FUMA – for GWAS follow-up

WINTER SCHOOL

Dr Fleur Garton

FUMA GWAS

Functional Mapping and Annotation of Genome-Wide Association Studies

FUMA is a platform that can be used to annotate, prioritize, visualize and interpret GWAS results.
The SNP2GENE function takes GWAS summary statistics as an input, and provides extensive functional annotation for all SNPs in genomic areas identified by lead SNPs.
The GENE2FUNC function takes a list of gene IDs (as identified by SNP2GENE or as provided manually) and annotates genes in biological context
To submit your own GWAS, login is required for security reason. If you have't registered yet, you can do from here.
You can browse public results of FUMA (including example jobs) from Browse Public Results without registration or login.

Please post any questions, suggestions and bug reports on Google Forum: FUMA GWAS users.
If you would like to be in the mailing list, please send an email to k.watanabe@vu.nl. Only major updates will be announced through email (low traffic).

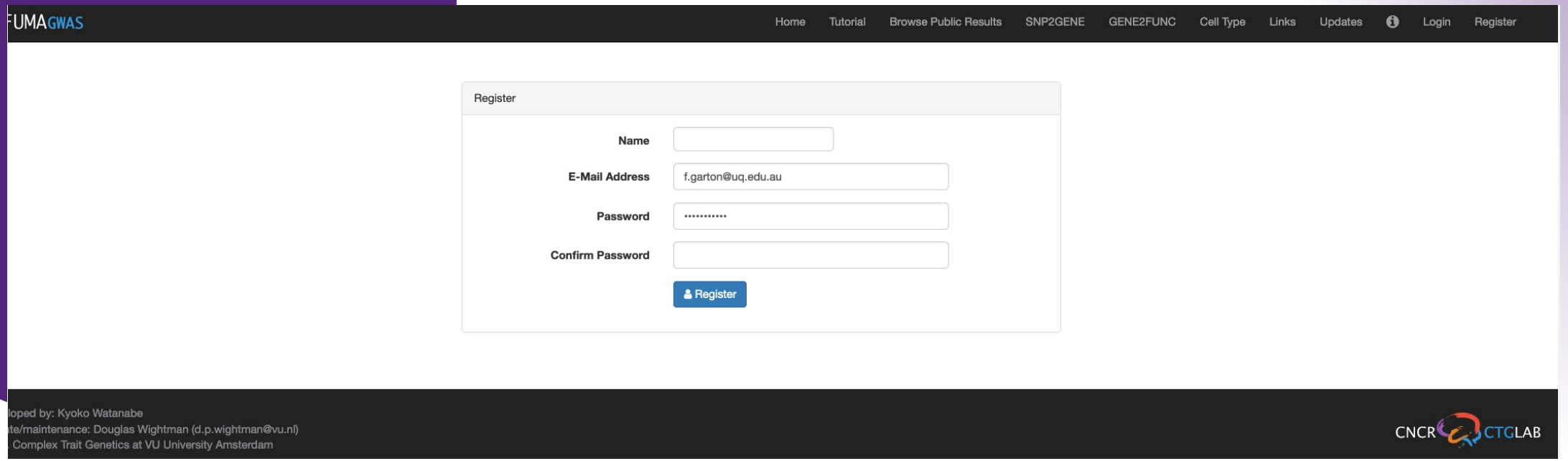
Citation:
When using FUMA, please cite the following.
K. Watanabe, E. Taskesen, A. van Bochoven and D. Posthuma. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**:1826. (2017).
https://www.nature.com/articles/s41467-017-01261-5
When using cell type analysis, please cite the following.
K. Watanabe, M. Umicevic Mirkov, C. de Leeuw, M. van den Heuvel and D. Posthuma. Genetic mapping of cell type specificity for complex traits. *Nat. Commun.* **10**:3222. (2019).
https://www.nature.com/articles/s41467-019-11181-1
Depending on which results you are going to report, please also cite the original study of data sources/tools used in FUMA (references are available at links or tutorial for the cell type specificity analysis for scRNA-seq data).

SNP2GENE

GENE2FUNC

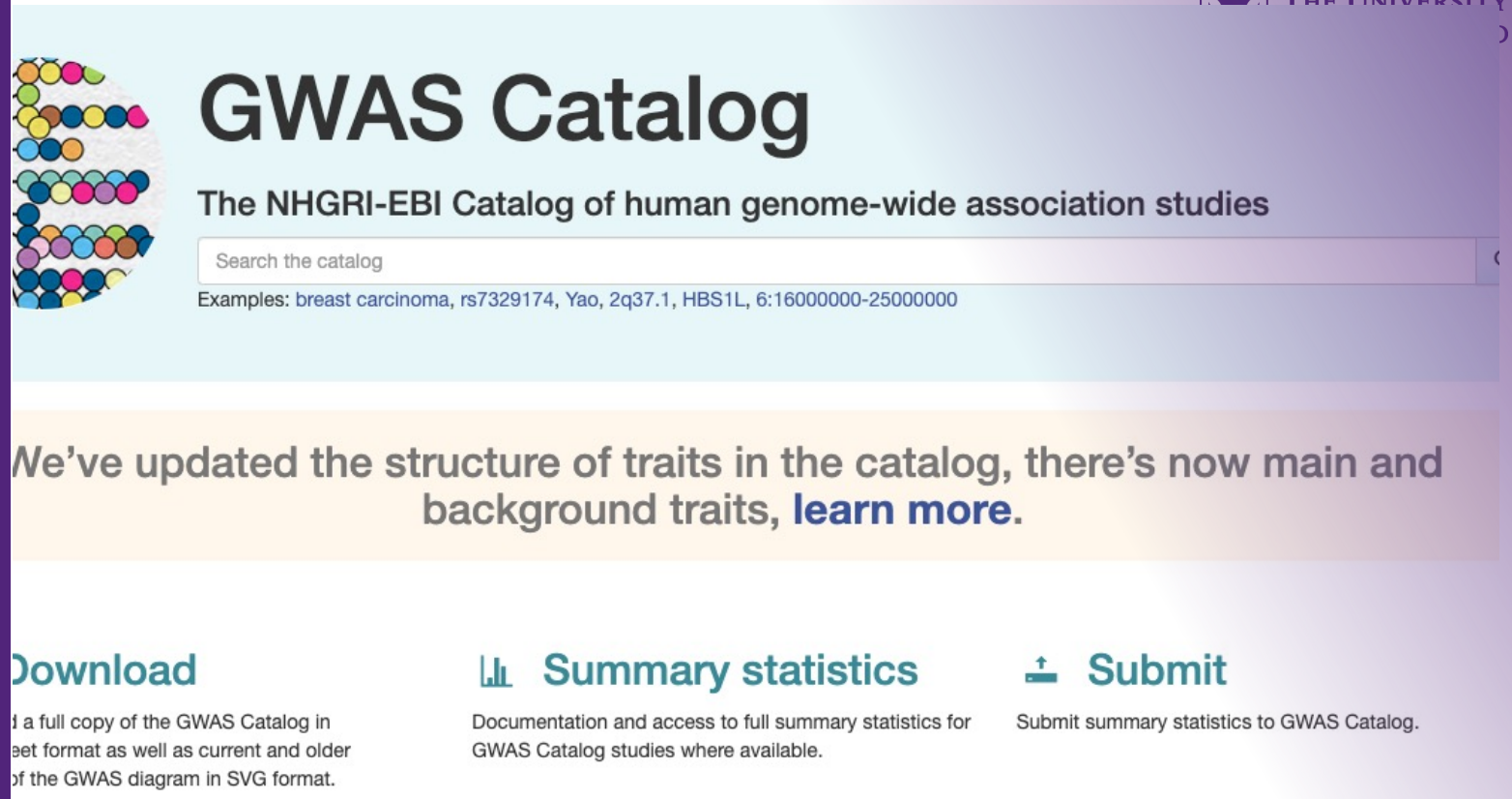
FUMA Registration

https://fuma.ctglab.nl/



Download data GWAS catalogue

Study: GCST90027163



Index of /pub/databases/gwas/summary_statistics/GCST90027001-GCST90028000/GCST90027163/

../		
harmonised/	29-Mar-2022 11:39	-
GCST90027163_buildGRCh37.tsv.gz	07-Sep-2021 23:22	176640530
README.txt	07-Sep-2021 23:22	28
md5sum.txt	07-Sep-2021 23:22	109

WHY?

- **Complex traits/diseases are generally highly polygenic, with many risk variants each contributing a small incremental effect**
- **To gaining biological insight from GWASs follow-up is needed**
- **Additional data integration…**

N_{cases}=29,612
$N_{controls}$=122,656

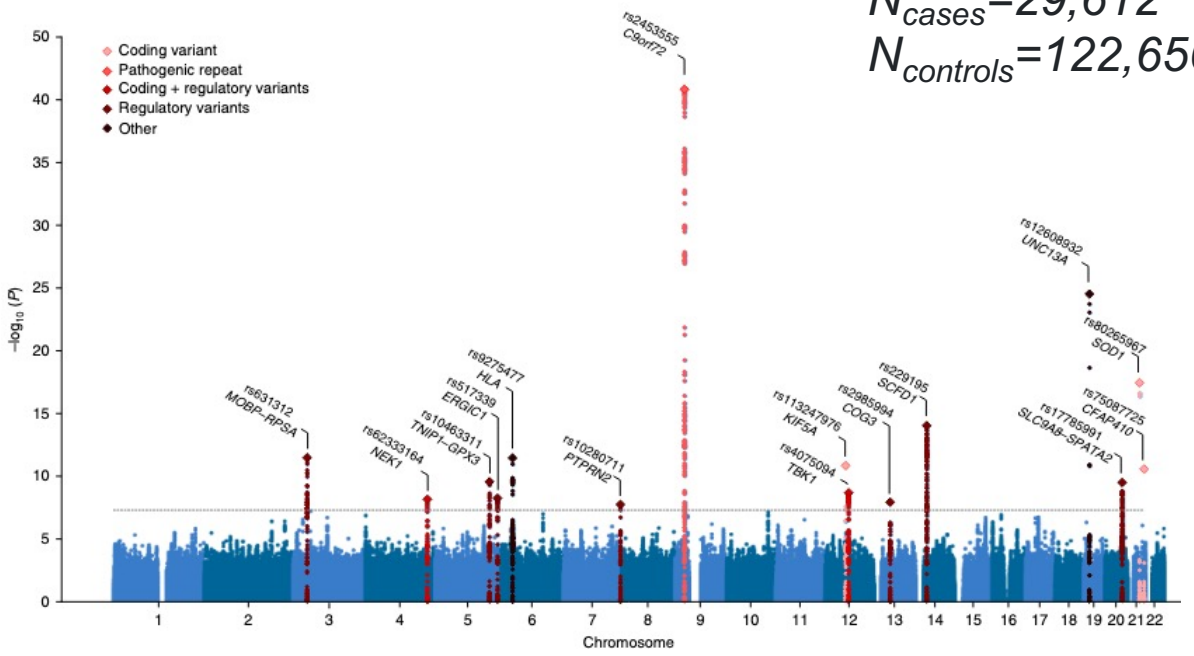


Figure 2. ALS Genome wide association study

Each dot represents a SNP, the x-axis shows the chromosomes where each SNP is located, and the y-axis shows $-log_{10}$ P-value of the association of each SNP with ALS

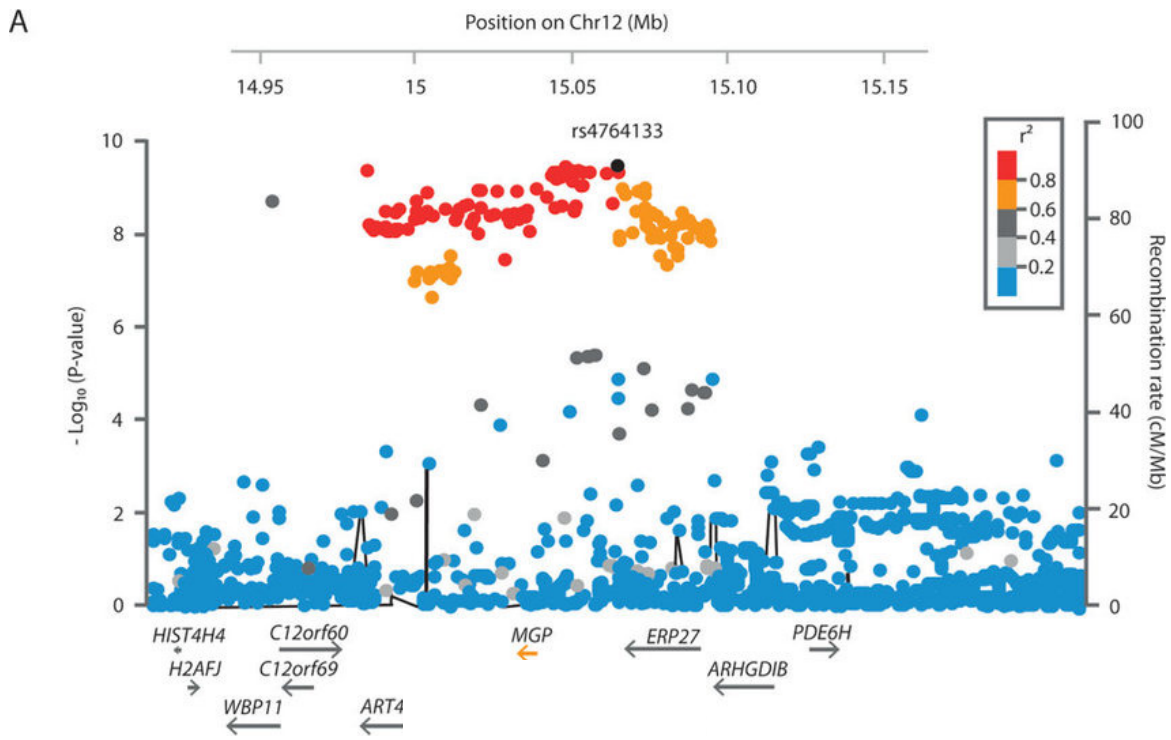
The horizontal dashed line reflects the genome-wide significant threshold (P= 5e-8)
Fifteen regions

Each locus is classified into one of the four classes: coding variants, pathogenic repeat, coding + regulatory variants and regulatory variants

EXAMPLE..

- **Hits are typically regions (‘loci’)**
  - multiple correlated SNPs
  - multiple closely-located genes

Pinpointing which and how genes are affected by SNPs associated with a trait /disease is necessary to increase insight into the biological mechanisms underlying that trait.

