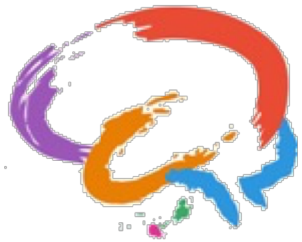
Amsterdam Neuroscience

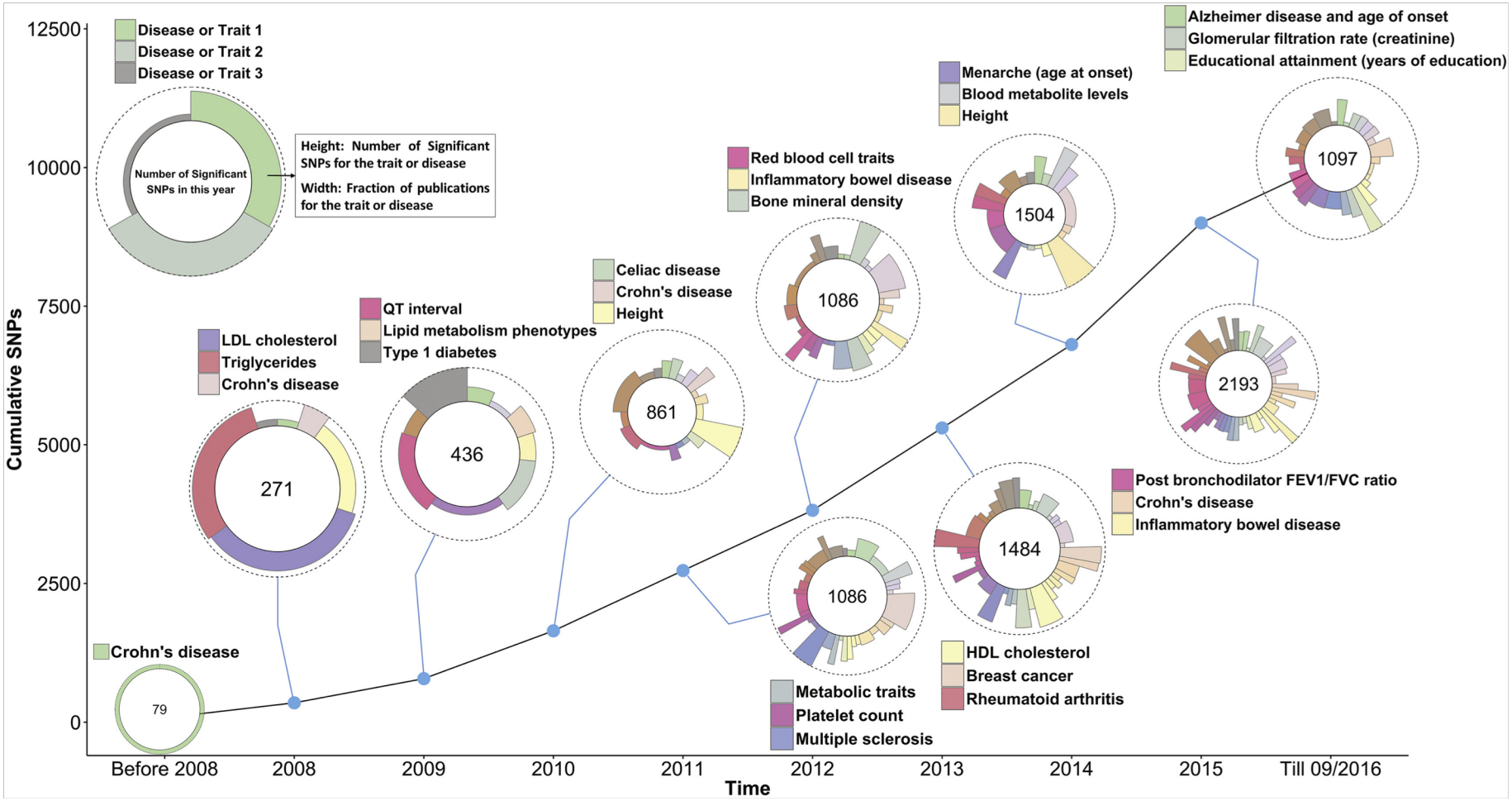
Functional mapping and annotation of genetic associations with FUMA

Kyoko Watanabe & Danielle Posthuma
8th March 2019

Center for Neurogenomics and Cognitive Research – CNCR
Vrije Universiteit & VU Medical Center, Amsterdam, The Netherlands



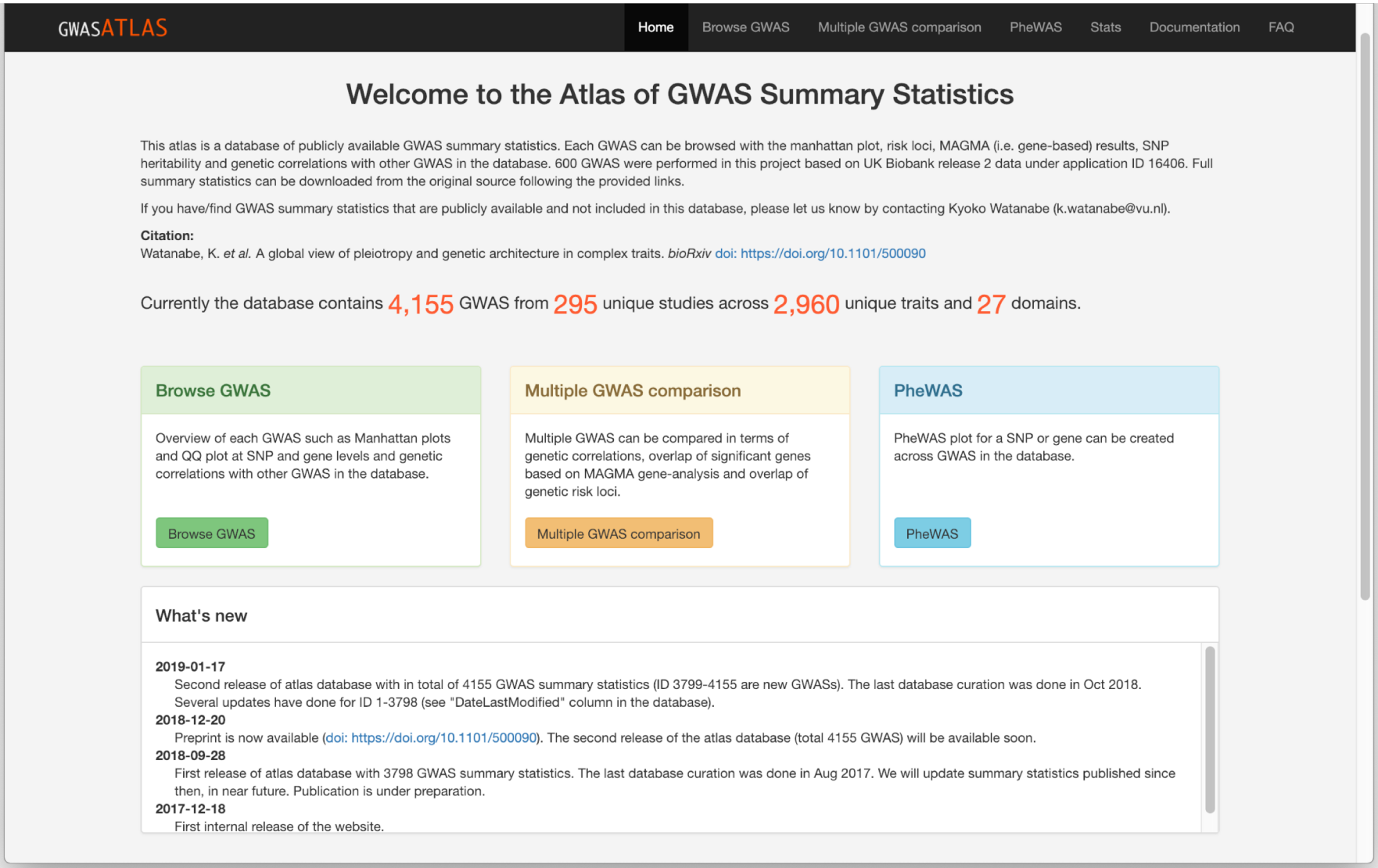
Genome-wide association studies (GWAS)



