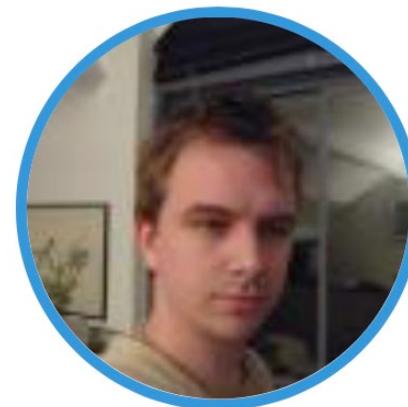


How do we go from genetic discoveries from GWAS/WGS/WES to mechanistic disease insight?

Danielle Posthuma

Part III – Pathway analysis using MAGMA



Christiaan de Leeuw

Practical session: Gene-set analysis using MAGMA

<https://ctg.cncr.nl/software/magma>

MAGMA gene-set analysis

- Software tool for gene and gene-set analysis
 - Command-line interface
- Input
 - Genotype and phenotype data
 - Or: (full) published GWAS results (plus reference data)
 - Gene definitions
 - Gene sets

Downloading summary statistics: GWAS ATLAS

<http://atlas.ctglab.nl>

GWASATLAS Home Browse GWAS Multiple GWAS comparison PheWAS Stats Documentation FAQ

Welcome to the Atlas of GWAS Summary Statistics

This atlas is a database of publicly available GWAS summary statistics. Each GWAS can be browsed with the manhattan plot, risk loci, MAGMA (i.e. gene-based) results, SNP heritability and genetic correlations with other GWAS in the database. 600 GWAS were performed in this project based on UK Biobank release 2 data under application ID 16406. Full summary statistics can be downloaded from the original source following the provided links.

If you have/find GWAS summary statistics that are publicly available and not included in this database, please let us know by contacting Kyoko Watanabe (k.watanabe@vu.nl).

Citation:
Watanabe, K. *et al.* A global view of pleiotropy and genetic architecture in complex traits. *bioRxiv* doi: <https://doi.org/10.1101/500090>

Currently the database contains **4,155** GWAS from **295** unique studies across **2,960** unique traits and **27** domains.

Browse GWAS

Overview of each GWAS such as Manhattan plots and QQ plot at SNP and gene levels and genetic correlations with other GWAS in the database.

[Browse GWAS](#)

Multiple GWAS comparison

Multiple GWAS can be compared in terms of genetic correlations, overlap of significant genes based on MAGMA gene-analysis and overlap of genetic risk loci.

[Multiple GWAS comparison](#)

PheWAS

PheWAS plot for a SNP or gene can be created across GWAS in the database.

[PheWAS](#)

What's new

2019-01-17
Second release of atlas database with in total of 4155 GWAS summary statistics (ID 3799-4155 are new GWASs). The last database curation was done in Oct 2018. Several updates have done for ID 1-3798 (see "DateLastModified" column in the database).

2018-12-20
Preprint is now available (doi: <https://doi.org/10.1101/500090>). The second release of the atlas database (total 4155 GWAS) will be available soon.

2018-09-28
First release of atlas database with 3798 GWAS summary statistics. The last database curation was done in Aug 2017. We will update summary statistics published since then, in near future. Publication is under preparation.

2017-12-18
First internal release of the website.

Downloading summary statistics: GWAS ATLAS

<http://atlas.ctglab.nl>

GWAS ATLAS

Home Browse GWAS Multiple GWAS comparison PheWAS Stats Documentation FAQ

atlas ID: 3785 Intelligence

GWAS information

Feature	Value
id	3785
PMID	PMID: 29942086
Year	2018
File ?	https://ctg.cncr.nl/documents/p1651/SavageJansen_intMeta_sumstats.zip
Website	https://ctg.cncr.nl/software/summary_statistics
Consortium	
Domain	Cognitive
ChapterLevel	Mental Functions
SubchapterLevel	Higher-Level Cognitive Functions
Trait	Intelligence
uniqTrait	Intelligence
Population ?	UKB2 (EUR meta)
Ncase	
Ncontrol	
N	269867
Nsnps ?	9295118

SNPs plots

Manhattan plot

i For plotting, overlapping data points are not drawn (filtering was performed only for SNPs with P-value $\geq 1e-5$, see documentation for more details of filtering).

Download the plot as [PNG](#) [JPG](#) [SVG](#) [PDF](#)

Q-Q plot

i For plotting purposes, overlapping data points are not drawn (filtering was performed only for SNPs with P-value $\geq 1e-5$, see documentation for details of filtering).