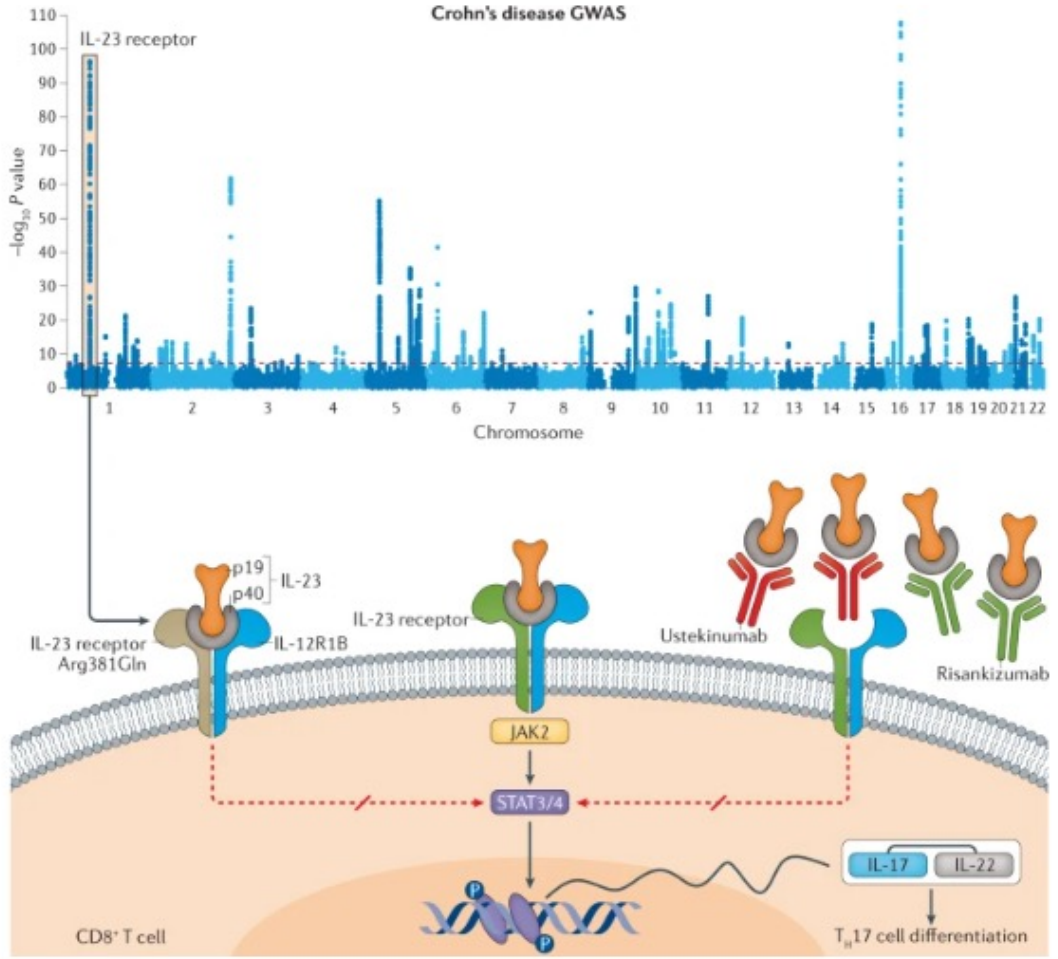


HOW?

To identify the most likely relevant, causal genes and variants required data integration from different repositories to interpret in broader context of genes and pathways

- **Linkage disequilibrium (LD) patterns**
- **Functional consequences of correlated SNPs (deleteriousness of variants, effects on gene expression, role in chromatin interaction sites)**
- **→ can lead to new treatments/drug-repositioning**
- rs11209026 missense - IL23 receptor Protective effect- pharmacological inhibition of this gene of value
- Two central monoclonal antibodies modulating IL-23 signalling - ustekinumab and Risankizumab (psoriasis)
- Ustekinumab now approved in United States, Europe and Australia

Fig. 1: Genome-wide significant variants associated with Crohn's disease spanning the IL-23 receptor provide drug repurposing opportunities.



Raey et al. 2021 https://www.nature.com/articles/s41576-021-00387-z

FUMA

PURPOSE: use functional, biological information to prioritize genes/mechanisms from GWAS

- **Annotate**
- **Prioritize**
- **Visualize**
- **Interpret**

SNP2GENE – input GWAS summary statistics for extensive functional annotation for all SNPs in genomic areas identified by lead SNPs

GENE2FUNC function takes a list of gene IDs (as identified by SNP2GENE or as provided manually) and annotates genes in biological context

FUMA version: v1.3.8
Total users: 10859 / Total SNP2GENE jobs: 190923/ Total GENE2FUNC jobs: 84227

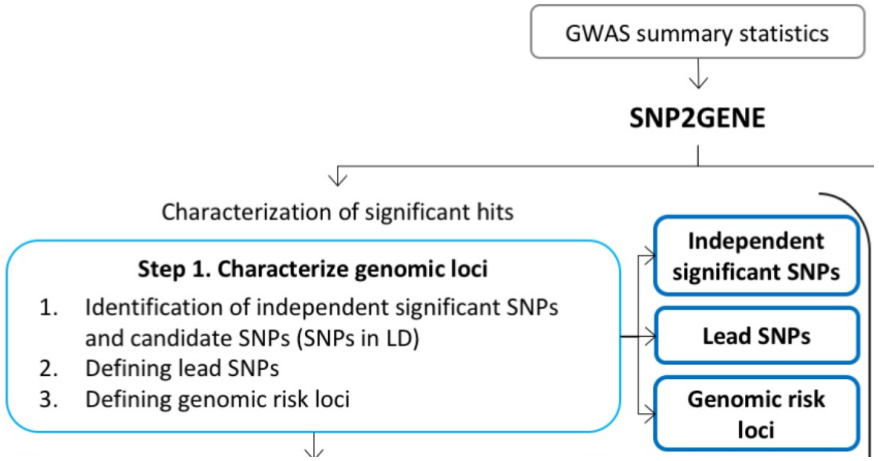
SNP2GENE is the core function

Tools/datasets incorporated into FUMA

Data source/tool	Used for
1000 Genome Project Phase 3	Reference panel used to compute r2 and MAF.
PLINK v1.9	Used to compute r2 and MAF.
MAGMA v1.08	Used for gene analysis and gene-set analysis.
ANNOVAR	A variant annotation tool used to obtain functional consequences of SNPs on gene functions.
CADD v1.4	A deleterious score of variants computed by integrating 63 functional annotations. The higher the score, the more deleterious.
RegulomeDB v1.1	A categorical score to guide interpretation of regulatory variants.
15-core chromatin state	Chromatin state for 127 epigenomes was learned by ChromHMM derived from 5 chromatin markers (H3K4me3, H3K4me1, H3K36me3, H3K27me3, H3K9me3).
GTEx v6/v7/v8	eQTLs and gene expression used in the pipeline were obtained from GTEx.
Blood eQTL Browser	eQTLs of blood cells. Only cis-eQTLs with FDR ≤ 0.05 are available in FUMA.
BIOS QTL browser	eQTLs of blood cells in Dutch population. Only cis-eQTLs (gene-level) with FDR ≤ 0.05 are available in FUMA.
BRAINEAC	eQTLs of 10 brain regions. Cis-eQTLs with nominal P-value < 0.05 are available in FUMA.
MuTHER	eQTLs in Adipose, LCL and Skin samples (only cis eQTLs).
xQTLServer	eQTLs in dorsolateral prefrontal cortex samples.
CommonMind Consortium	eQTLs in brain samples. Both cis and trans eQTLs are available
eQTLGen	Meta-analysis of cis and trans eQTLs based on 37 data sets (in total of 31,684 individuals).
DICE	eQTLs of 15 types of immune cells.
van der Wijst et al. scRNA eQTLs	eQTLs based on scRNA-seq of 9 cell types.
PsychENCODE	SNP annotations (enhancer, H3K27ac markers), eQTLs and HiC based enhancer-promoter interactions.
eQTL Catalogue	Gene level eQTL data generated from a variety of studies, where all of the eQTL datasets were produced in a uniform manner.
EyeGEx	cis-eQTLs from retina.
FANTOM5	SNP annotations (enhancer and promoter) and enhancer-promoter correlations.
BrainSpan	Gene expression data of developmental brain samples.
GSE87112 (Hi-C)	Hi-C data (significant loops) of 21 tissue/cell types. Pre-processed data (output of Fit-Hi-C) is used in FUMA.
Giusti-Rodriguez et al. 2019 (Hi-C)	Hi-C data (significant loops) of adult and fetal cortex. Only significant loops after Bonferroni correction (Pbon < 0.001) are available.
Enhancer and promoter regions	Predicted enhancer and promoter regions (including dyadic) from Roadmap Epigenomics Projects. 111 epigenomes are available.
MsigDB v7.0	Collection of publicly available gene sets. Data sets include e.g. KEGG, Reactome, BioCarta, GO terms and so on.
WikiPathways v20191010	The curated biological pathways.
GWAS-catalog e104_2021-09-15	A database of reported SNP-trait associations.
DrugBank v5.1.4	Targeted genes (protein) of drugs in DrugBank was obtained to assign drug ID for input genes.
pLI	A gene score annotated to prioritized genes. The score is the probability of being loss-of-function intolerance.
ncRVIS	A gene score annotated to prioritized genes. The score is the non-coding residual variation intolerance score.

Yellow = expression data
Purple = chromatin
Pink = annotation (SNPs, genes, other traits)
Grey = computation/reference panels
Green = pathways/drug targets

Overview



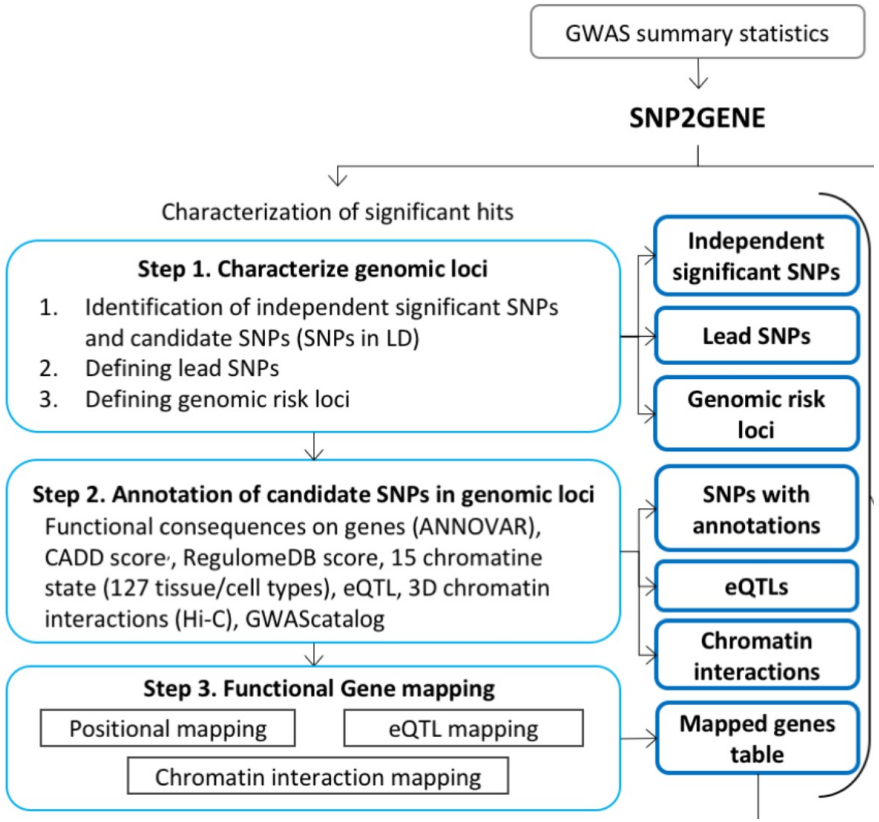
GWAS - Significant SNPs – generally p< 5x10-8
FUMA –

- Using LD structure from 1000G of the relevant population – the independent (sig) SNPs are identified.
- Then - all known (i.e., regardless of being available in the GWAS input) SNPs that have $r^2 \geq 0.6$ are annotated as candidates

Means that SNPs in LD with significant loci not available in the GWAS input, (but in 1000G reference panel) can be annotated

Default MAF >=0.01.
User – can define loci/snps to be annotated

Overview



CADD- combined annotation-dependant depletion - scores the deleteriousness of single nucleotide variants as well as insertion/deletions variants in the human genome (high=bad, 12.37 is threshold)

Chromatin state- 127 tissue/cell types – i.e. is the region typically 'open/transcribed'

RegulomeDB – whether likely to affect TF binding states 1-6

eQTL – associated with expression of gene (different datasets – blood/tissues/cells)

Chromatin-interactions (Hi-C on different cells)

GWAS catalogue- reported SNPS with trait/disease

Overview

Gene mapping

3 ways:

Positional (10kb)

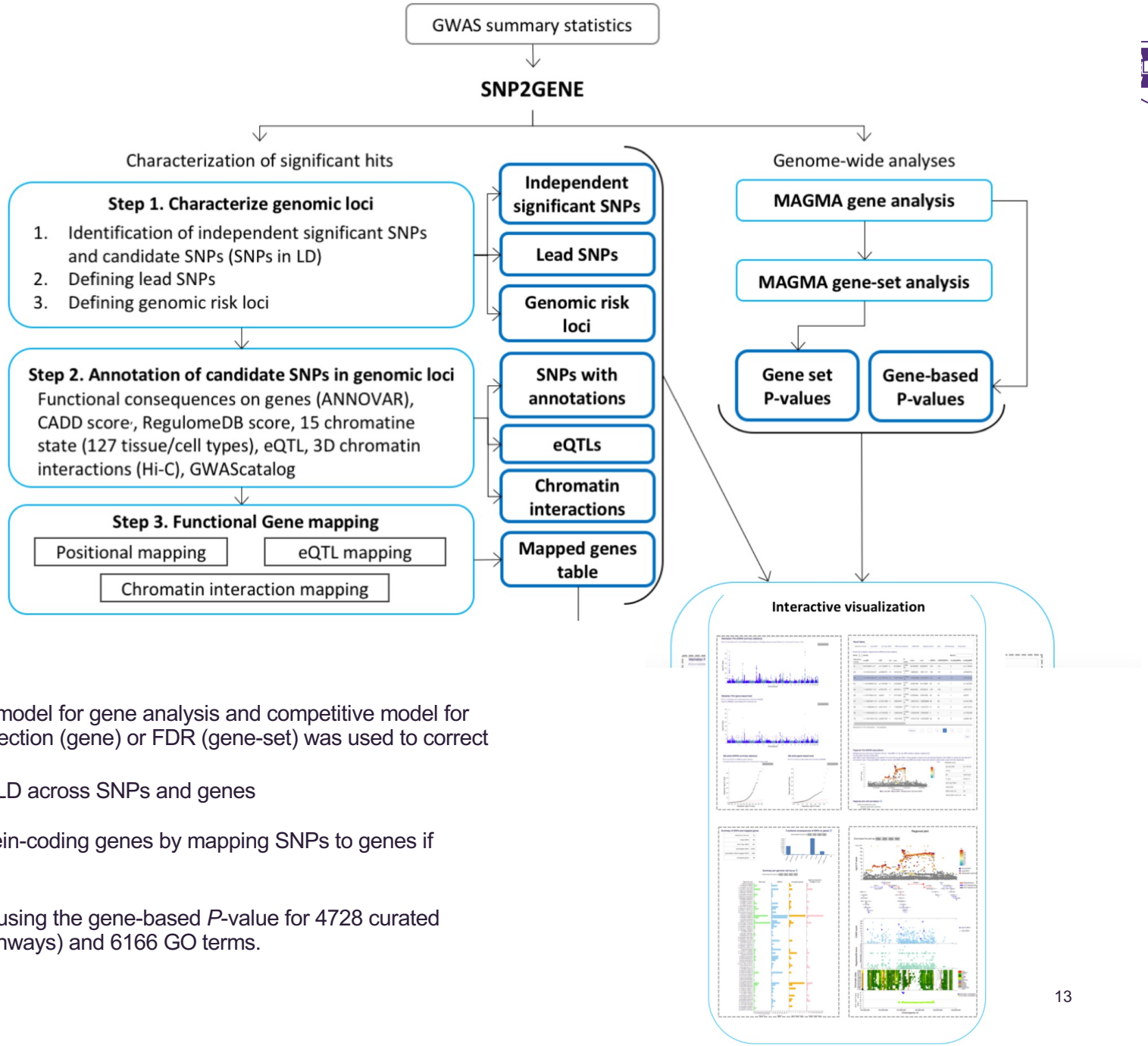
eQTL (FDR <0.05)