Visscher et al. 2017. AJHG

GWAS ATLAS

http://atlas.ctglab.nl



CNCR CTGLAB

GWAS ATLAS

Summary of results for a single GWAS

GWASATLAS | Home | Browse GWAS | Multiple GWAS comparison | PheWAS | Stats | Documentation | FAQ

atlas ID: 3785 Intelligence

GWAS information

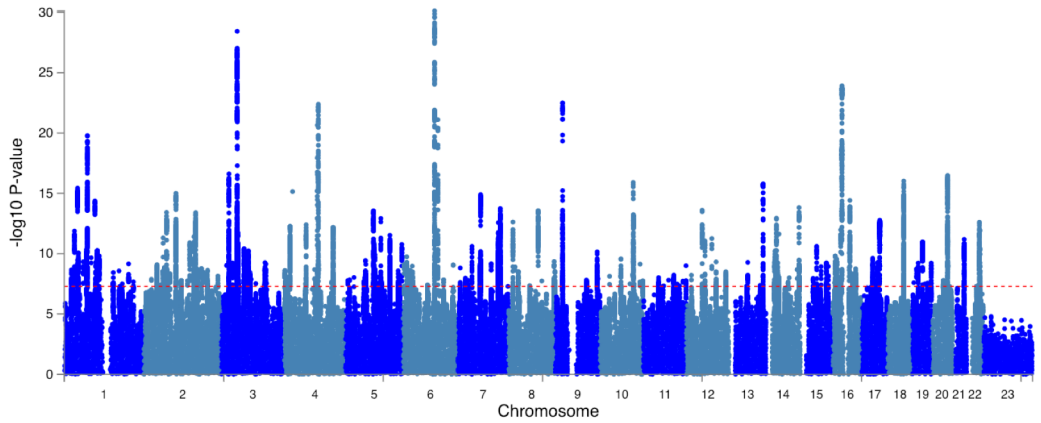
Feature	Value
id	3785
PMID	PMID: 29942086
Year	2018
File	https://ctg.cncr.nl/documents/p1651/SavageJansen_IntMeta_sumstats.zip
Website	https://ctg.cncr.nl/software/summary_statistics
Consortium	
Domain	Cognitive
ChapterLevel	Mental Functions
SubchapterLevel	Higher-Level Cognitive Functions
Trait	Intelligence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

SNPs plots

Manhattan plot

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

Download the plot as PNG JPG SVG PDF



Q-Q plot

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

Download the plot as PNG JPG SVG PDF

Top SNPs

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

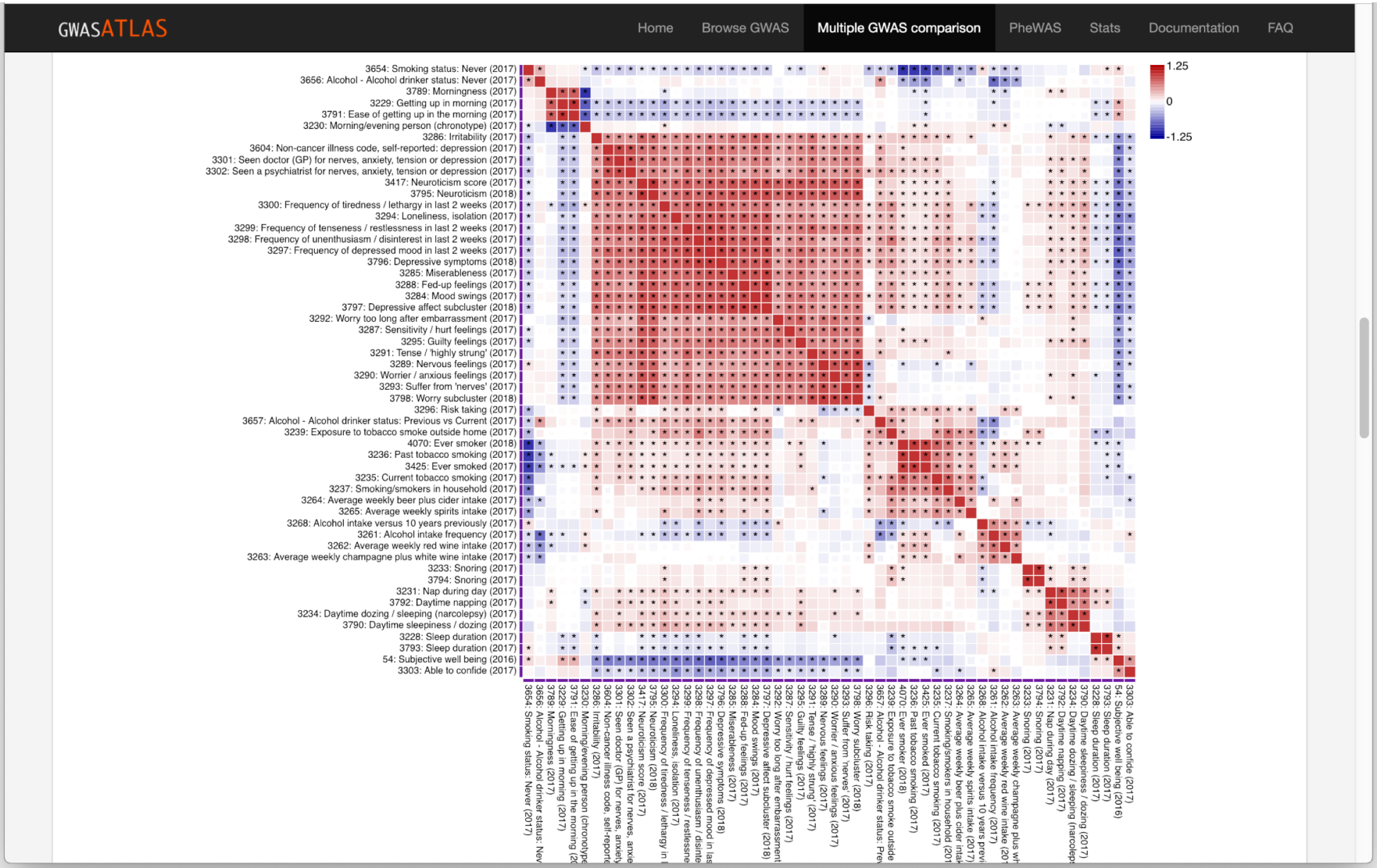
Show 10 entries Search:

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

CNCR CTGLAB

GWAS ATLAS

Multiple GWAS comparison (Genetic Correlation)



CNCR CTGLAB

GWAS ATLAS

Multiple GWAS comparison (Loci overlap)

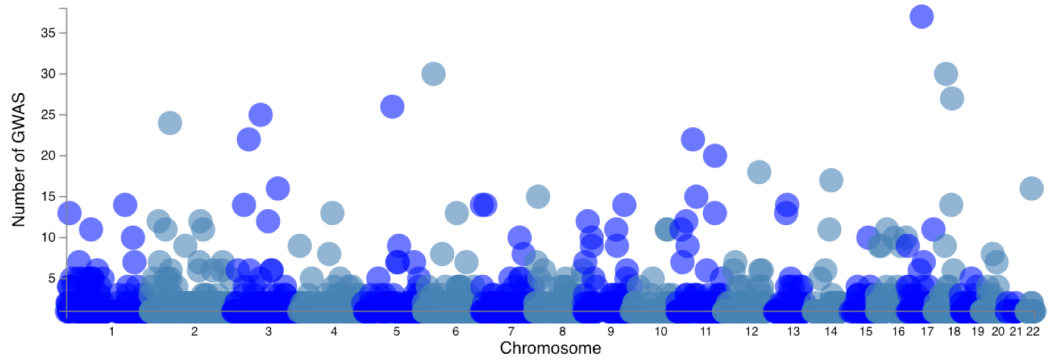
GWASATLAS | Home | Browse GWAS | Multiple GWAS comparison | PheWAS | Stats | Documentation | FAQ

