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

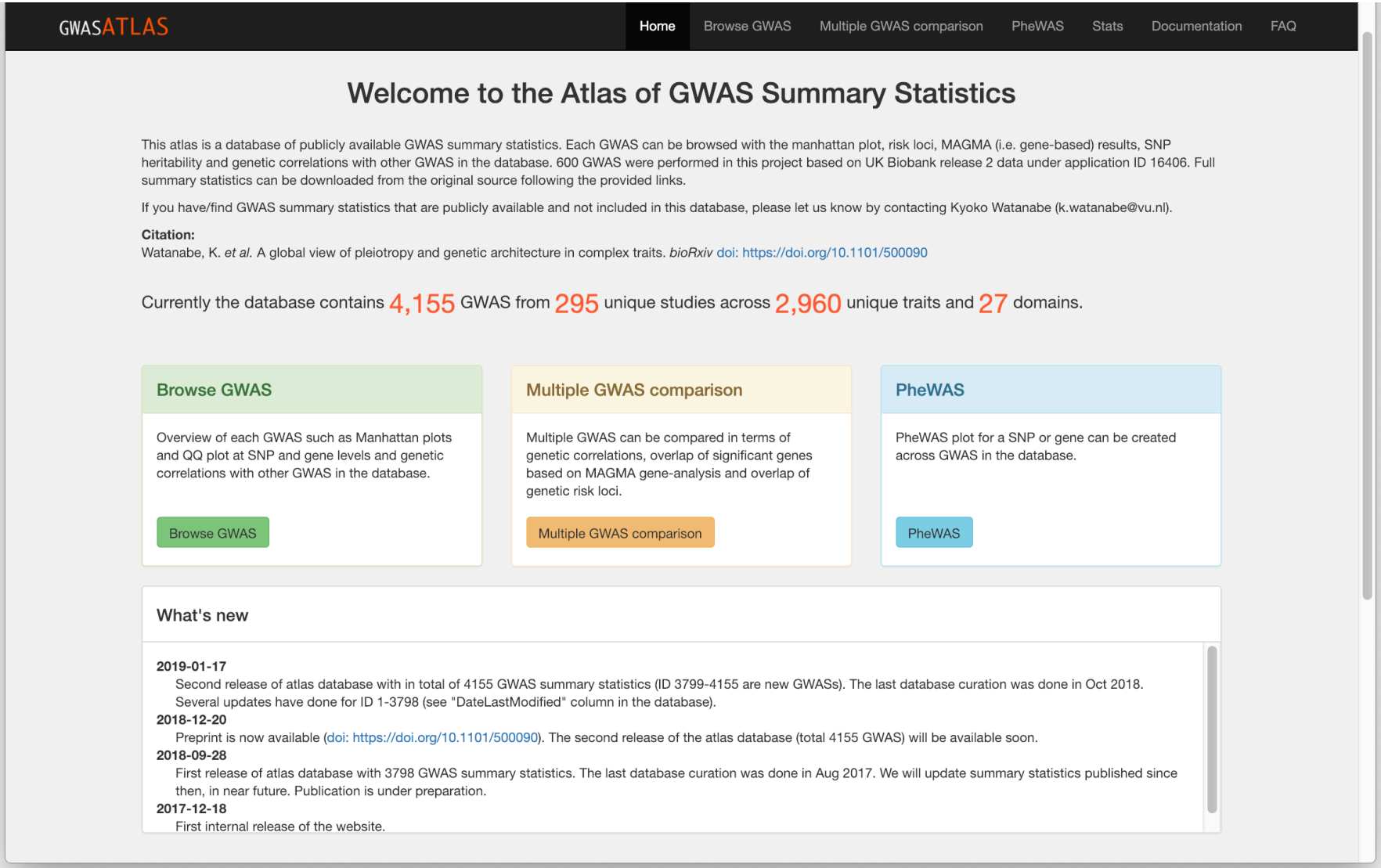
MAGMA gene-set analysis

- 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

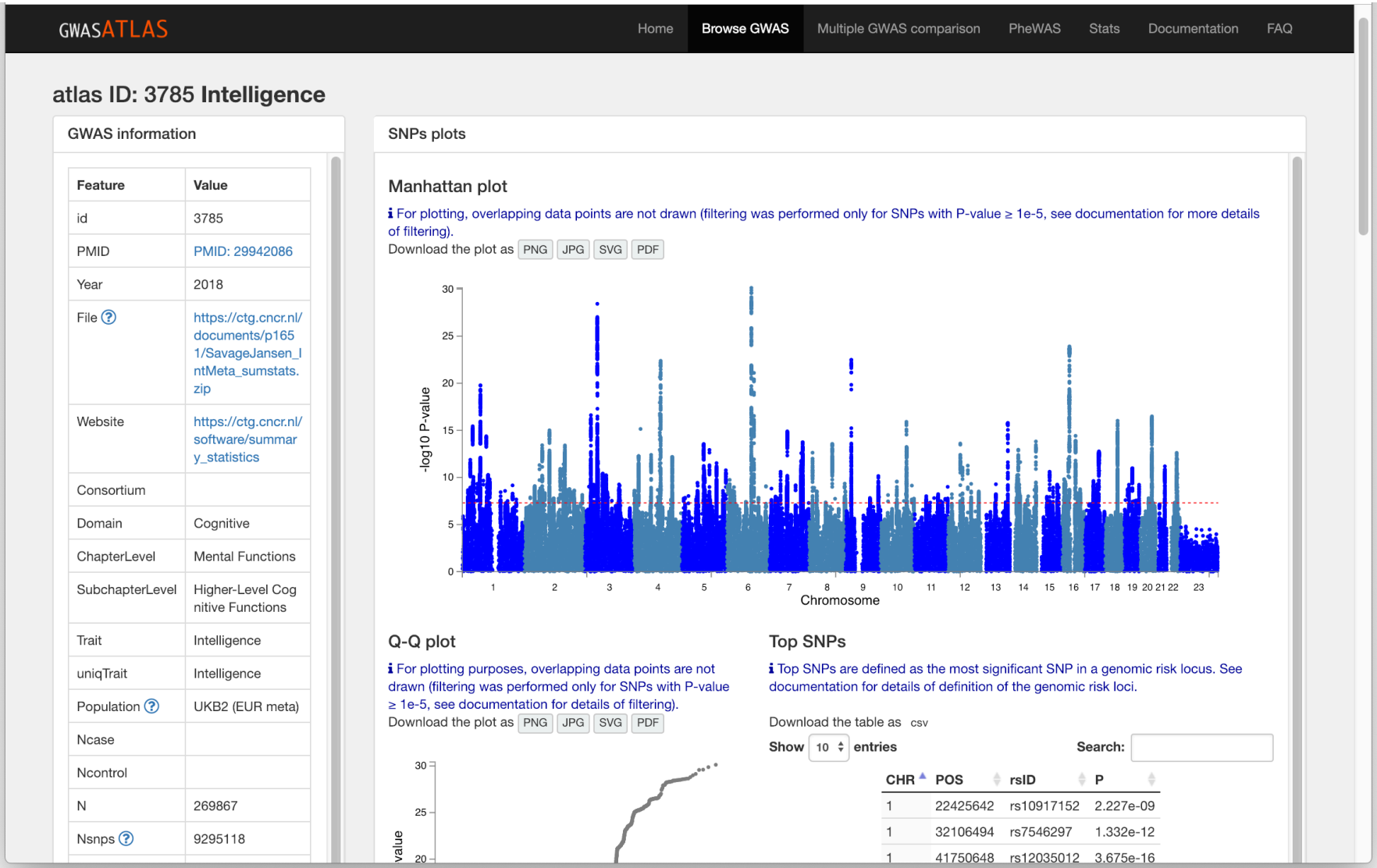
Downloading summary statistics: GWAS ATLAS

http://atlas.ctglab.nl



Downloading summary statistics: GWAS ATLAS

http://atlas.ctglab.nl



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	μ_S
4	1.12	Yes	
5	-0.02	Yes	
6	-1.04	No	
7	0.86	No	
8	-1.27	No	μ_0
9	0.41	No	
10	0.11	No	

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	μ_S
2	-0.76	Yes	
3	0.48	Yes	
4	1.12	Yes	
5	-0.02	Yes	
6	-1.04	No	μ_0
7	0.86	No	
8	-1.27	No	
9	0.41	No	
10	0.11	No	

Self-contained analysis:

- Is the [illegible] 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	μ_S
2	-0.76	Yes	
3	0.48	Yes	
4	1.12	Yes	
5	-0.02	Yes	
6	-1.04	No	μ_0
7	0.86	No	
8	-1.27	No	
9	0.41	No	
10	0.11	No	

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!