Chromatin
(significantly interacting regions in user-selected tissue/cell types)

Mapped genes are annotated pLI and ncRVIS

Category	Mandatory/Optional	Parameter	Description	Default
Gene mapping	Mandatory	positional mapping	Whether perform positional mapping	TRUE
	Optional	distance based mapping	To map SNPs to genes based on physical distance	Yes
	Mandatory if distance based mapping is active	window (<)	The maximum distance to map SNPs to genes	10kb
	Optional	annotation	SNPs positional annotation to map to genes	all
	Mandatory	eQTL mapping	Whether perform eQTL mapping	FALSE
	Mandatory if eQTL mapping is true	tissue	Tissue type to use for eQTL mapping	all
	Mandatory if eQTL mapping is true	significant eQTL only	Whether map only significant eQTL at FDR < 0.05	TRUE
	Optional if significant eQTL only is FALSE	eqtlP (<)	The maximum (nominal) P-value of eQTL	1.00E-03
	Mandatory	chromatin interaction mapping	Whether perform chromatin interaction mapping	FALSE
	Optional	build in chromatin interaction data	Build in Hi-C data for chromatin interaction mapping	None
	Mandatory if chromatin interaction mapping is true	FDR of interaction (<)	The maximum FDR of significance of chromatin interactions	1.00E-06
	Mandatory if chromatin interaction mapping is true	Promoter regions	User defined promoter region around TSS of genes	250bp up- and 500bp down-stream of TSS
	Optional	Cell/tissue types of enhancer and promoter regions	Cell/tissue types of 111 epigenomes to annotate predicted enhancer and promoter regions to interacting regions	None
	Optional	Use enhancers for filtering	Whether filter SNPs in one end of a significant interaction on which are overlapped with enhancer regions of selected epigenomes	FALSE
	Optional	Use promoters for filtering	Whether filter genes whose promoter regions are overlapped with predicted promoter regions of selected epigenomes	FALSE
	Optional*	CADD (>=)	The minimum CADD score	0
	Optional*	RDB (<=)	The minimum RegulomeDB score	7
	Optional*	Chromatin state filtering	Whether filter SNPs based on chromatin 15 states	FALSE
	Optional*	Cell/tissue types	Cell/tissue type of chromatin state	none
	Optional*	chromatin state (<=)	The minimum chromatin state	7
	Optional*	Method for chromatin state filtering	When multiple tissue/cell types are selected, this method will apply to filter SNPs	none

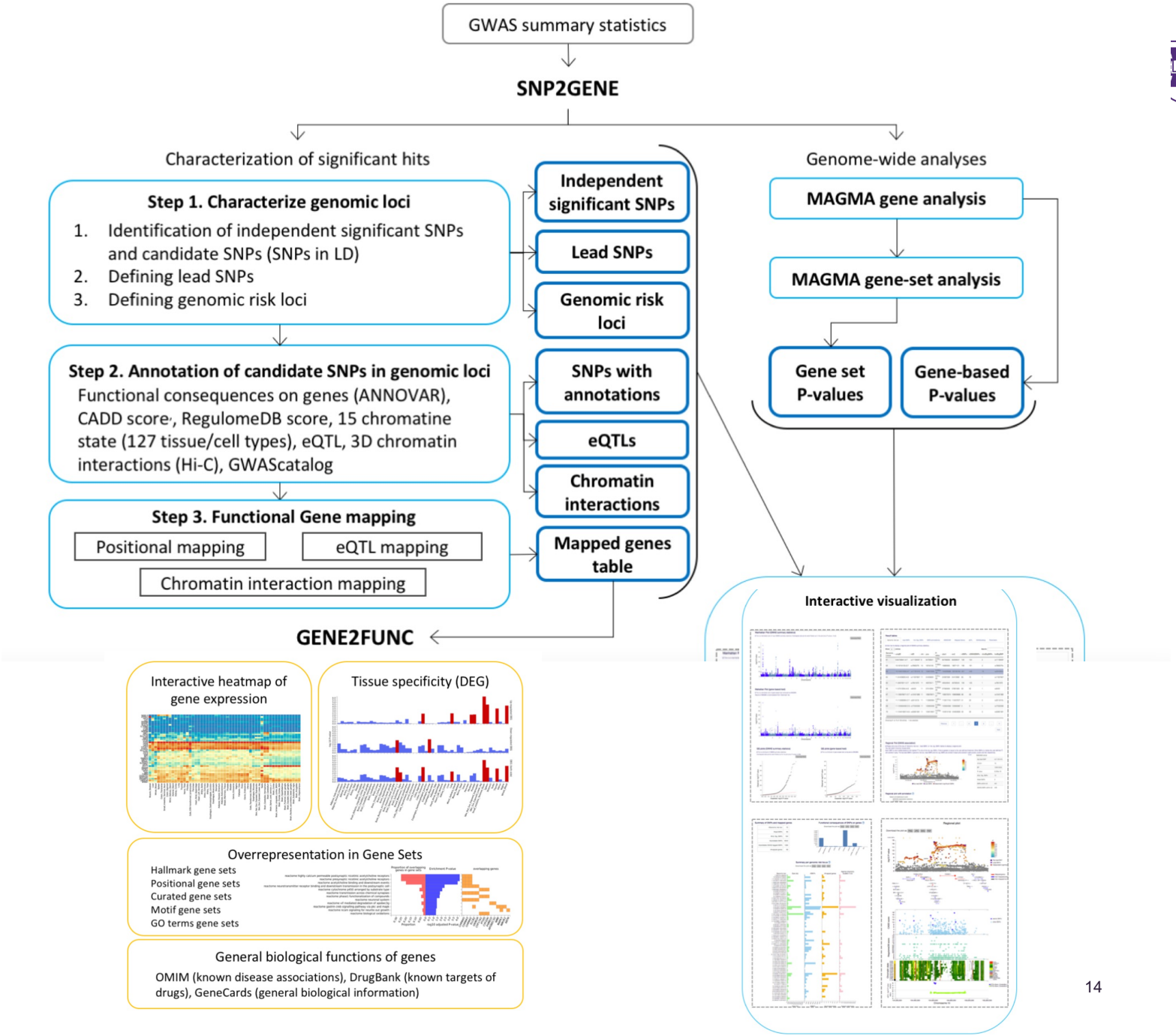
Overview



- MAGMA (not the only gene-based test!)
- Default settings from MAGMA (SNP-wise model for gene analysis and competitive model for gene set analysis) and the Bonferroni correction (gene) or FDR (gene-set) was used to correct for multiple testing.
- 1000G phase 3 is the reference panel for LD across SNPs and genes

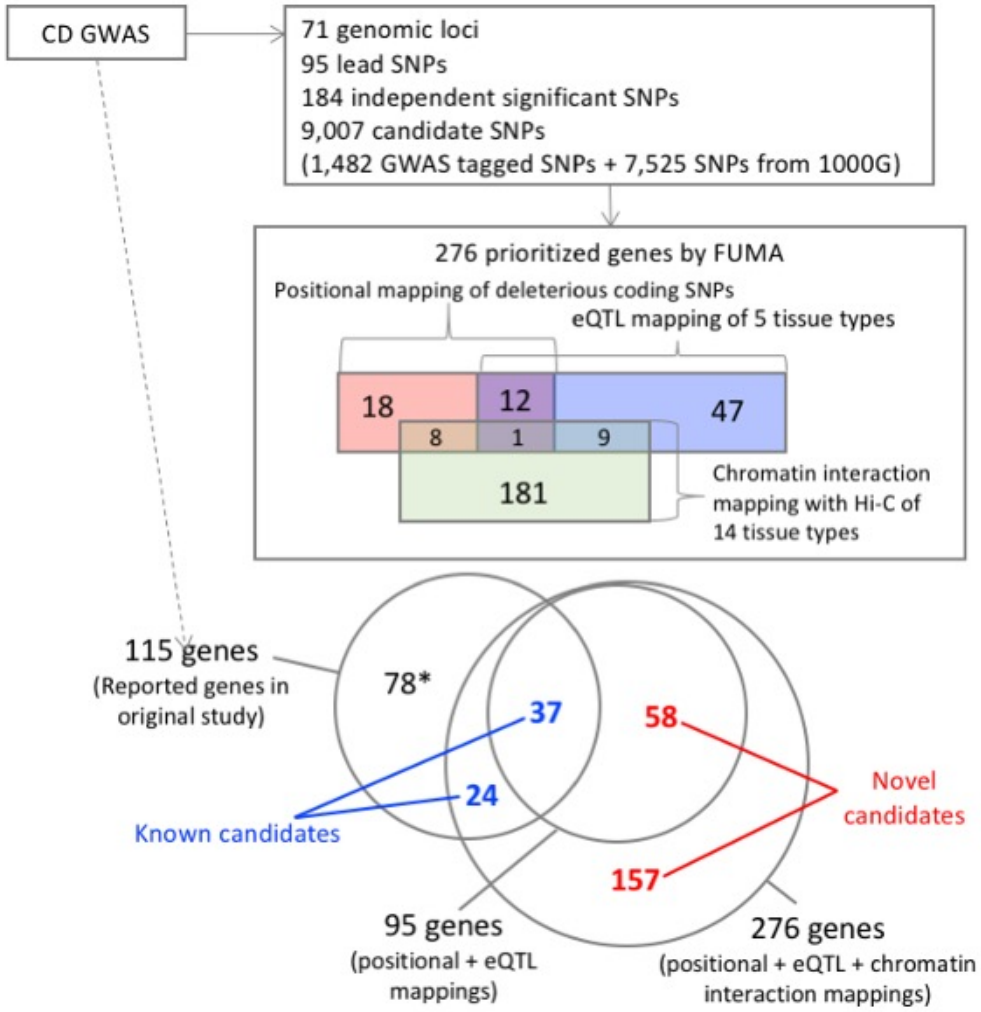
- Gene-based *P*-value is computed for protein-coding genes by mapping SNPs to genes if SNPs are located within the genes
- For gene set analysis, MsigDB v5.2.
  - the gene set *P*-value is computed using the gene-based *P*-value for 4728 curated gene sets (including canonical pathways) and 6166 GO terms.
  - For both analyses, the.

Overview


GWAS summary statistics
SNP2GENE
Characterization of significant hits
Step 1. Characterize genomic loci
1. Identification of independent significant SNPs and candidate SNPs (SNPs in LD)
2. Defining lead SNPs
3. Defining genomic risk loci
Step 2. Annotation of candidate SNPs in genomic loci
Functional consequences on genes (ANNOVAR), CADD score, RegulomeDB score, 15 chromatine state (127 tissue/cell types), eQTL, 3D chromatin interactions (Hi-C), GWAScatalog
Step 3. Functional Gene mapping
Positional mapping
eQTL mapping
Chromatin interaction mapping
Independent significant SNPs
Lead SNPs
Genomic risk loci
SNPs with annotations
eQTLs
Chromatin interactions
Mapped genes table
Genome-wide analyses
MAGMA gene analysis
MAGMA gene-set analysis
Gene set P-values
Gene-based P-values
Interactive visualization
GENE2FUNC
Interactive heatmap of gene expression
Tissue specificity (DEG)
Overrepresentation in Gene Sets
Hallmark gene sets
Positional gene sets
Curated gene sets
Motif gene sets
GO terms gene sets
General biological functions of genes
OMIM (known disease associations), DrugBank (known targets of drugs), GeneCards (general biological information)

Rationale for use?