Pleiotropic risk loci

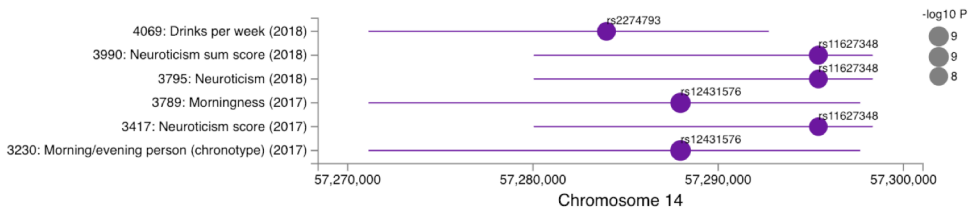
ℹ Each dot represents a group of physically overlapping risk loci. See Documentation for details. The genome wide plot can be zoomed in and out by scroll. When Y-axis is the number of associated domains, the dots are sized by the number of associated GWAS and vice versa. Note that the number of associated GWAS is the number of associated unique summary statistics which does not necessary reflect the number of associated unique traits when multiple summary statistics for a single trait are selected. By clicking a dot, another plot for a specific group of risk loci will be plotted (only if the number of GWAS in the grouped locus is > 1). Note that P-value < 1e-300 is replaced with 1e-300 (maximum -log10 P-value is 300 in this plot).

Y-axis: number of associated GWAS

Download the plot as PNG JPG SVG PDF



Download the plot as PNG JPG SVG PDF



Pleiotropic genes

GWAS ATLAS

Create PheWAS plot

GWASATLAS | Home | Browse GWAS | Multiple GWAS comparison | PheWAS | Stats | Documentation | FAQ

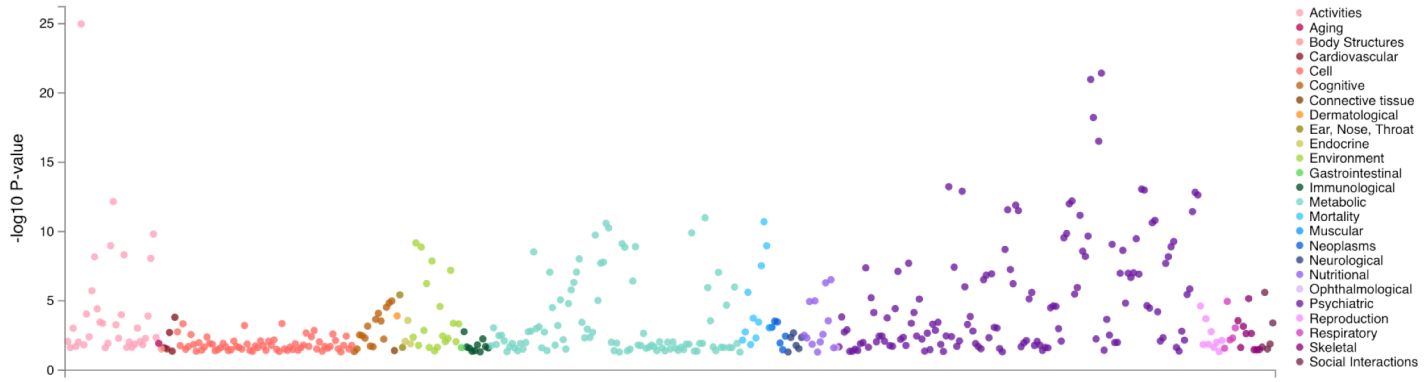
PheWAS plot

Sort traits by: Domain Clear text labels

max P-value: 0.05 Update

ℹ For SNPs, 0.05 is the maximum P-value. SNPs with P-value > 0.05 are not displayed regardless of this option due to low performance of the website.

Download the plot as PNG JPG SVG PDF



PheWAS table

Show 10 rows CSV Search:

atlas ID	PMID	Year	Domain	Trait	P-value	N
3312	This study	2017	Activities	Electronic device use - Plays computer games	1.0511e-25	386152
3990	29500382	2018	Psychiatric	Neuroticism sum score	3.7068e-22	380506
3795	29942085	2018	Psychiatric	Neuroticism	1.0576e-21	390278
4087	29255261	2018	Psychiatric	Neuroticism	5.9264e-19	329821
3417	This study	2017	Psychiatric	Neuroticism score	3.0434e-17	312740
3797	29942085	2018	Psychiatric	Depressive affect subcluster	5.8062e-14	357957
3287	This study	2017	Psychiatric	Sensitivity / hurt feelings	8.7375e-14	375272

CNCR CTGLAB

GWAS ATLAS

Cold Spring Harbor Laboratory
bioRxiv
THE PREPRINT SERVER FOR BIOLOGY
HOME | ABO
Search

New Results

Comment on this paper

A global view of pleiotropy and genetic architecture in complex traits

Kyoko Watanabe, Sven Stringer, Oleksandr Frei, Masa Umićević Mirkov, Tinca J.C. Polderman, Sophie van der Sluis, Ole A. Andreassen, Benjamin M. Neale, Danielle Posthuma

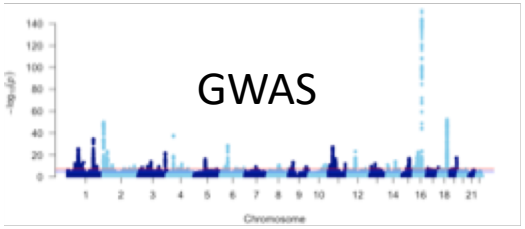
doi: https://doi.org/10.1101/500090

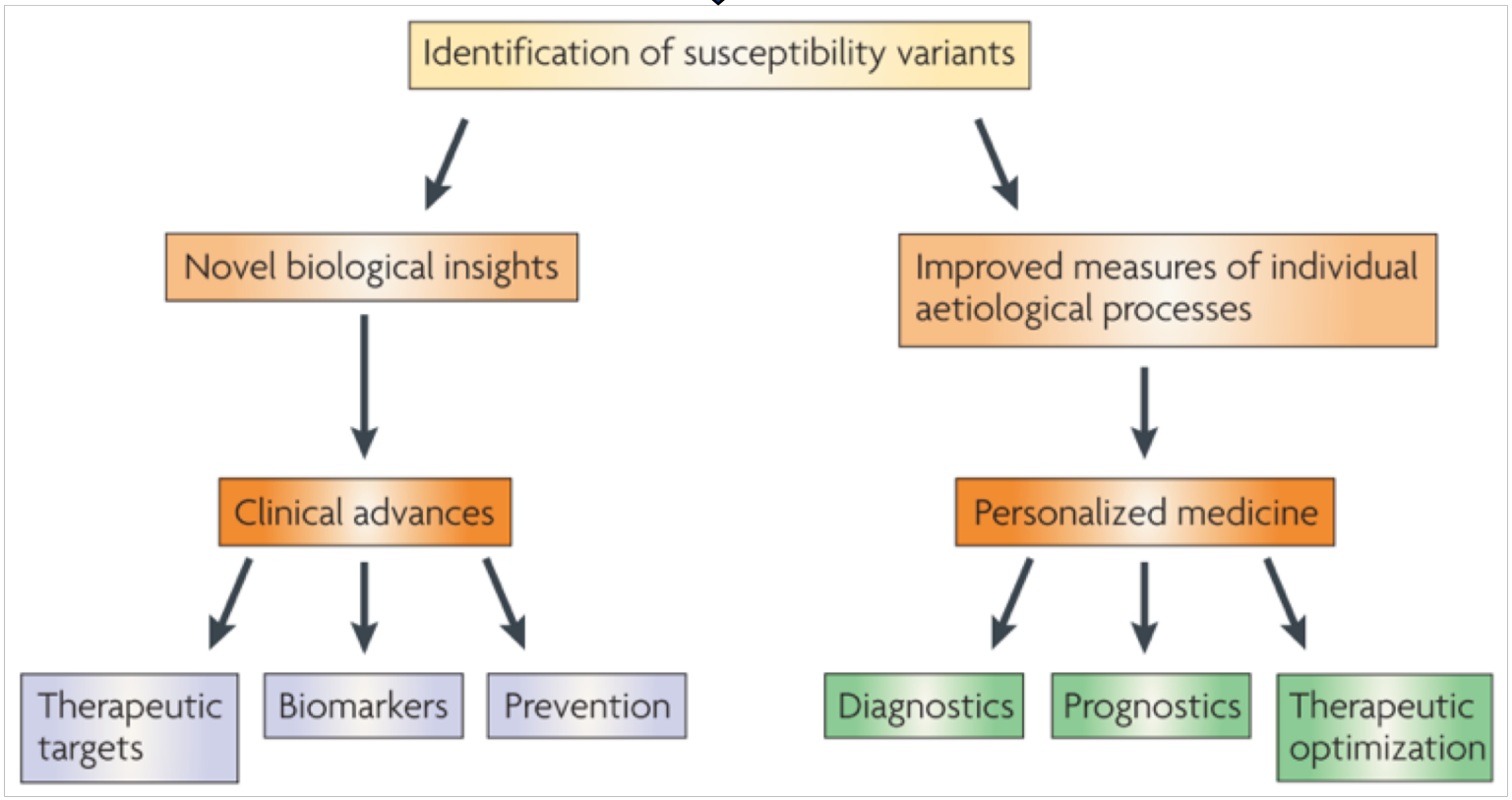
This article is a preprint and has not been peer-reviewed [what does this mean?].