Download the plot as [PNG](#) [JPG](#) [SVG](#) [PDF](#)

Top SNPs

i Top SNPs are defined as the most significant SNP in a genomic risk locus. See documentation for details of definition of the genomic risk loci.

Download the table as [csv](#)

Show entries Search:

CHR	POS	rsID	P
1	22425642	rs10917152	2.227e-09
1	32106494	rs7546297	1.332e-12
1	41750648	rs12035012	3.675e-16

MAGMA gene-set analysis

- Three main steps
 - Annotation: map SNPs onto genes
 - Gene analysis: compute association of genes with phenotype
 - Gene-set analysis: test gene associations in gene sets
- Generalized gene-set analysis
 - Continuous 'sets'
 - Conditional and joint analysis
 - Interaction analysis

Annotation

Map SNPs to a gene based on physical location

- If located inside the transcription region of the gene
- Optionally, if located in window around the gene
 - Especially upstream of transcription start site
- A SNP can be mapped to multiple genes

Gene-based analysis

- Four models available in MAGMA
 - Principal component linear regression
 - Performs test on explained phenotypic variance (F-test)
 - Requires raw genotype data
 - SNP-wise models: compute SNP associations with phenotype first
 - SNP-wise Mean: performs test on mean SNP association
 - SNP-wise Top: performs test on strongest SNP association
 - SNP-wise Multi: combines SNP-wise Top and Mean
- The main question for all gene analysis: what sensitivity do you want?
 - Also with regard to allele frequency, functional annotation, etc.
 - In general, no 'best' or most powerful model

Gene-set analysis

Gene ID	Association	In gene set
1	1.32	Yes
2	-0.76	Yes
3	0.48	Yes
4	1.12	Yes
5	-0.02	Yes
6	-1.04	No
7	0.86	No
8	-1.27	No
9	0.41	No
10	0.11	No

An analysis of genes:

- Genes are data points in the analysis
- The gene set is a grouping variable
- Genetic association with the phenotype is the outcome variable

So: gene-set analysis is like a t-test

- Testing the mean association of genes in the gene set
 - One-sided test
- Two kinds of t-test...

Gene-set analysis

Gene ID	Association	In gene set
1	1.32	Yes
2	-0.76	Yes
3	0.48	Yes
4	1.12	Yes
5	-0.02	Yes
6	-1.04	No
7	0.86	No
8	-1.27	No
9	0.41	No
10	0.11	No

Diagram illustrating the grouping of genes into two sets for analysis:

- Genes 1 through 5 are grouped together, labeled μ_S (the gene set).
- Genes 6 through 10 are grouped together, labeled μ_0 (genes outside the gene set).

Self-contained analysis:

- Is the mean genetic association of genes in the gene set greater than 0?

- $H_0: \mu_S = 0$

Competitive analysis:

- Is the mean genetic association of genes in the gene set greater than that of genes outside the gene set?

- $H_0: \mu_S = \mu_0$

Clinical study

Patient	Improvement	Treatment
1	1.32	Yes
2	-0.76	Yes
3	0.48	Yes
4	1.12	Yes
5	-0.02	Yes
6	-1.04	No
7	0.86	No
8	-1.27	No
9	0.41	No
10	0.11	No

μ_T

μ_C

Self-contained analysis:

- Is the mean **improvement of patients in the treatment group** greater than 0?
 - $H_0: \mu_T = 0$

Competitive analysis:

- Is the mean **improvement of patients in the treatment group** greater than that of **patients in the control group**?
 - $H_0: \mu_T = \mu_C$

Gene-set analysis

Gene ID	Association	In gene set
1	1.32	Yes
2	-0.76	Yes
3	0.48	Yes
4	1.12	Yes
5	-0.02	Yes
6	-1.04	No
7	0.86	No
8	-1.27	No
9	0.41	No
10	0.11	No

μ_S

μ_0

Self-contained analysis:

- Is the mean genetic association of genes in the gene set greater than 0?

- $H_0: \mu_S = 0$

You need a baseline or reference group to say anything about the treatment, or the gene set

Gene-set analysis

Gene ID	Association	In gene set
1	1.32	Yes
2	-0.76	Yes
3	0.48	Yes
4	1.12	Yes
5	-0.02	Yes
6	-1.04	No
7	0.86	No
8	-1.27	No
9	0.41	No
10	0.11	No

μ_S

μ_0

Competitive analysis:

- Is the mean genetic association of genes in the gene set greater than that of genes outside the gene set?

- $H_0: \mu_S = \mu_0$

Only competitive analysis allows any inference about the gene set itself

See you at the MAGMA practical!