- **Sanity check**
- **Pick up additional genes – that could be relevant**
- **Change parameters of original study**
- **Quick easy/minimal resources required**



Limitations

- Time
- Tools

Tips

- Changes from default you might want to make
  - i.e. positional mapping - selecting exonic and splicing SNPs with CADD score ≥ 12.37
  - i.e. eQTL mapping -using GTEx eQTLs with FDR<0.05
  - i.e. Chromatin mapping Hi-C data from Schmitt et al. (14 for BMI, 5 for CD) & interactions filtered by FDR<1e-6
  - Gene-mapping - protein-coding only, test enrichment of DEG in 53 tissue types, Canonical Pathways & GO terms

- Input – understand what each dataset is (highly recommend supplement of initial publication)
  - & if you want to include your own input – the supplement explains the format
  - Also recommend reading relevant linked publication

- Results download –
  - Lots of txt files..
  - each output i.e. “genes.txt” has headers described in the supplementary information file from the paper

PRACTICAL

1. **Set-up and queue your own job**

2. **Explore results**
    1. SNP2GENE
    2. GENE2FUNC

3. At home - Alter parameters by re-running job, explore different datasets, functionalities

Lots of ‘user’ input opportunities
Difficult to make decisions without knowing what the data is
Take the input selection process slowly

PRACTICAL

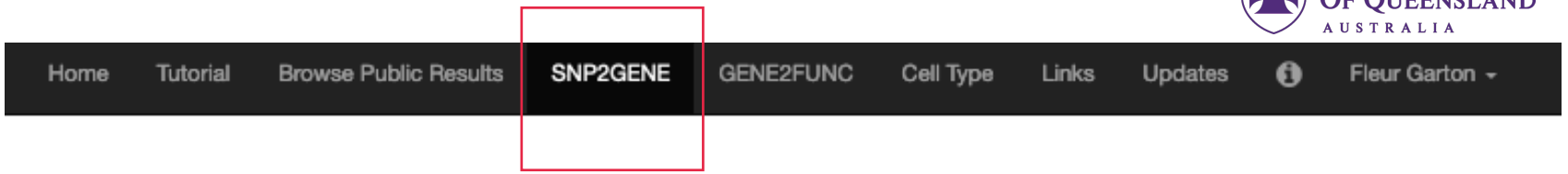
.gz

Gunzip – to unzip and check file headers

Gzip – to compress

```
imb20-015314-lt:Downloads uqfgarto$ head GCST90027163_buildGRCh37.tsv
rsid    chromosome      base_pair_location      effect_allele   other_allele    beta    standard_error  p_value N_effective
rs10900604      1       798400  a       g       -0.0198 0.0171  0.2471  47571
rs11240777      1       798959  a       g       0.0195  0.0171  0.252   47571
rs11240779      1       808631  a       g       -0.0181 0.0186  0.3292  47571
rs11240780      1       808928  t       c       -0.0189 0.0187  0.3113  47571
rs117086422     1       845635  t       c       0.0289  0.0178  0.1044  57846
```

PRACTICAL



1. Upload input files

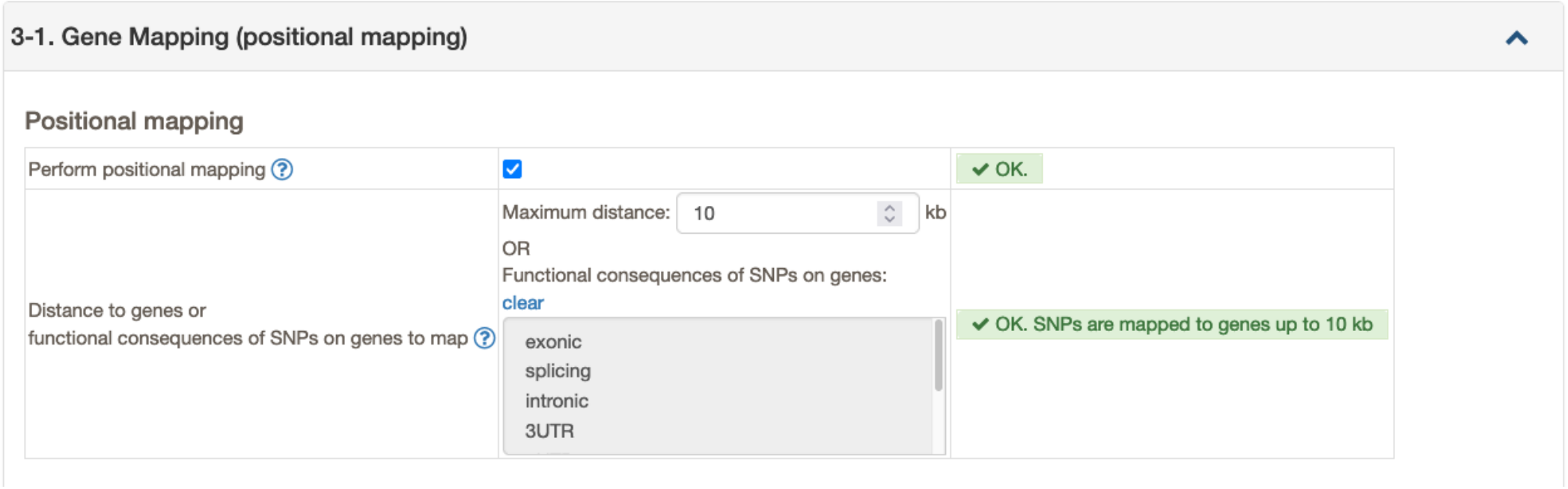
GWAS summary statistics	Browse... GCST90027163_buildGRCh37.tsv.gz Or ☐ : Use example input (Crohn's disease, Franke et al. 2010).	✔ OK. Please check your input file format.
GWAS summary statistics file columns	ℹ case insensitive Chromosome: chromosome Position: base_pair_location rsID: rsid P-value: p_value Effect allele*: effect_allele * "A1" is effect allele by default Non effect allele: other_allele OR: Beta: beta SE: standard_error	❗ Optional. Please fill as much as you can. It is not necessary to fill all column names.
Pre-defined lead SNPs	Browse... No file selected.	❗ Optional.
Identify additional independent lead SNPs	☑	❗ Optional. This is only valid when predefined lead SNPs are provided.
Predefined genomic region	Browse... No file selected.	❗ Optional.

PRACTICAL

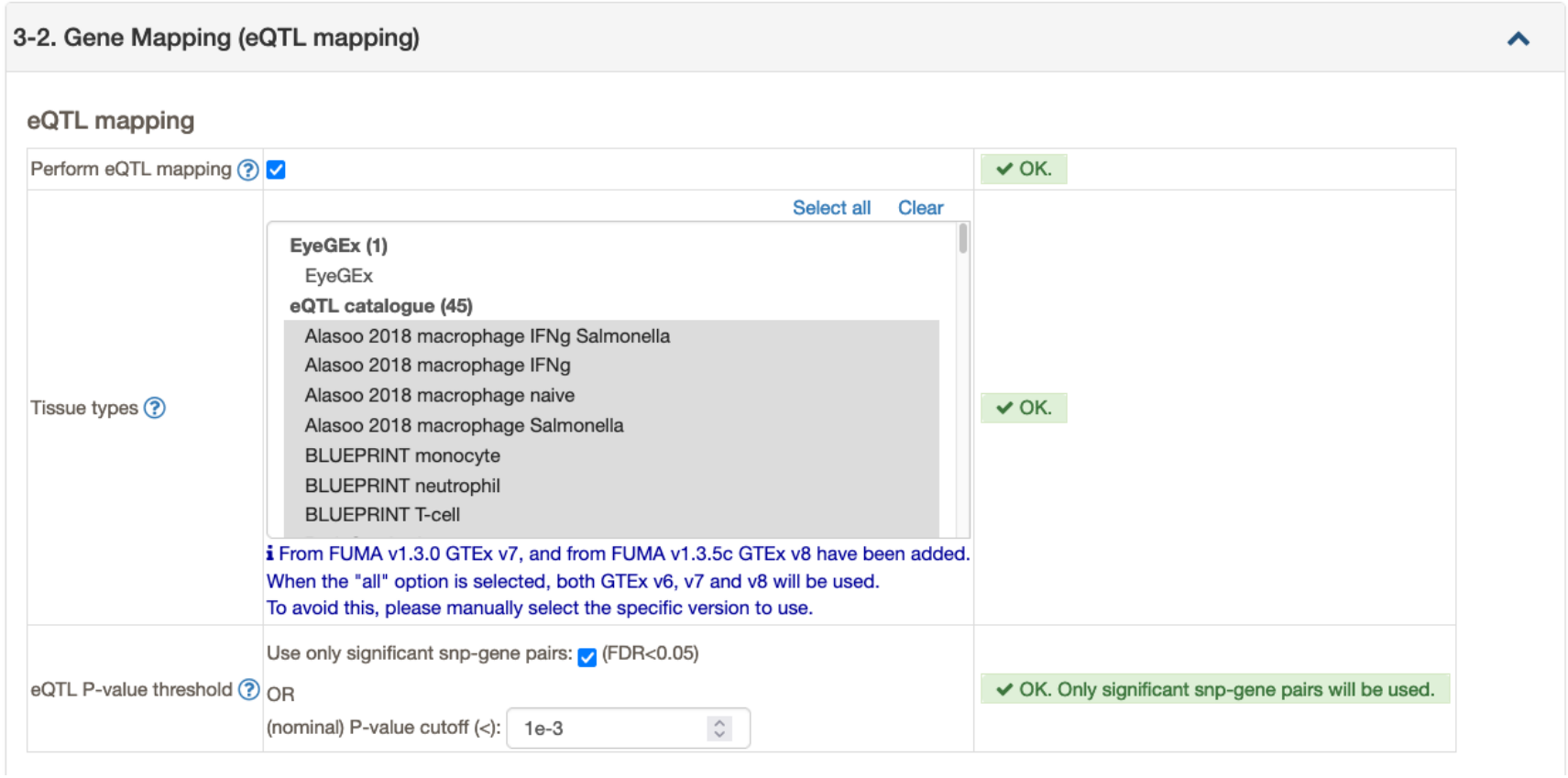
2. Parameters for lead SNPs and candidate SNPs identification

Sample size (N)	Total sample size (integer): OR Column name for N per SNP (text): N_effective	✔ OK. The defined column will be used for sample size per SNP.
Maximum P-value of lead SNPs (<)	5e-8	✔ OK.
Maximum P-value cutoff (<)	0.05	✔ OK.
r^2 threshold to define independent significant SNPs (≥)	0.6	✔ OK.
2nd r^2 threshold to define lead SNPs (≥)	0.1	✔ OK.
Reference panel population	1000G Phase3 ALL	✔ OK.
Include variants in reference panel (non-GWAS tagged SNPs in LD)	Yes	✔ OK.
Minimum Minor Allele Frequency (≥)	0	✔ OK.
Maximum distance between LD blocks to merge into a locus (< kb)	250 kb	

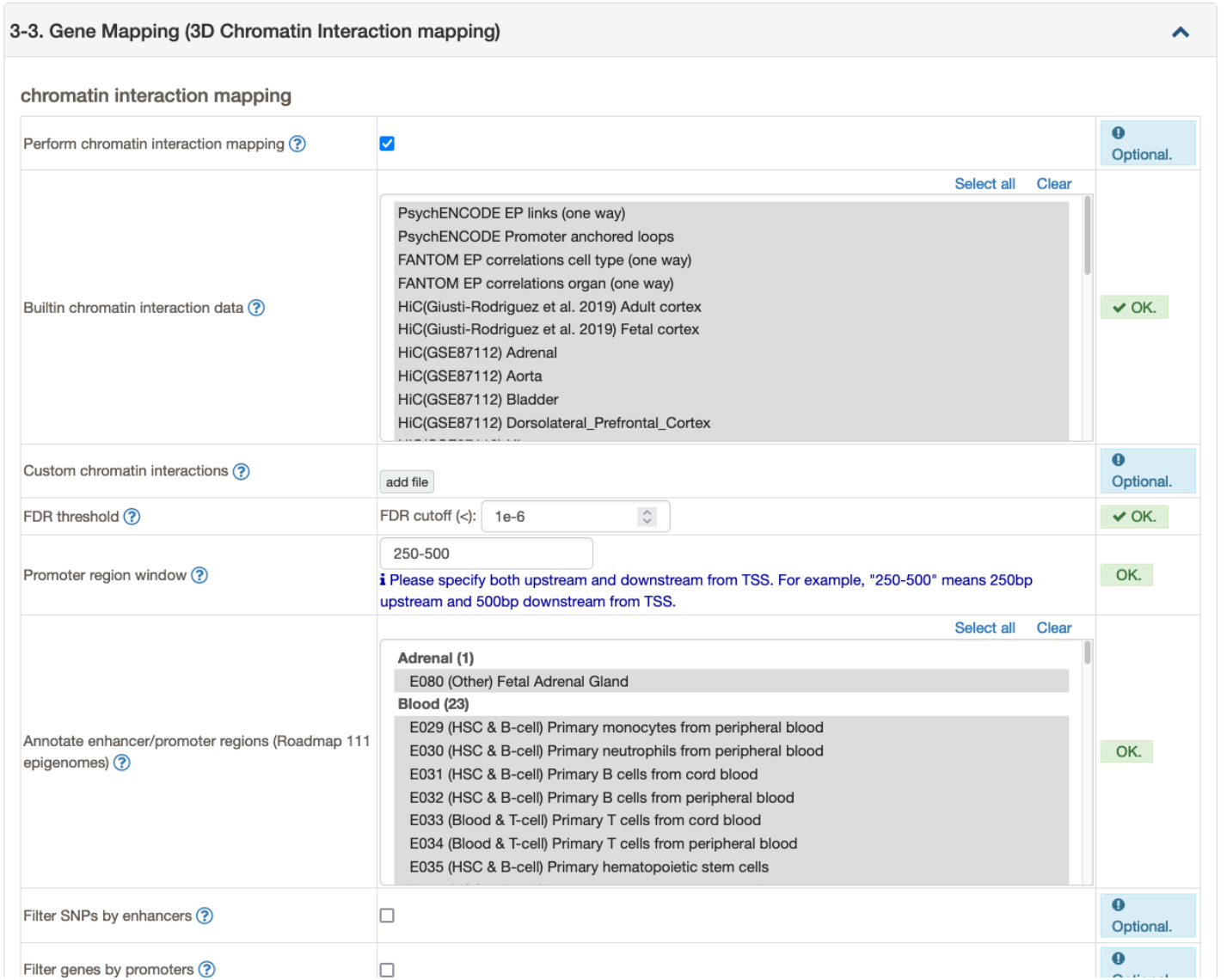
PRACTICAL


3-1. Gene Mapping (positional mapping)
Positional mapping
Perform positional mapping
OK.
Maximum distance: 10 kb
OR
Functional consequences of SNPs on genes:
clear
Distance to genes or
functional consequences of SNPs on genes to map
exonic
splicing
intronic
3UTR
OK. SNPs are mapped to genes up to 10 kb