https://www.biorxiv.org/content/10.1101/500090v1

CNCR CTGLAB

Genome-wide association studies (GWAS)



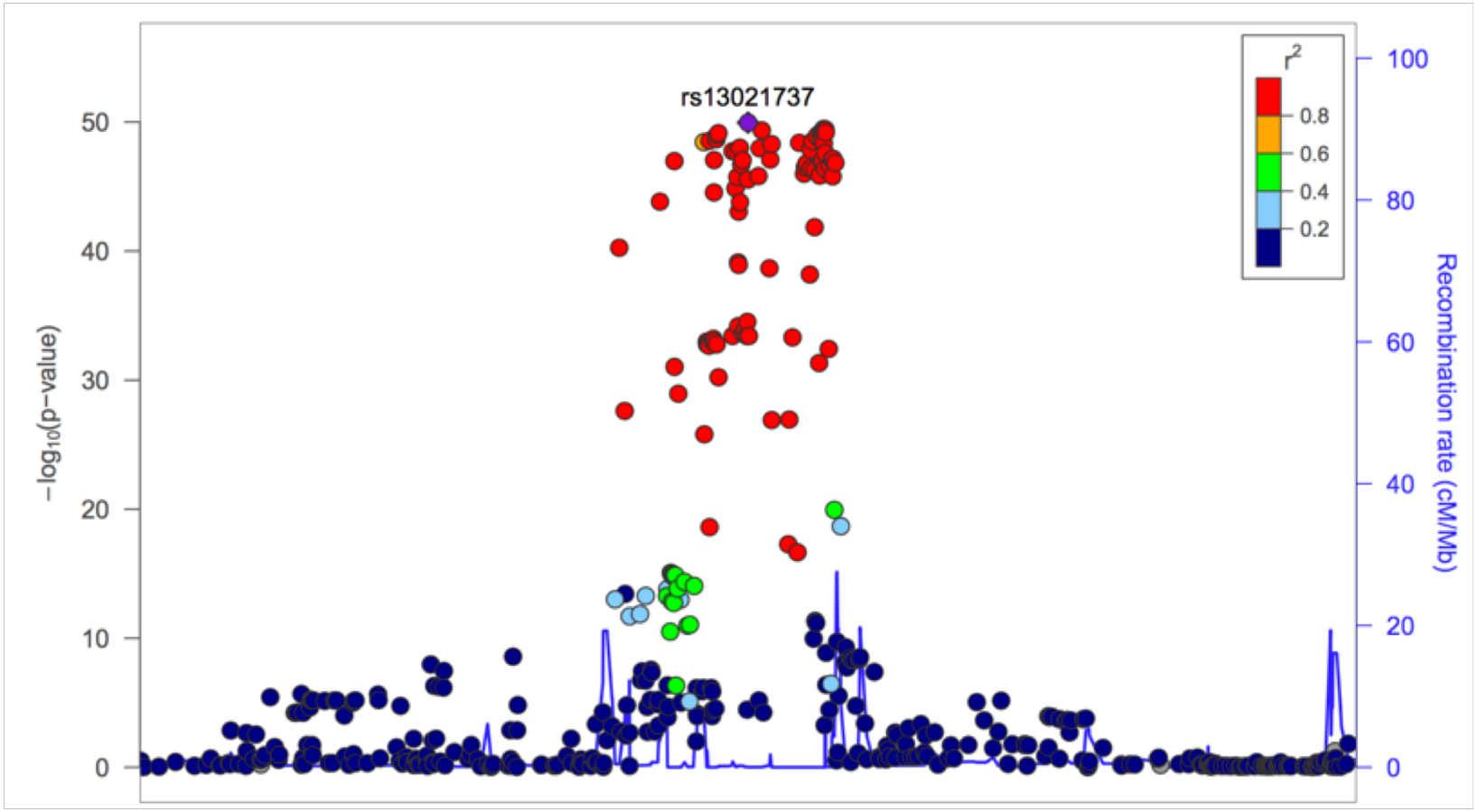


McCarthy et al. *Nat. Rev. Gent.* (2008)

