</td>
<td>-0.249</td>
<td>0.803</td>
</tr>
<tr>
<td>ADHD</td>
<td>Anorexia</td>
<td>0.192</td>
<td>0.1162</td>
<td>1.649</td>
<td>0.099</td>
</tr>
<tr>
<td>ADHD</td>
<td>Autism Spectrum</td>
<td>-0.164</td>
<td>0.1438</td>
<td>-1.144</td>
<td>0.253</td>
</tr>
<tr>
<td>ADHD</td>
<td>BMI</td>
<td>0.287</td>
<td>0.08913</td>
<td>3.222</td>
<td>0.001</td>
</tr>
</tbody>
</table>
LD Hub practical

Sharing and exchanging GWAS results
We provided a list of existing GWAS resources here: (columns 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](#).

<table>
<thead>
<tr>
<th>File name</th>
<th>Trait name</th>
<th>Consortium/first_author/databse</th>
<th>Sample size</th>
<th>PMID</th>
<th>Publish year</th>
<th>Ethnicity</th>
</tr>
</thead>
<tbody>
<tr>
<td>adipogen.discovery.eur_meta_public.release.txt.nomHC.sumstats_deGC.gz</td>
<td>Adiponectin</td>
<td>ADIPOGen</td>
<td>39883</td>
<td>22479202</td>
<td>2012</td>
<td>Mixed</td>
</tr>
<tr>
<td>Age_of_smoking.sumstats.gz</td>
<td>Age of smoking initiation</td>
<td>TAG</td>
<td>47961</td>
<td>20418890</td>
<td>2010</td>
<td>European</td>
</tr>
<tr>
<td>Birthlength.sumstats.gz</td>
<td>Child birth length</td>
<td>EGG</td>
<td>28459</td>
<td>25281659</td>
<td>2015</td>
<td>European</td>
</tr>
<tr>
<td>Birthweight.sumstats.gz</td>
<td>Child birth weight</td>
<td>EGG</td>
<td>26836</td>
<td>23202124</td>
<td>2013</td>
<td>European</td>
</tr>
<tr>
<td>BMI_2010.sumstats_deGC.gz</td>
<td>Body mass index</td>
<td>GIANT</td>
<td>123912</td>
<td>20935630</td>
<td>2010</td>
<td>European</td>
</tr>
<tr>
<td>body_fat_percentage_GWAS.PLUS_MC.ALL_ancestry sexe_Sex_combined_for_locus_zoom_plot.TBL.txt.tab.sumstats.gz</td>
<td>Body fat</td>
<td>Lu</td>
<td>100716</td>
<td>26833246</td>
<td>2016</td>
<td>Mixed</td>
</tr>
</tbody>
</table>