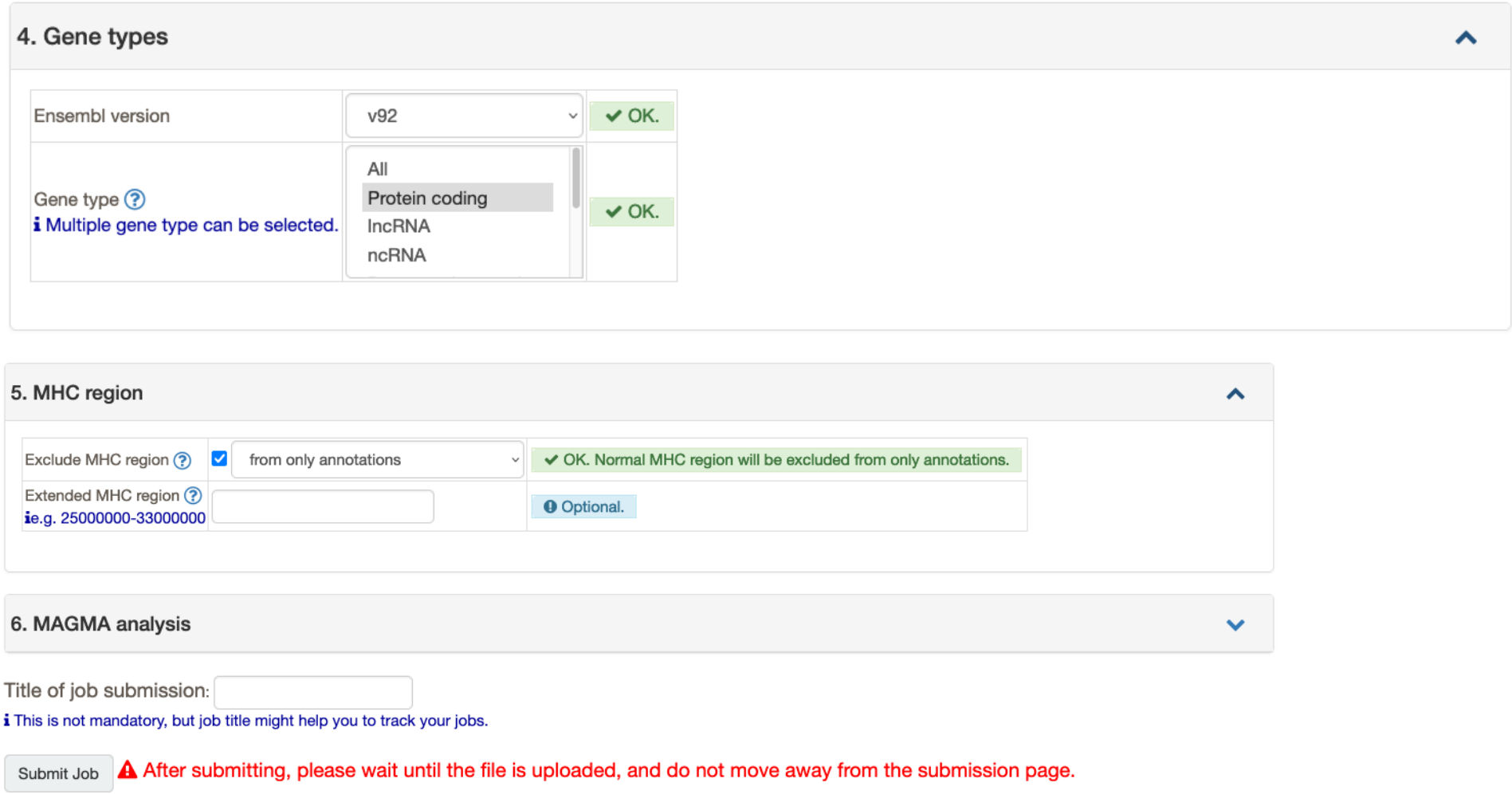
PRACTICAL


3-2. Gene Mapping (eQTL mapping)
eQTL mapping
Perform eQTL mapping
OK.
Select all
Clear
EyeGEx (1)
EyeGEx
eQTL catalogue (45)
Alasoo 2018 macrophage IFNg Salmonella
Alasoo 2018 macrophage IFNg
Alasoo 2018 macrophage naive
Alasoo 2018 macrophage Salmonella
BLUEPRINT monocyte
BLUEPRINT neutrophil
BLUEPRINT T-cell
Tissue types
OK.
From FUMA v1.3.0 GTEx v7, and from FUMA v1.3.5c GTEx v8 have been added.
When the "all" option is selected, both GTEx v6, v7 and v8 will be used.
To avoid this, please manually select the specific version to use.
Use only significant snp-gene pairs: (FDR<0.05)
OR
(nominal) P-value cutoff (<): 1e-3
eQTL P-value threshold
OK. Only significant snp-gene pairs will be used.

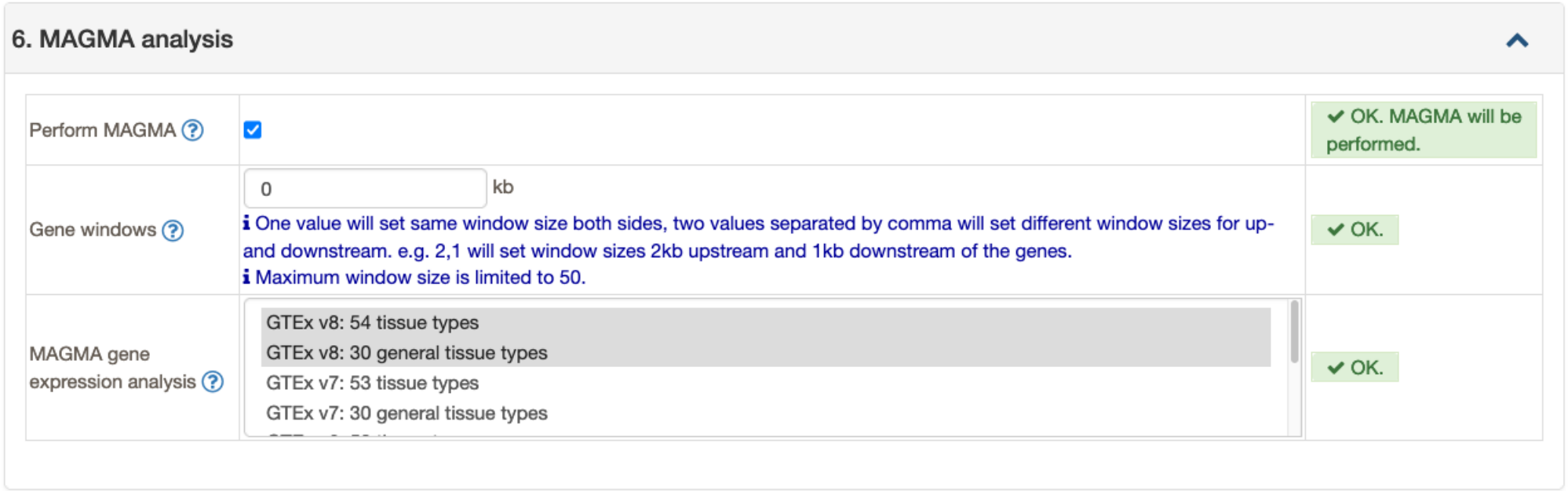
PRACTICAL


3-3. Gene Mapping (3D Chromatin Interaction mapping)
chromatin interaction mapping
Perform chromatin interaction mapping
Optional.
Select all
Clear
Builtin chromatin interaction data
PsychENCODE EP links (one way)
PsychENCODE Promoter anchored loops
FANTOM EP correlations cell type (one way)
FANTOM EP correlations organ (one way)
HiC(Giusti-Rodriguez et al. 2019) Adult cortex
HiC(Giusti-Rodriguez et al. 2019) Fetal cortex
HiC(GSE87112) Adrenal
HiC(GSE87112) Aorta
HiC(GSE87112) Bladder
HiC(GSE87112) Dorsolateral_Prefrontal_Cortex
OK.
Custom chromatin interactions
add file
Optional.
FDR threshold
FDR cutoff (<):
1e-6
OK.
Promoter region window
250-500
Please specify both upstream and downstream from TSS. For example, "250-500" means 250bp upstream and 500bp downstream from TSS.
OK.
Select all
Clear
Annotate enhancer/promoter regions (Roadmap 111 epigenomes)
Adrenal (1)
E080 (Other) Fetal Adrenal Gland
Blood (23)
E029 (HSC & B-cell) Primary monocytes from peripheral blood
E030 (HSC & B-cell) Primary neutrophils from peripheral blood
E031 (HSC & B-cell) Primary B cells from cord blood
E032 (HSC & B-cell) Primary B cells from peripheral blood
E033 (Blood & T-cell) Primary T cells from cord blood
E034 (Blood & T-cell) Primary T cells from peripheral blood
E035 (HSC & B-cell) Primary hematopoietic stem cells
OK.
Filter SNPs by enhancers
Optional.
Filter genes by promoters

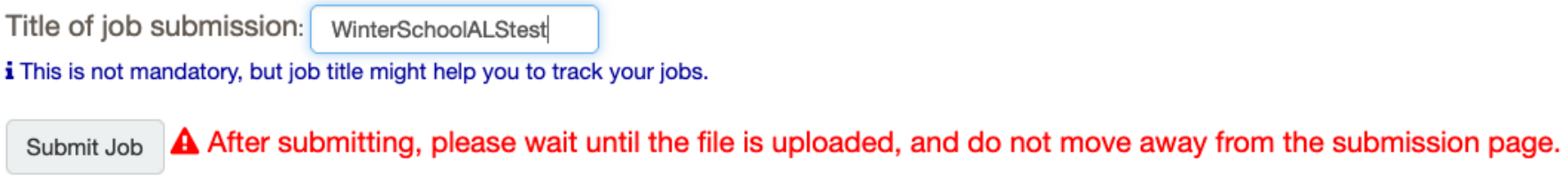
PRACTICAL


4. Gene types
Ensembl version
v92
OK.
Gene type
Multiple gene type can be selected.
All
Protein coding
lncRNA
ncRNA
OK.
5. MHC region
Exclude MHC region
from only annotations
OK. Normal MHC region will be excluded from only annotations.
Extended MHC region
e.g. 25000000-33000000
Optional.
6. MAGMA analysis
Title of job submission:
This is not mandatory, but job title might help you to track your jobs.
Submit Job
After submitting, please wait until the file is uploaded, and do not move away from the submission page.

PRACTICAL


6. MAGMA analysis
Perform MAGMA
OK. MAGMA will be performed.
Gene windows
0
kb
One value will set same window size both sides, two values separated by comma will set different window sizes for up- and downstream. e.g. 2,1 will set window sizes 2kb upstream and 1kb downstream of the genes.
Maximum window size is limited to 50.
OK.
MAGMA gene expression analysis
GTEx v8: 54 tissue types
GTEx v8: 30 general tissue types
GTEx v7: 53 tissue types
GTEx v7: 30 general tissue types
OK.

PRACTICAL


Title of job submission:
WinterSchoolALStest
This is not mandatory, but job title might help you to track your jobs.
Submit Job
After submitting, please wait until the file is uploaded, and do not move away from the submission page.

PRACTICAL

My Jobs

List of Jobs

Delete selected jobs

Job ID	Job name	Submit date	Status	Jump to GENE2FUNC	Publish	Select
191056	WinterSchoolALStest	2022-06-17 03:33:00	NEW	Not available	Not available	☐

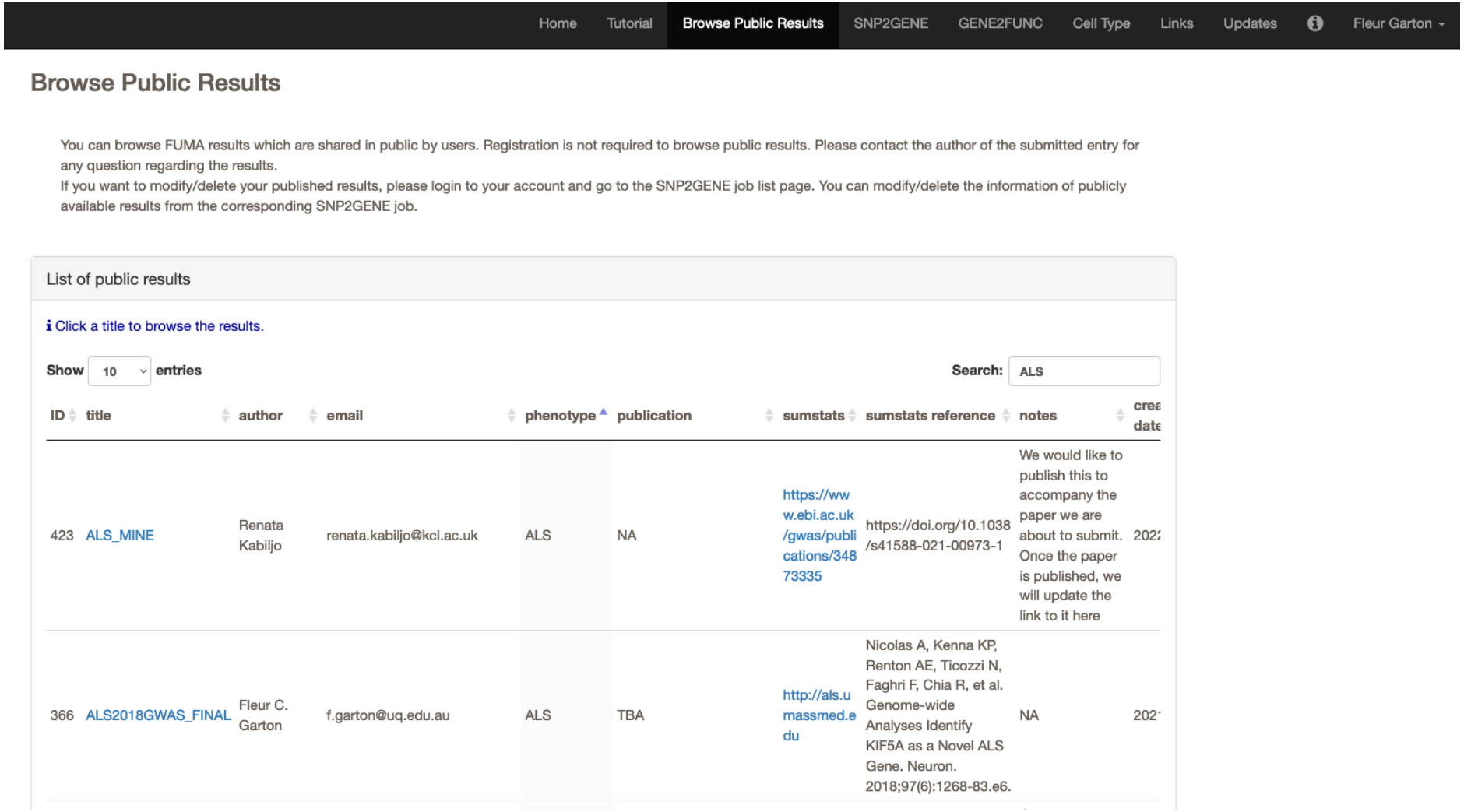
My Jobs

List of Jobs

Delete selected jobs

Job ID	Job name	Submit date	Status	Jump to GENE2FUNC	Publish	Select
191056	WinterSchoolALStest	2022-06-17 03:33:00	QUEUED	Not available	Not available	☐

PRACTICAL


Home
Tutorial
Browse Public Results
SNP2GENE
GENE2FUNC
Cell Type
Links
Updates
Fleur Garton
Browse Public Results
You can browse FUMA results which are shared in public by users. Registration is not required to browse public results. Please contact the author of the submitted entry for any question regarding the results.
If you want to modify/delete your published results, please login to your account and go to the SNP2GENE job list page. You can modify/delete the information of publicly available results from the corresponding SNP2GENE job.
List of public results
Click a title to browse the results.
Show 10 entries
Search: ALS
ID title author email phenotype publication sumstats sumstats reference notes crea date
423 ALS_MINE Renata Kabiljo renata.kabiljo@kcl.ac.uk ALS NA https://www.ebi.ac.uk/gwas/publications/34873335 https://doi.org/10.1038/s41588-021-00973-1 We would like to publish this to accompany the paper we are about to submit. Once the paper is published, we will update the link to it here 202
366 ALS2018GWAS_FINAL Fleur C. Garton f.garton@uq.edu.au ALS TBA http://als.umassmed.edu Nicolas A, Kenna KP, Renton AE, Ticozzi N, Faghri F, Chia R, et al. Genome-wide Analyses Identify KIF5A as a Novel ALS Gene. Neuron. 2018;97(6):1268-83.e6. NA 202

FUMAGWAS

GWAS list

Example input page

New Job

SNP2GENE

Genome-wide plots

Summary of results

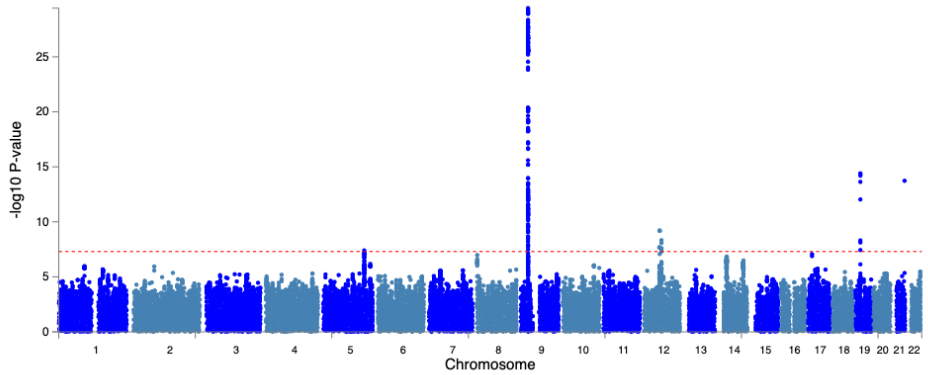
Results

Download

Manhattan Plot (GWAS summary statistics)

ℹ Manhattan plot of the input GWAS summary statistics.
For plotting, overlapping data points are not drawn (filtering was performed only for SNPs with P-value ≥ 1e-5, see tutorial for details).