CNCR CTGLAB

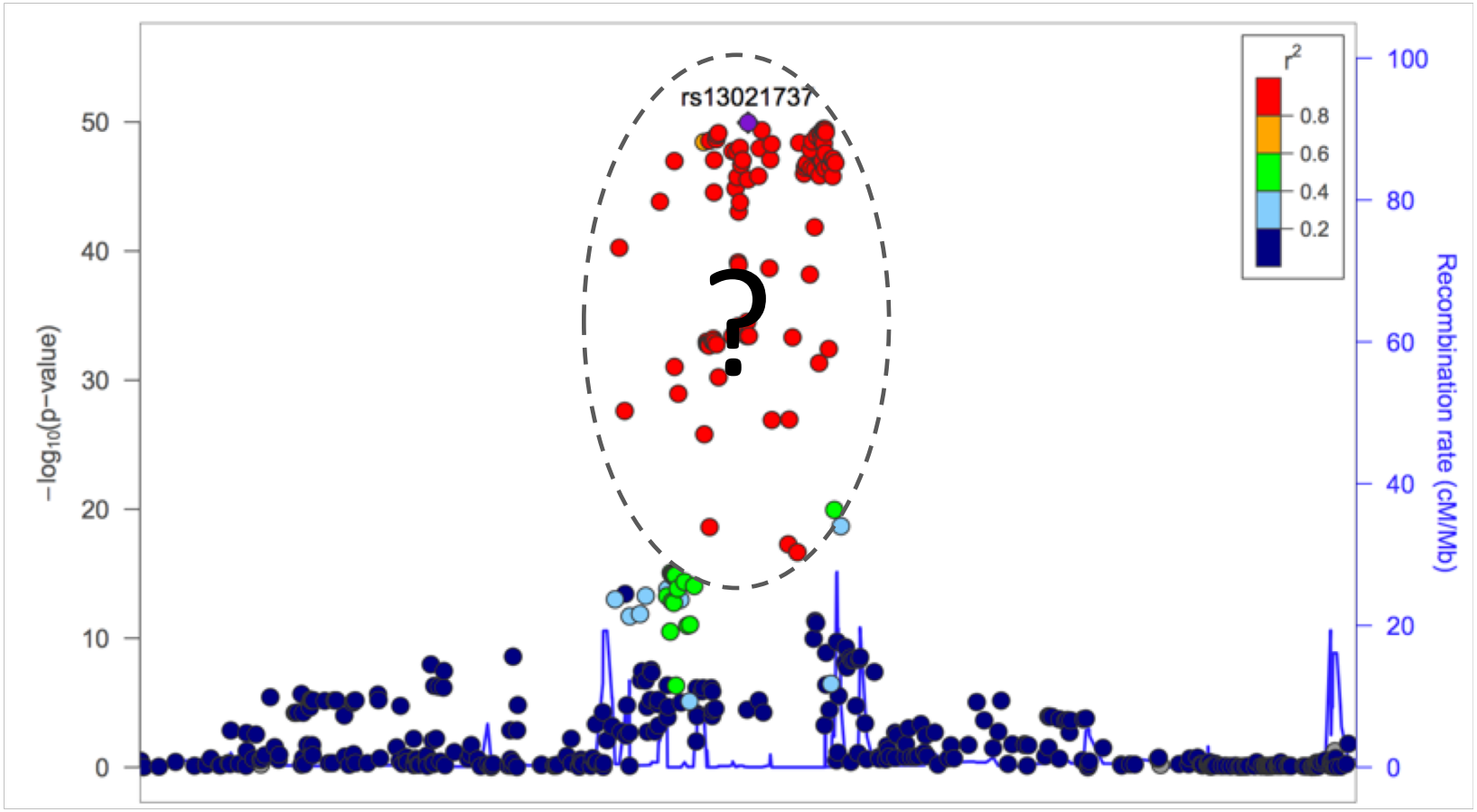
Challenge 1: Linkage disequilibrium

1. Linkage disequilibrium (LD)
   Non-random occurrence of SNPs



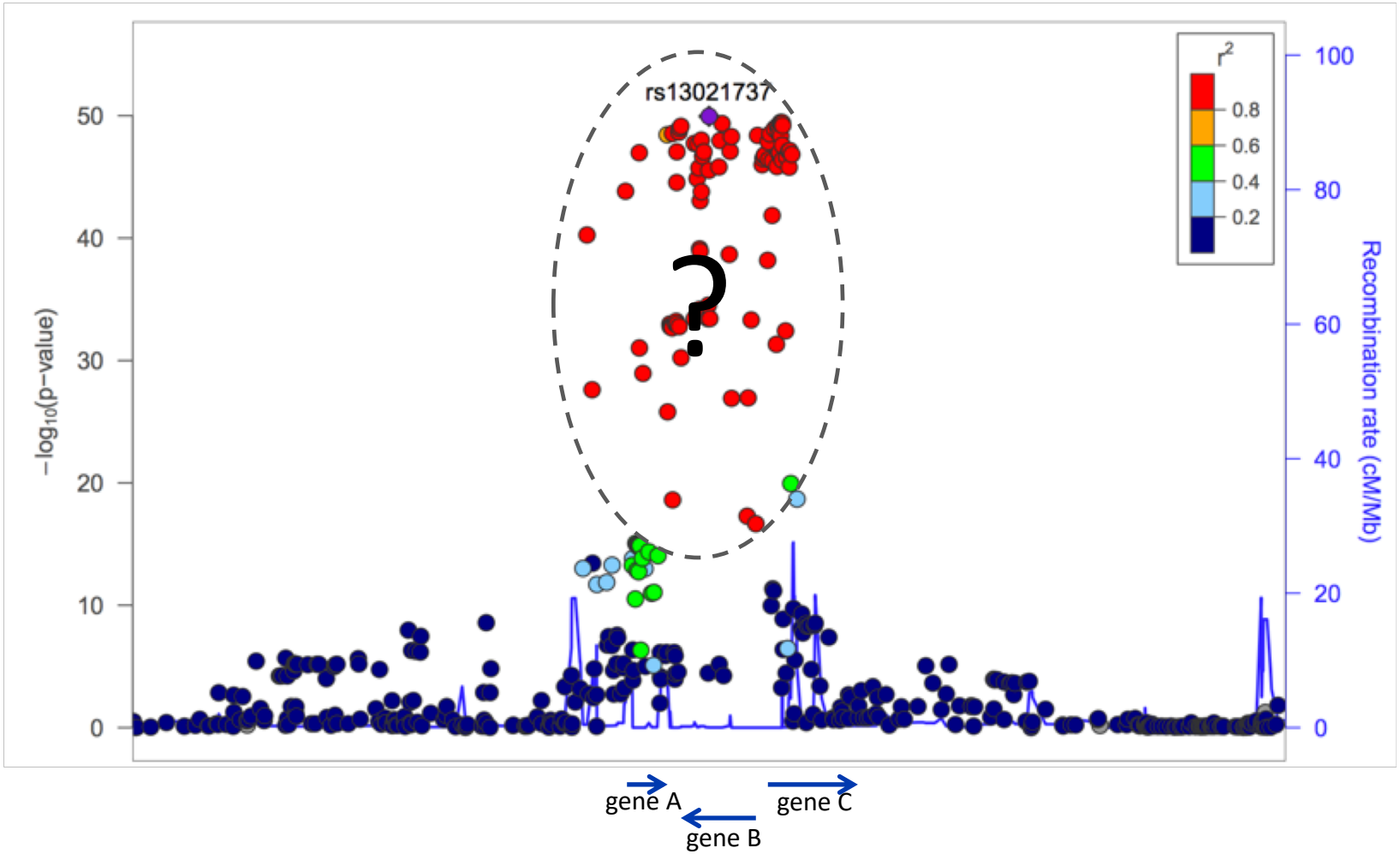
Challenge 1: Linkage disequilibrium

1. Linkage disequilibrium (LD)
   Non-random occurrence of SNPs



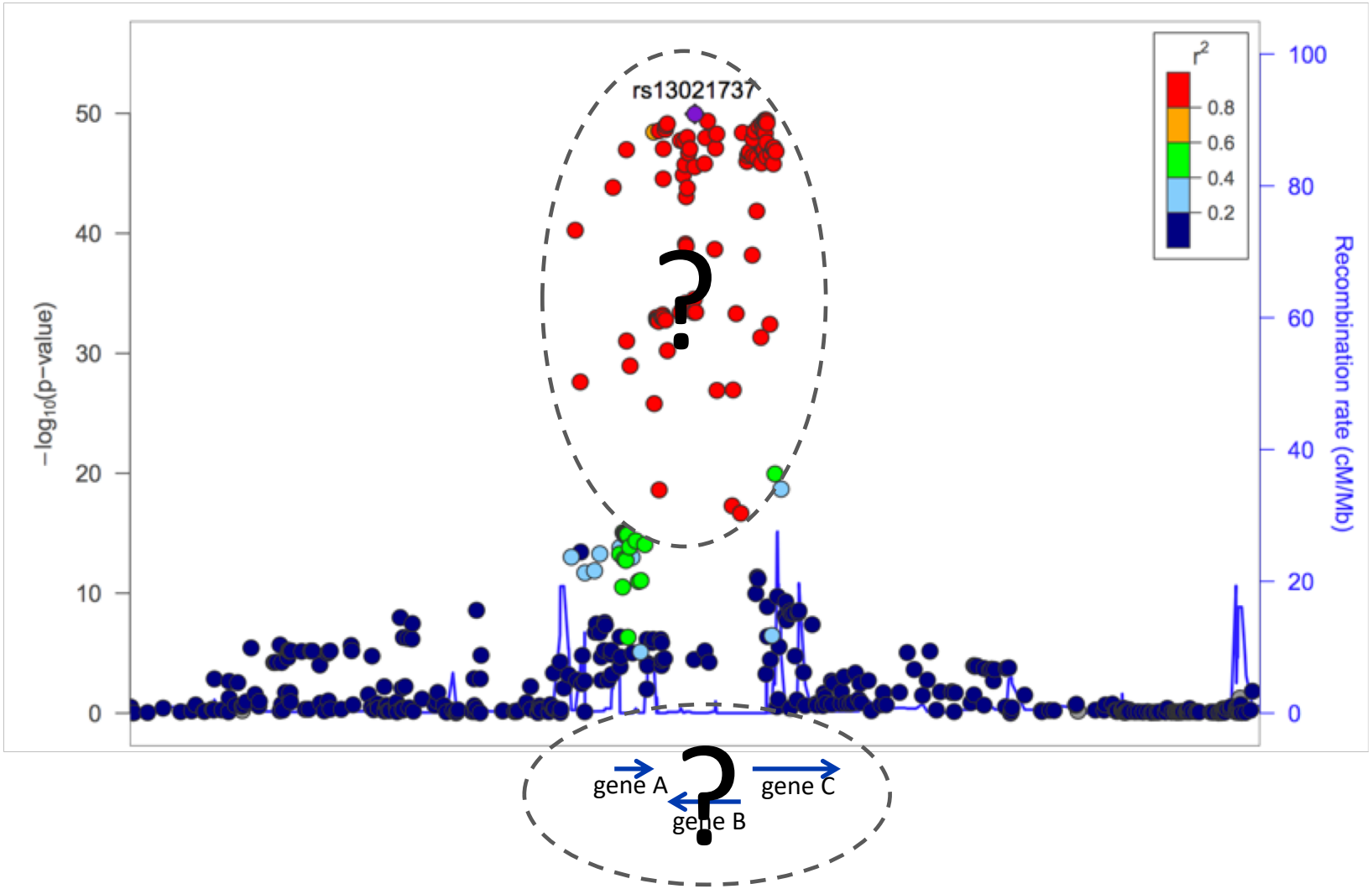
Challenge 1: Linkage disequilibrium

1. Linkage disequilibrium (LD)
   Non-random occurrence of SNPs



Challenge 1: Linkage disequilibrium

1. Linkage disequilibrium (LD)
   Non-random occurrence of SNPs



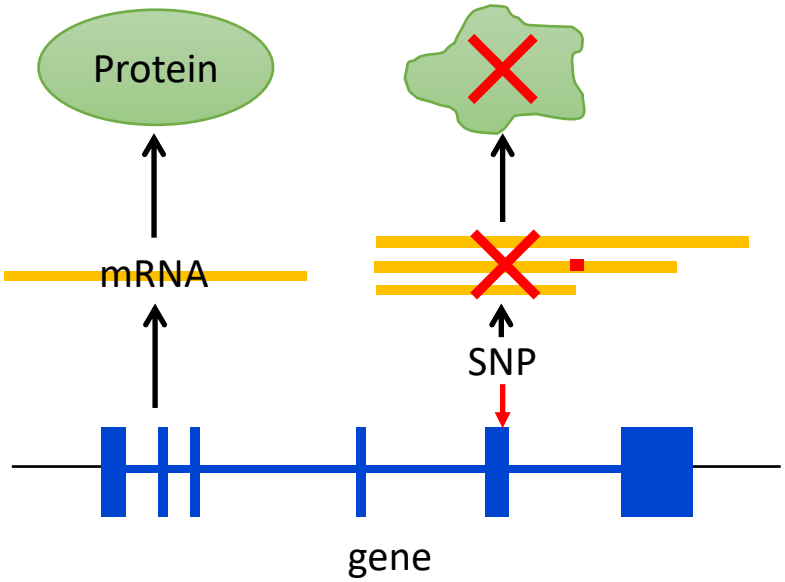
Challenge 2: Non-coding SNPs

2. non-coding SNPs

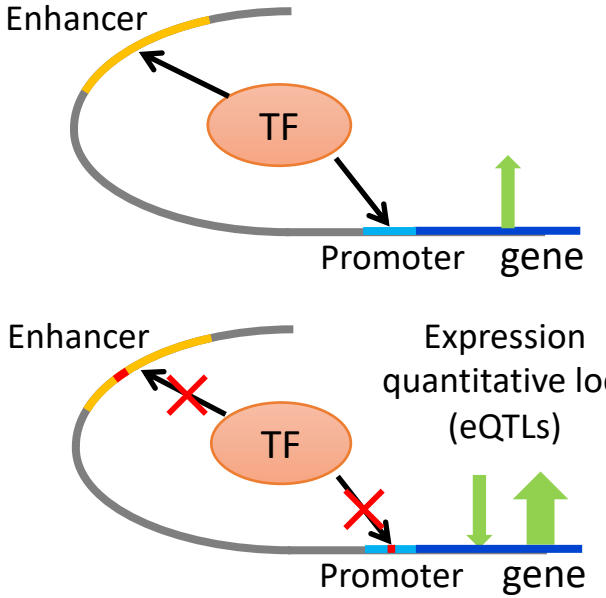
Coding SNPs
Non-synonymous SNPs
Loss of Function (Stop gain, Stop loss)
Splicing SNPs

Non-coding SNPs
Enhancer/Promoter regions
Transcription Factor Binding Sites (TFBSs)