Download the plot as PNG JPG SVG PDF



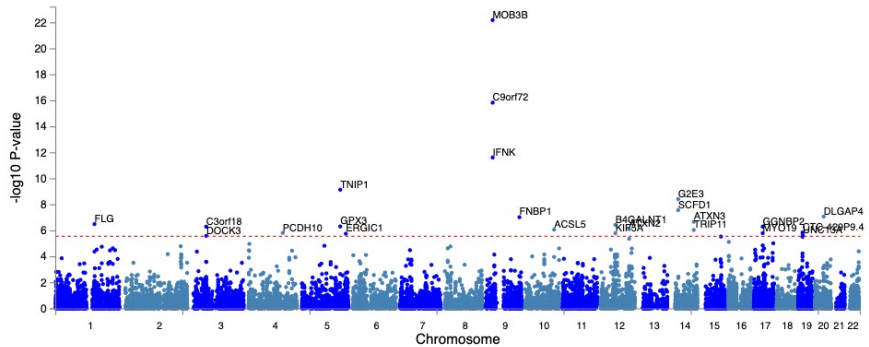
Mahattan Plot (gene-based test)

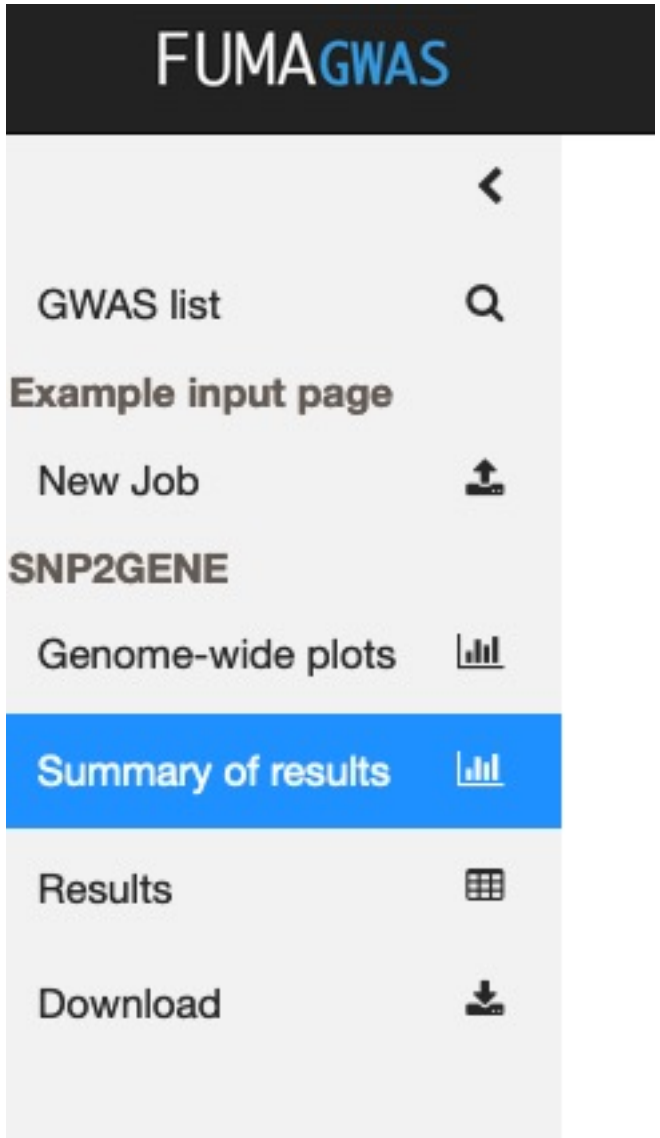
ℹ This is a manhattan plot of the gene-based test as computed by MAGMA based on your input GWAS summary statistics.
The gene-based P-value is downloadable from 'Download' tab from the left side bar.

Input SNPs were mapped to 19071 protein coding genes. Genome wide significance (red dashed line in the plot) was defined at P = 0.05/19071 = 2.622e-6.

Download the plot as PNG JPG SVG PDF

Label top 24 genes.





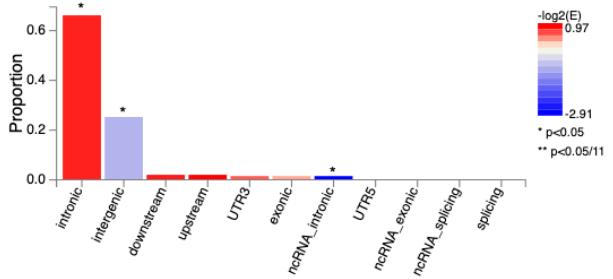
Summary of SNPs and mapped genes

#Genomic risk loci	6
#lead SNPs	16
#Ind. Sig. SNPs	43
#candidate SNPs	201
#candidate GWAS tagged SNPs	195
#mapped genes	92

Functional consequences of SNPs on genes

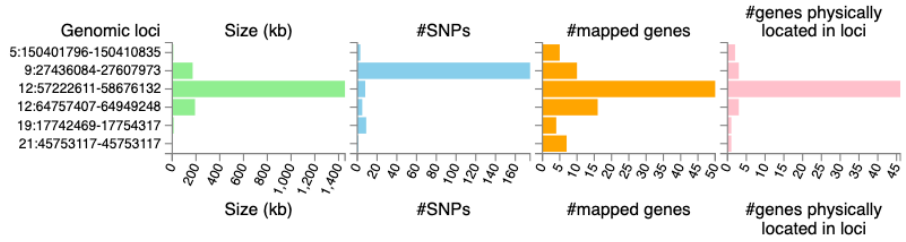
Statistics are available in "annov.stats.txt". The file is downloadable from the "Download" tab.

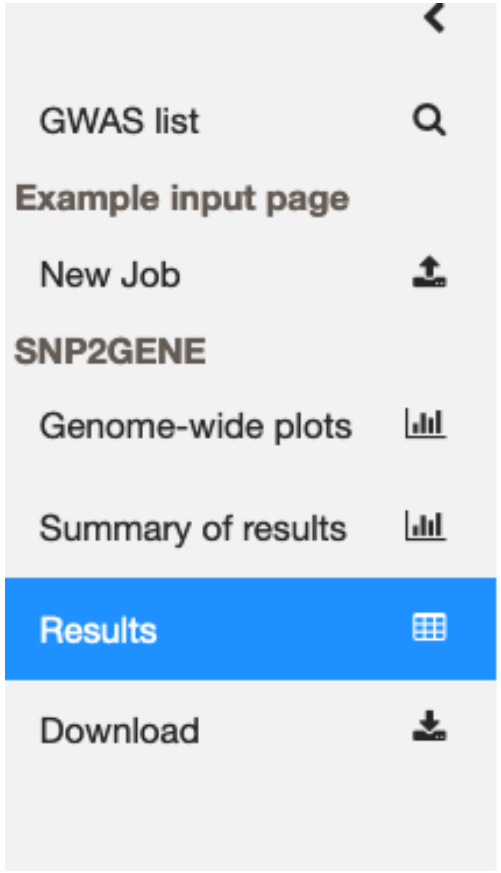
Download the plot as PNG JPG SVG PDF



Summary per genomic risk locus

Download the plot as PNG JPG SVG PDF





Result tables

Genomic risk loci | lead SNPs | Ind. Sig. SNPs | SNPs (annotations) | ANNOVAR | Mapped Genes | eQTL | Chromatin interactions | GWAScatalog | Parameters

ℹ Click row to display a regional plot of GWAS summary statistics.

Show 10 entries Search:

Genomic Locus	uniqID	rsID	chr	pos	P-value	start	end	nSNPs	nGWASSNPs	nIndSigSNPs	IndSigSNP
1	5:150410835:C:T	rs10463311	5	150410835	3.999e-08	150401796	150410835	3	3	1	rs1046331
2	9:27543382:C:T	rs3849943	9	27543382	3.77e-30	27436084	27607973	175	169	32	rs1472118
3	12:58676132:C:G	rs142321490	12	58676132	6.147e-10	57222611	58676132	8	8	3	rs1180825
4	12:64881967:A:G	rs74654358	12	64881967	4.658e-09	64757407	64949248	5	5	2	rs7465435
5	19:17753239:C:G	rs12973192	19	17753239	3.916e-15	17742469	17754317	9	9	4	rs1297319
6	21:45753117:A:C	rs75087725	21	45753117	1.848e-14	45753117	45753117	1	1	1	rs7508772

Showing 1 to 6 of 6 entries

Previous 1 Next

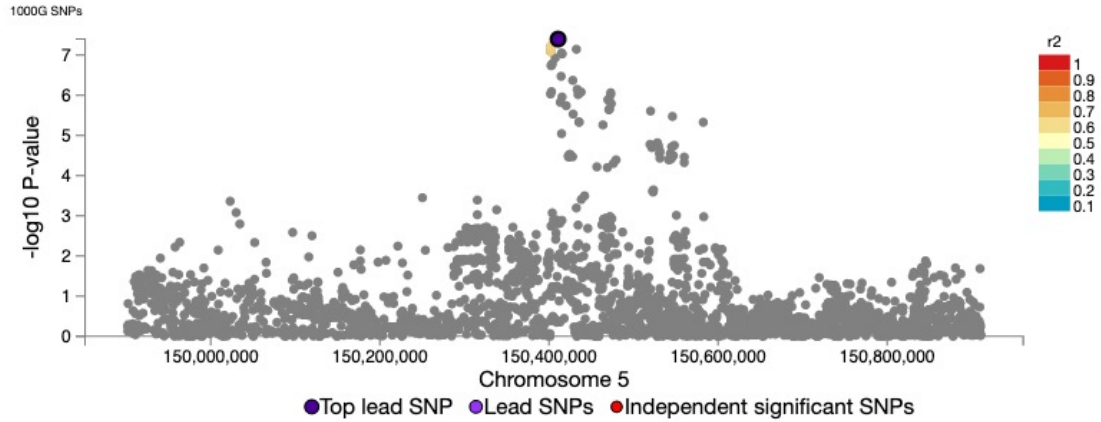
Regional Plot (GWAS association)

ℹ Please click one of the row of 'Genomic risk loci', 'lead SNPs' or 'ind. sig. SNPs' tables to display a regional plot.
You can zoom in/out by mouse scroll.
Each SNP is color-coded based on the highest r^2 to one of the ind. sig. SNPs, if that is greater or equal to the user defined threshold. Other SNPs (i.e. below the user-defined r^2) are colored in grey. The top lead SNPs in genomic risk loci, lead SNPs and ind. sig. SNPs are circled in black and colored in dark-purple, purple and red, respectively.

Regional Plot (GWAS association)

ℹ Please click one of the row of 'Genomic risk loci', 'lead SNPs' or 'ind. sig. SNPs' tables to display a regional plot.
You can zoom in/out by mouse scroll.
Each SNP is color-coded based on the highest r^2 to one of the ind. sig. SNPs, if that is greater or equal to the user defined threshold. Other SNPs (i.e. below the user-defined r^2) are colored in grey. The top lead SNPs in genomic risk loci, lead SNPs and ind. sig. SNPs are circled in black and colored in dark-purple, purple and red, respectively.

Clear



Selected Locus

top lead SNP	rs10463311
Chrom	5
BP	150410835
P-value	3.999e-08
#Ind. Sig. SNPs	1
#lead SNPs	1
SNPs within LD	3
GWAS SNPs within LD	3

Regional plot with annotation

Select annotation(s) to plot:

- [x] GWAS association statistics
- [x] CADD score
- [x] RegulomeDB score
- [ ] Chromatine 15 state
- [x] eQTL
- [x] Chromatin interaction

Plot

OK. Good to go. Click "Plot" to create regional plot with selected annotations.

PRACTICAL

Detail of output options

Regulome DB

RegulomeDB Categorical Scores

Category	Description
	Likely to affect binding and linked to expression of a gene target
1a	eQTL + TF binding + matched TF motif + matched DNase footprint + DNase peak
1b	eQTL + TF binding + any motif + DNase footprint + DNase peak
1c	eQTL + TF binding + matched TF motif + DNase peak
1d	eQTL + TF binding + any motif + DNase peak
1e	eQTL + TF binding + matched TF motif
1f	eQTL + TF binding/DNase peak
	Likely to affect binding
2a	TF binding + matched TF motif + matched DNase footprint + DNase peak
2b	TF binding + any motif + DNase footprint + DNase peak
2c	TF binding + matched TF motif + DNase peak
	Less likely to affect binding
3a	TF binding + any motif + DNase peak
3b	TF binding + matched TF motif
	Minimal binding evidence
4	TF binding + DNase peak
5	TF binding or DNase peak
6	Motif hit
	No binding evidence
7	No evidence
NA	the variant does not exist in RegulomeDB

*External link to RegulomeDB from SNP table (when one of the SNPs is clicked) will open a new tab. rsID does not always match since RegulomeDB used dbSNP build 141 (the rsID in FUMA is dbSNP build 146). Genomic position (bp on hg19) shown in the link of RegulomeDB is the position shown in the SNP table - 1, since RegulomeDB used 0 based coordinate.