Direct effect to genes/proteins

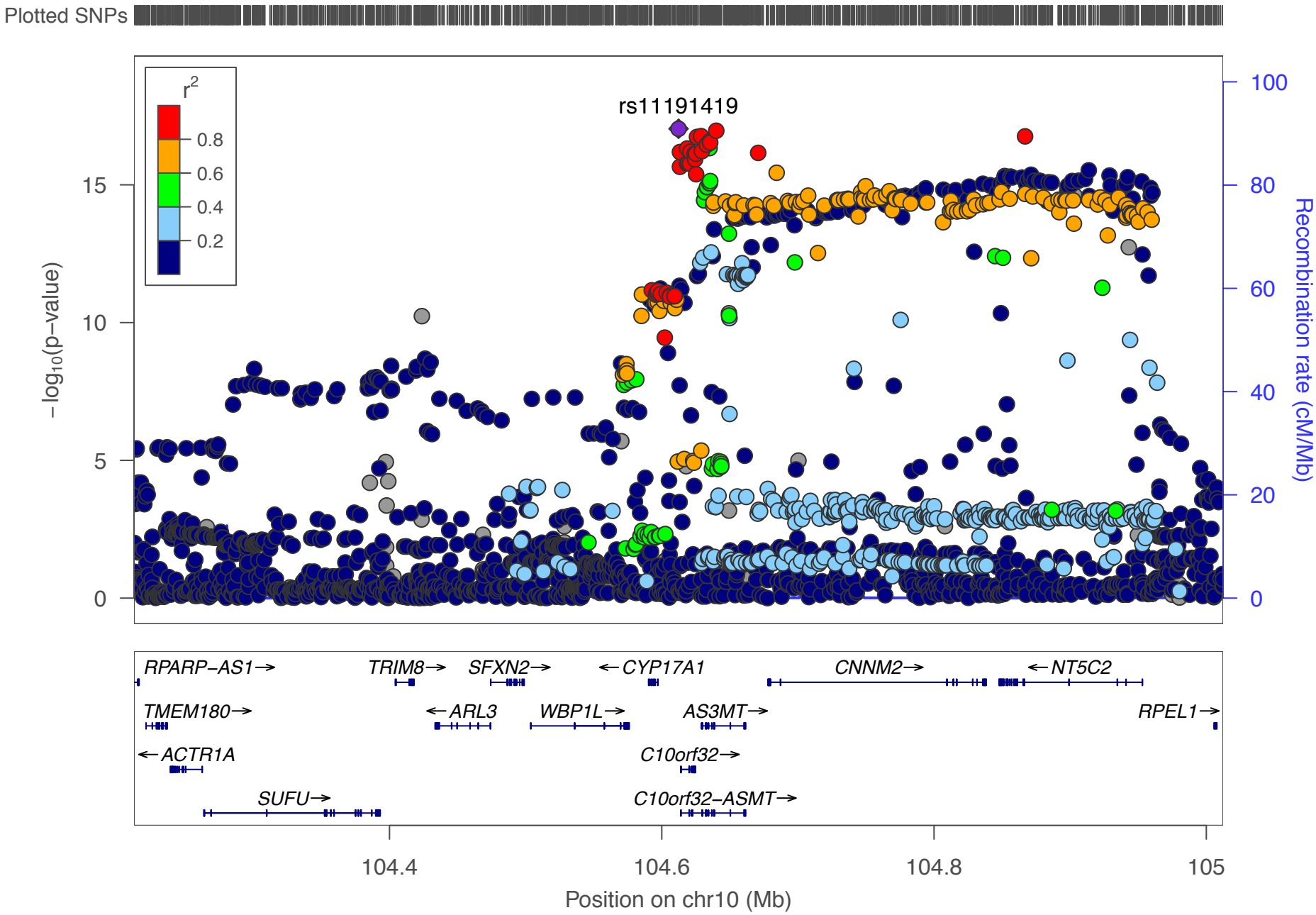


Indirect effect to genes

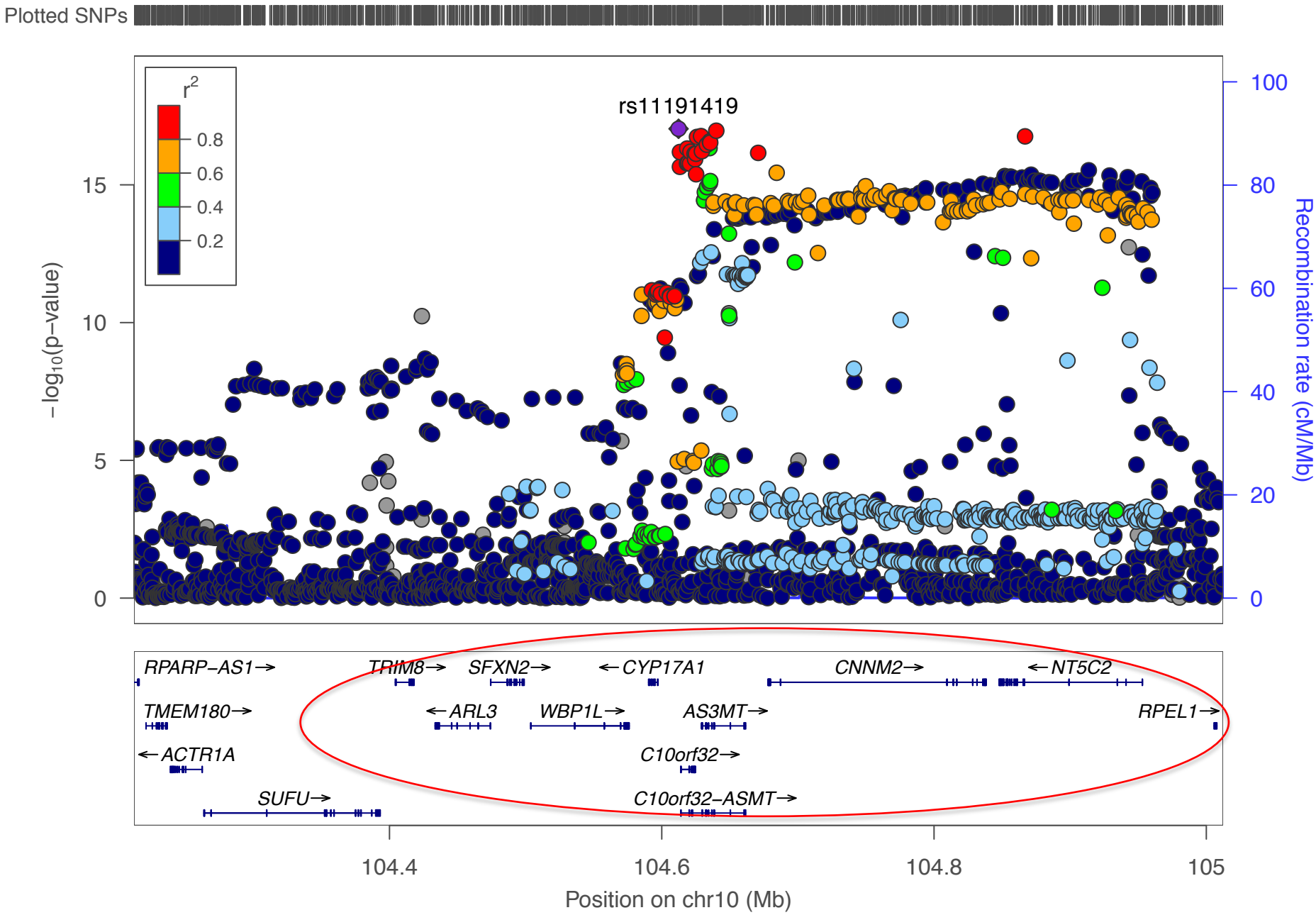


CNCR CTGLAB