15-core chromatin state

*When 15-core chromatin state is included in the plot and >30 cell types are selected, the labels of Y-axis are omitted. The order of the cell types is same as the legend table.

eQTLs

The color of eQTLs are arbitrary. When P-value is not available (i.e. for CMC eQTLs), -log10 FDR is plotted in stead of P-value.

SNPs colored grey in the plots

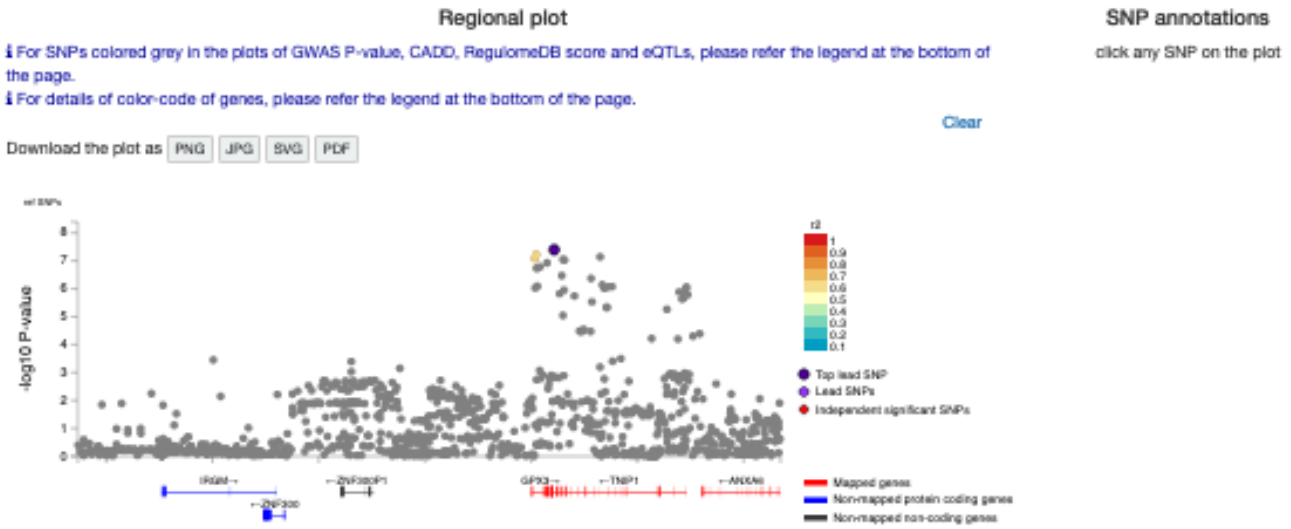
GWAS P-value: SNPs which are not in LD of any of significant independent lead SNPs in the selected region are colored grey.
CADD score: Only SNPs which are in LD of any of significant independent lead SNPs are displayed in the plot. Of those SNPs, SNPs which did not used for mapping (SNPs that were filtered by user defined parameters) are colored grey.
When positional mapping is performed, SNPs used for positional mapping are always colored non-grey colors.
When eQTL mapping is performed and eQTLs are plotted, SNPs used for eQTL mapping are also colored non-grey colors. If the option of eQTLs is not selected for the plot, SNPs which are not used for other mappings are colored grey even if they are used for eQTL mapping.
When chromatin interaction mapping is performed and chromatin interactions are plotted, SNPs used for chromatin interaction mapping are also colored non-grey colors. If the option of chromatin interactions is not selected for the plot, SNPs which are not used for other mappings are colored grey even if they are used for chromatin interaction mapping.
RegulomeDB score: Same as CADD score.
eQTLs: When eQTL mapping was performed and if there is any eQTL in the selected region, all eQTLs with user defined P-value threshold and tissue types are displayed. Of those eQTLs, eQTLs which did not used for eQTL mapping (eQTLs that were filtered by user defined parameters) are colored grey.

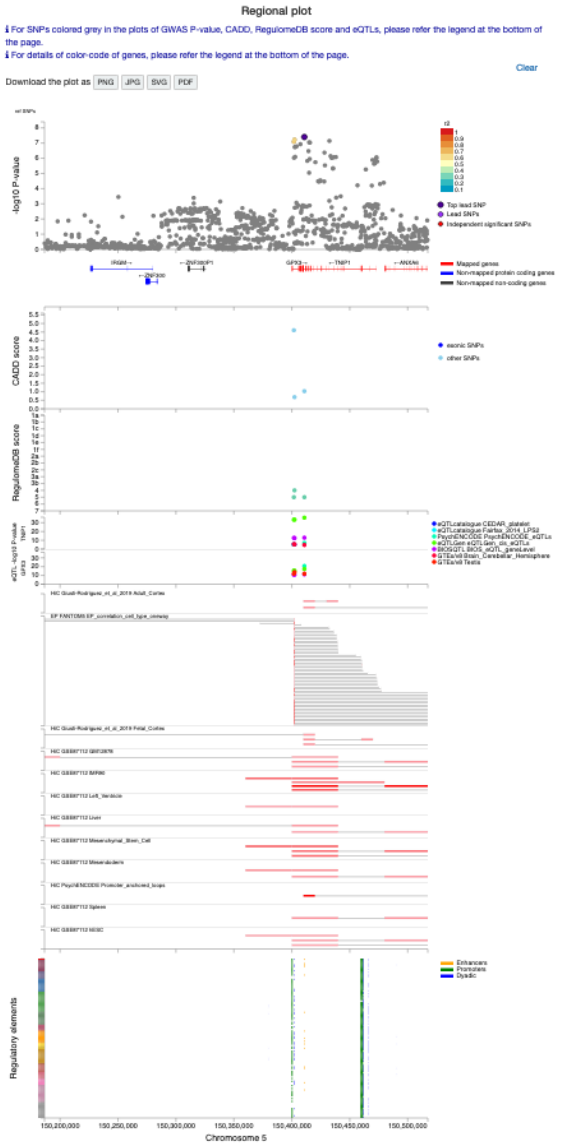
Color-code for genes

Red : Mapped genes. Genes mapped by positional mapping are always colored red. Genes mapped by eQTL mapping are colore red only when the option of eQTLs is selected for the plot, otherwise those genes are considered as non-mapped genes. Genes mapped by chromatin interaction are colored red only when the option of chromatin interactions is selected for the plot, otherwise those genes are considered as non-mapped genes.
Blue : Non-mapped protein-coding genes.
Dark grey : Non-mapped non-coding genes.

PRACTICAL

Regional plotting – locus specific





PRACTICAL

Result tables

Genomic risk loci | lead SNPs | Ind. Sig. SNPs | SNPs (annotations) | ANNOVAR | Mapped Genes | eQTL | Chromatin interactions | GWAScatalog | Parameters

PRACTICAL

•Manhattan plot: The most outer layer. Only SNPs with P < 0.05 are displayed. SNPs in genomic risk loci are color-coded as a function of their maximum r^2 to the one of the independent significant SNPs in the locus, as follows: red ($r^2 > 0.8$), orange ($r^2 > 0.6$), green ($r^2 > 0.4$) and blue ($r^2 > 0.2$). SNPs that are not in LD with any of the independent significat SNPs (with $r^2 \leq 0.2$) are grey.The rsID of the top SNPs in each risk locus are displayed in the most outer layer. Y-axis are raned between 0 to the maximum -log10(P-value) of the SNPs.

•Chromosome ring: The second layer. Genomic risk loci are highlighted in blue.

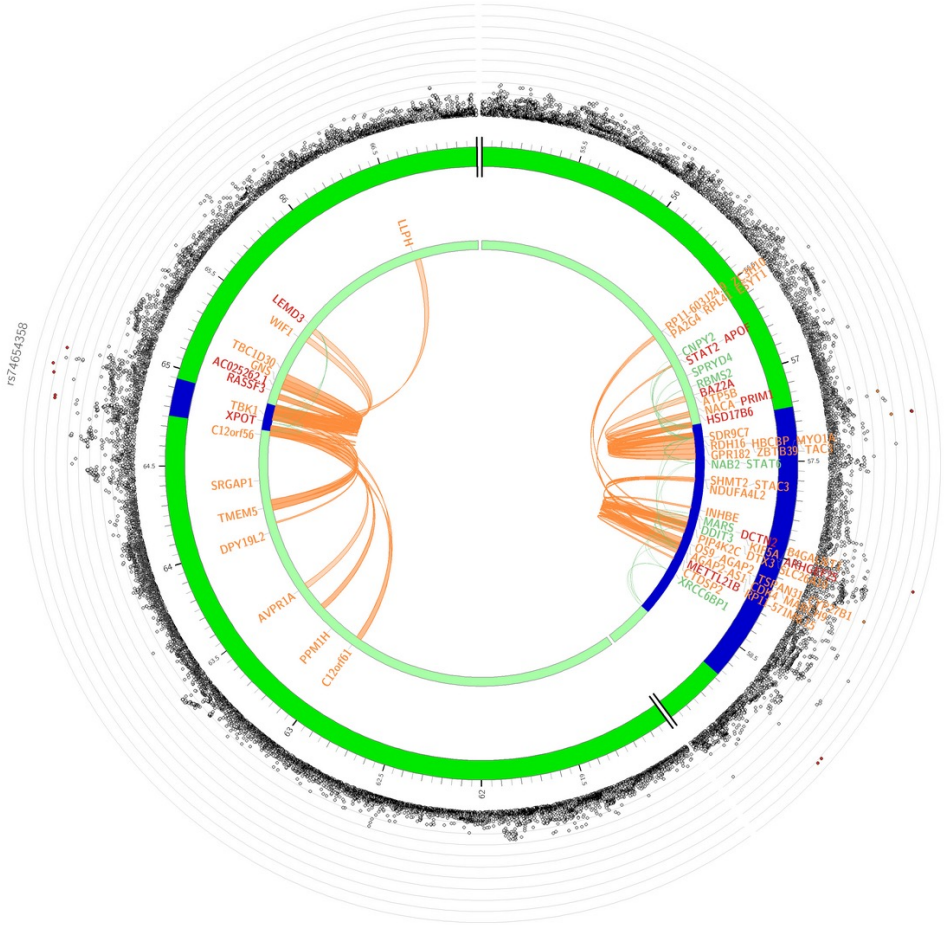
•Mapped genes -
•If the gene is mapped only by chromatin interactions or only by eQTLs, it is colored orange or green, respectively.
•When the gene is mapped by both, it is colored red.

Since creating a circos plot might take long time with a large number of points and links, the maximum number of points and links are limited to 50,000 and 10,000 per plot (chromosome), respectively, in the default plot.

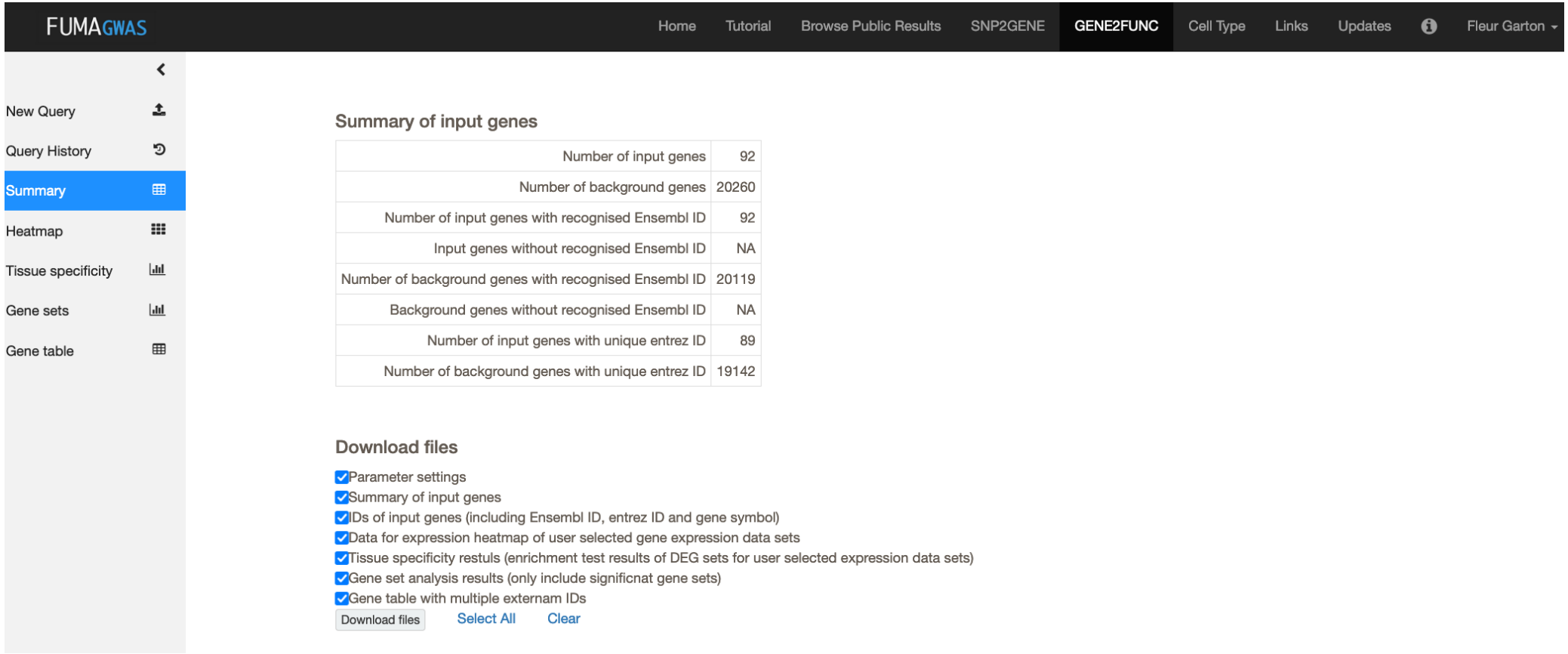
Therefore, if there are more than 50,000 SNPs with P-value < 0.05 in a chromosome, top 50,000 SNPs (sorted by P-value) are displayed in the plot.

This is same for eQTLs and chromatin interactions, e.g. if there are more than 10,000 eQTLs in a chromosome, top 10,000 eQTLs (sorted by P-value for eQTLs, FDR for chromatin interactions) are displayed in the plot. These can be optimized by downloading config file and re-creating input text files for SNPs and links.

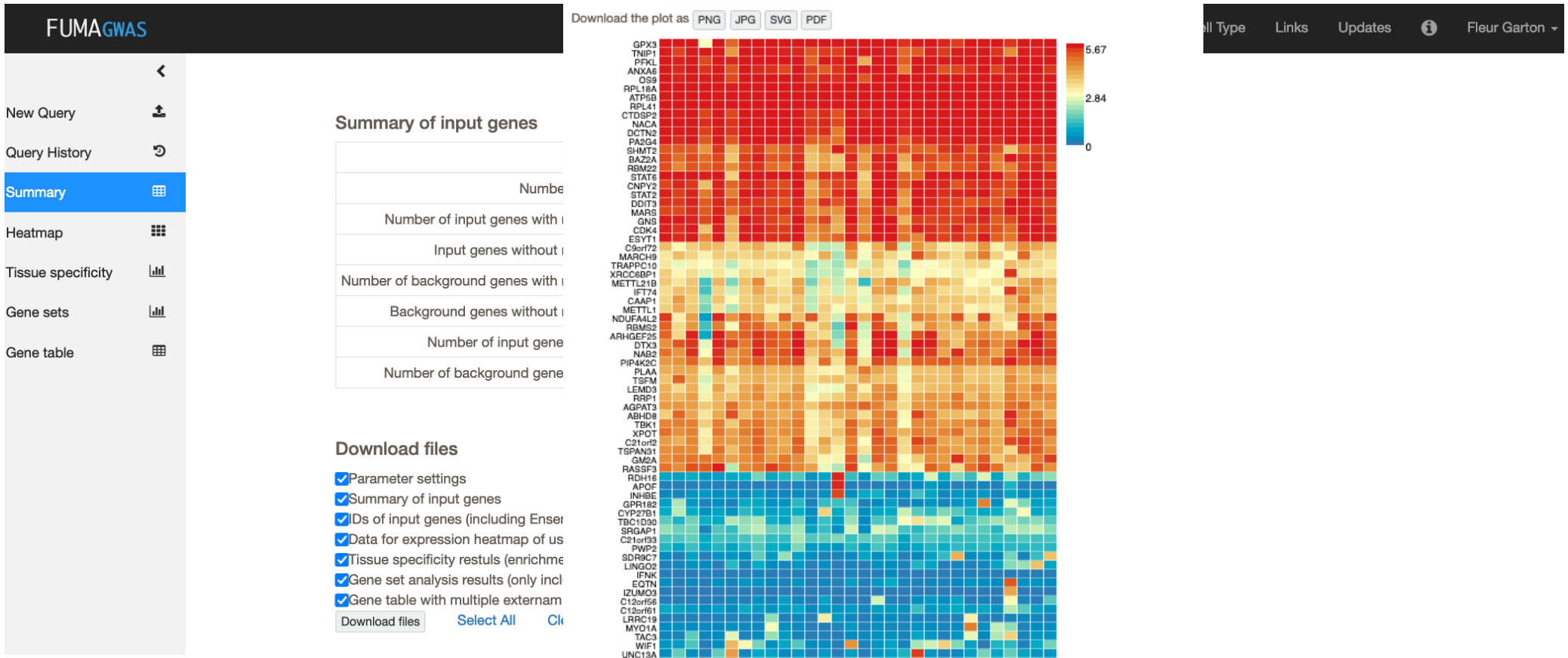
Please refer github repository FUMA circos plot for details.

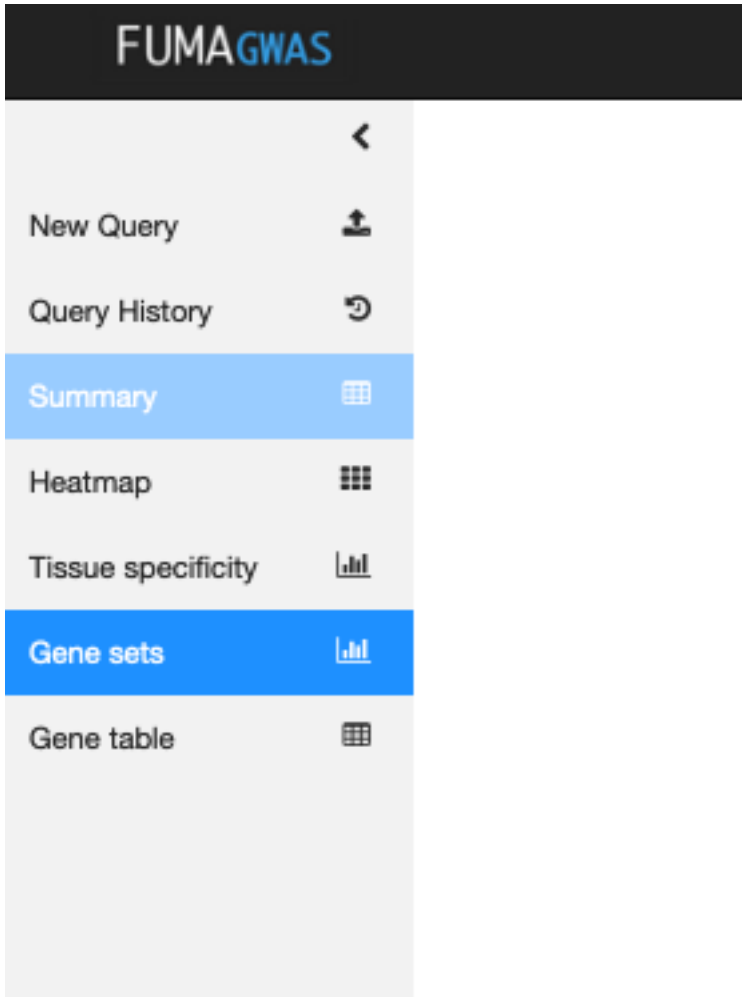


PRACTICAL


FUMAGWAS
Home
Tutorial
Browse Public Results
SNP2GENE
GENE2FUNC
Cell Type
Links
Updates
Fleur Garton
New Query
Query History
Summary
Heatmap
Tissue specificity
Gene sets
Gene table
Summary of input genes
Number of input genes 92
Number of background genes 20260
Number of input genes with recognised Ensembl ID 92
Input genes without recognised Ensembl ID NA
Number of background genes with recognised Ensembl ID 20119
Background genes without recognised Ensembl ID NA
Number of input genes with unique entrez ID 89
Number of background genes with unique entrez ID 19142
Download files
Parameter settings
Summary of input genes
IDs of input genes (including Ensembl ID, entrez ID and gene symbol)
Data for expression heatmap of user selected gene expression data sets
Tissue specificity restuls (enrichment test results of DEG sets for user selected expression data sets)
Gene set analysis results (only include significnat gene sets)
Gene table with multiple externam IDs
Download files
Select All
Clear

PRACTICAL


FUMAGWAS
Type
Links
Updates
Fleur Garton
New Query
Query History
Summary
Heatmap
Tissue specificity
Gene sets
Gene table
Summary of input genes
Numbe
Number of input genes with
Input genes without
Number of background genes with
Background genes without
Number of input gene
Number of background gene
Download files
Parameter settings
Summary of input genes
IDs of input genes (including Ense
Data for expression heatmap of us
Tissue specificity restuls (enrichme
Gene set analysis results (only incl
Gene table with multiple externam
Download files
Select All
Gene expression heatmap
Data set: GTEx v8 30 general tissue types
Expression Value: Average expression per label (log2 transformed)
Order genes by: Cluster
Order tissues by: Alphabetical order
Download the plot as PNG JPG SVG PDF
5.67
2.84
0
Adipose_Tissue
Adrenal_Gland
Bladder
Blood
Blood_Vessel
Brain
Breast
Cervix_Uteri
Colon
Esophagus
Fallopian_Tube
Heart
Kidney
Liver
Lung
Muscle
Nerve
Ovary
Pancreas
Pituitary
Prostate
Salivary_Gland
Skin
Small_Intestine
Spleen
Stomach
Testis
Thyroid
Uterus
Vagina

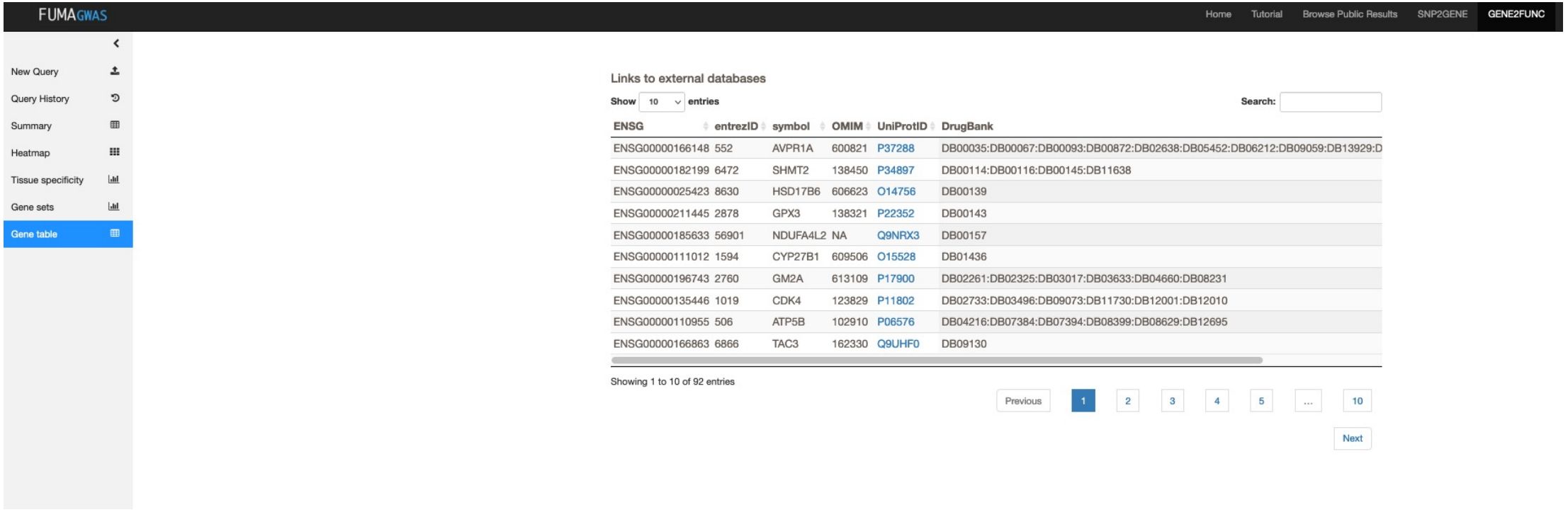


Enrichment of input genes in Gene Sets

ℹ Plots and tables only display gene sets with adjusted P-value < 0.05. When adjusted P-value threshold is set to > 0.05, all results passed threshold are included in the GS.txt field downloadable from "Summary" tab.
ℹ If there is no significant gene sets (adjusted P-value < 0.05) in user provided custom gene sets, they are not displayed in this page, but all results passed threshold are included in the GS.txt field downloadable from "Summary" tab.

Hallmark gene sets (MsigDB h)	(0)
Positional gene sets (MsigDB c1)	(5)
Curated_gene_sets	(2)
Chemical and Genetic pertubation gene sets (MsigDB c2)	(2)
All Canonical Pathways (MsigDB c2)	(0)
BioCarta (MsigDB c2)	(0)
KEGG (MsigDB c2)	(0)
Reactome (MsigDB c2)	(0)
microRNA targets (MsigDB c3)	(0)
TF targets (MsigDB c3)	(0)
All computational gene sets (MsigDB c4)	(0)
Cancer gene neighborhoods (MsigDB c4)	(0)
Cancer gene modules (MsigDB c4)	(0)
GO biological processes (MsigDB c5)	(0)
GO cellular components (MsigDB c5)	(0)
GO molecular functions (MsigDB c5)	(0)
Oncogenic signatures (MsigDB c6)	(0)
Immunologic signatures (MsigDB c7)	(0)
WikiPathways	(0)
GWAS catalog reported genes	(5)

PRACTICAL


FUMAGWAS
Home
Tutorial
Browse Public Results
SNP2GENE
GENE2FUNC
New Query
Query History
Summary
Heatmap
Tissue specificity
Gene sets
Gene table
Links to external databases
Show 10 entries
Search:
ENSG entrezID symbol OMIM UniProtID DrugBank
ENSG00000166148 552 AVPR1A 600821 P37288 DB00035:DB00067:DB00093:DB00872:DB02638:DB05452:DB06212:DB09059:DB13929:D
ENSG00000182199 6472 SHMT2 138450 P34897 DB00114:DB00116:DB00145:DB11638
ENSG00000025423 8630 HSD17B6 606623 O14756 DB00139
ENSG00000211445 2878 GPX3 138321 P22352 DB00143
ENSG00000185633 56901 NDUFA4L2 NA Q9NRX3 DB00157
ENSG00000111012 1594 CYP27B1 609506 O15528 DB01436
ENSG00000196743 2760 GM2A 613109 P17900 DB02261:DB02325:DB03017:DB03633:DB04660:DB08231
ENSG00000135446 1019 CDK4 123829 P11802 DB02733:DB03496:DB09073:DB11730:DB12001:DB12010
ENSG00000110955 506 ATP5B 102910 P06576 DB04216:DB07384:DB07394:DB08399:DB08629:DB12695
ENSG00000166863 6866 TAC3 162330 Q9UHF0 DB09130
Showing 1 to 10 of 92 entries
Previous
1
2
3
4
5
...
10
Next

Reading list/references:

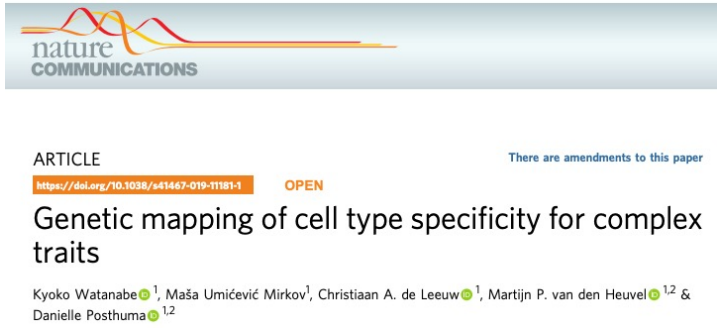

nature COMMUNICATIONS

ARTICLE

DOI: 10.1038/s41467-017-01261-5 OPEN

Functional mapping and annotation of genetic associations with FUMA

Kyoko Watanabe[1], Erdogan Taskesen[1,2], Arjen van Bochoven[3] & Danielle Posthuma[1,4]


nature COMMUNICATIONS

ARTICLE

There are amendments to this paper

https://doi.org/10.1038/s41467-019-11181-1 OPEN

Genetic mapping of cell type specificity for complex traits

Kyoko Watanabe[1], Maša Umićević Mirkov[1], Christiaan A. de Leeuw[1], Martijn P. van den Heuvel[1,2] & Danielle Posthuma[1,2]

Source code -

- https://github.com/Kyoko-wtnb/FUMA-webapp

SOURCE DATA PUBLICATIONS