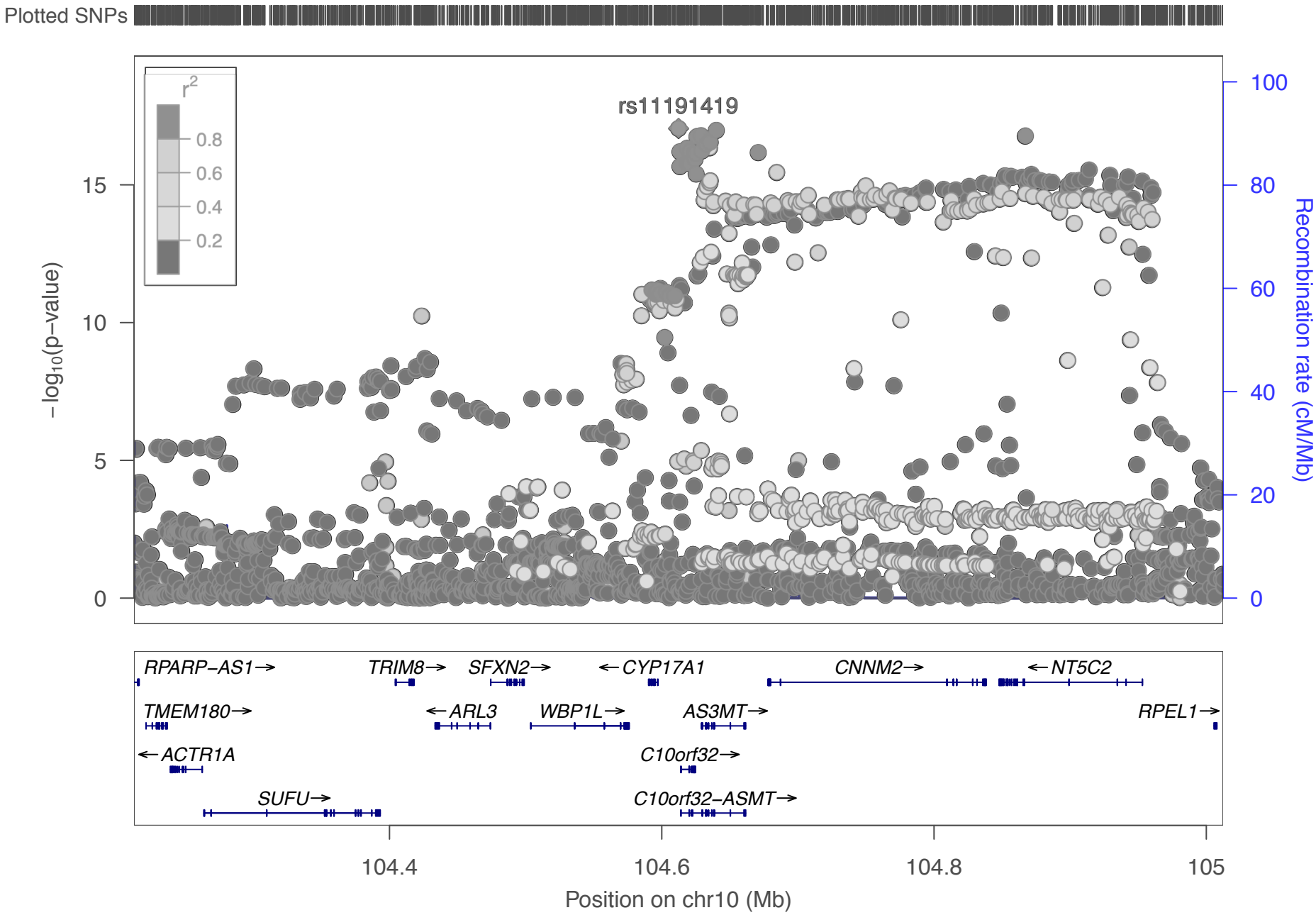
GWAS risk locus



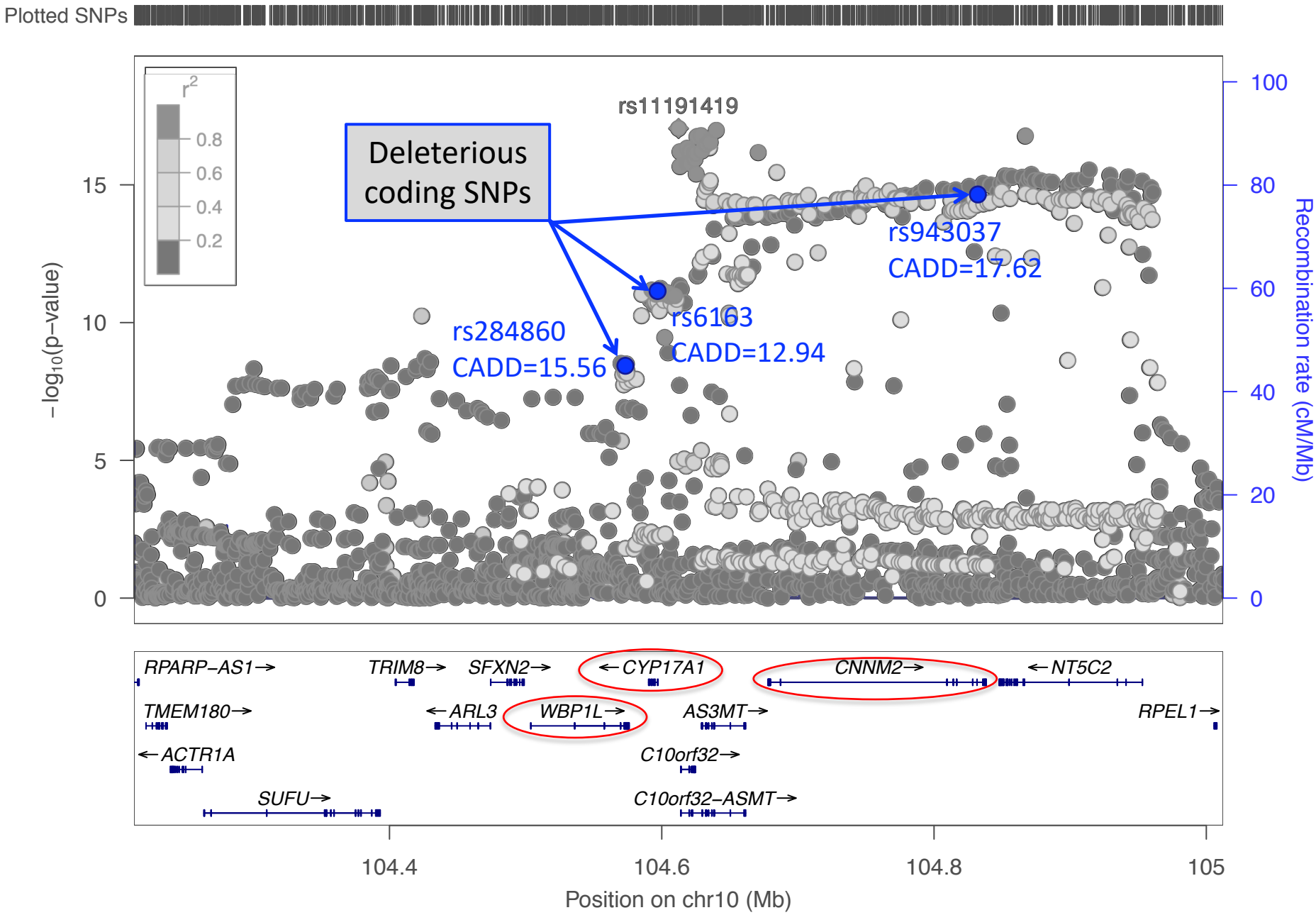
GWAS risk locus



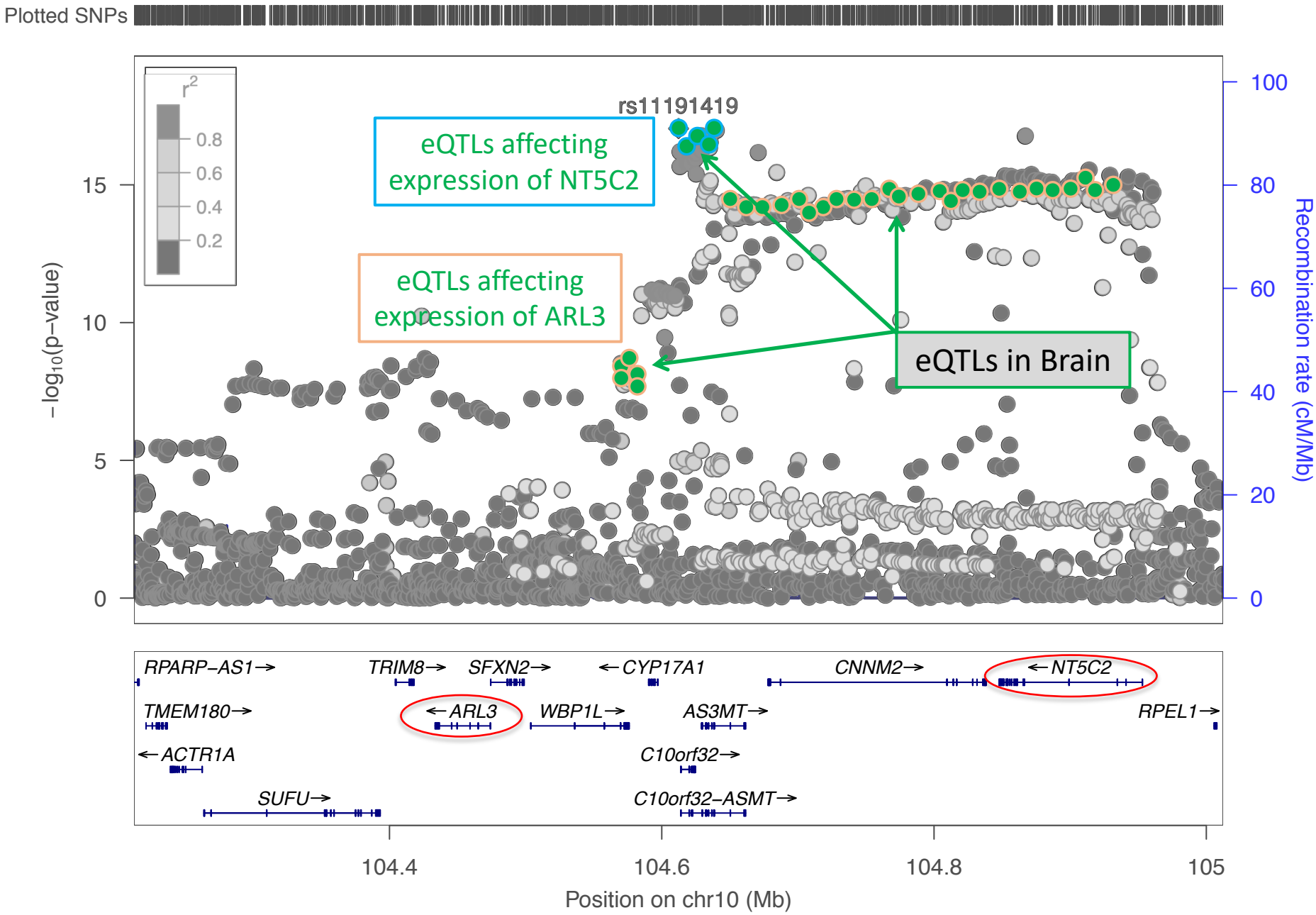
GWAS risk locus



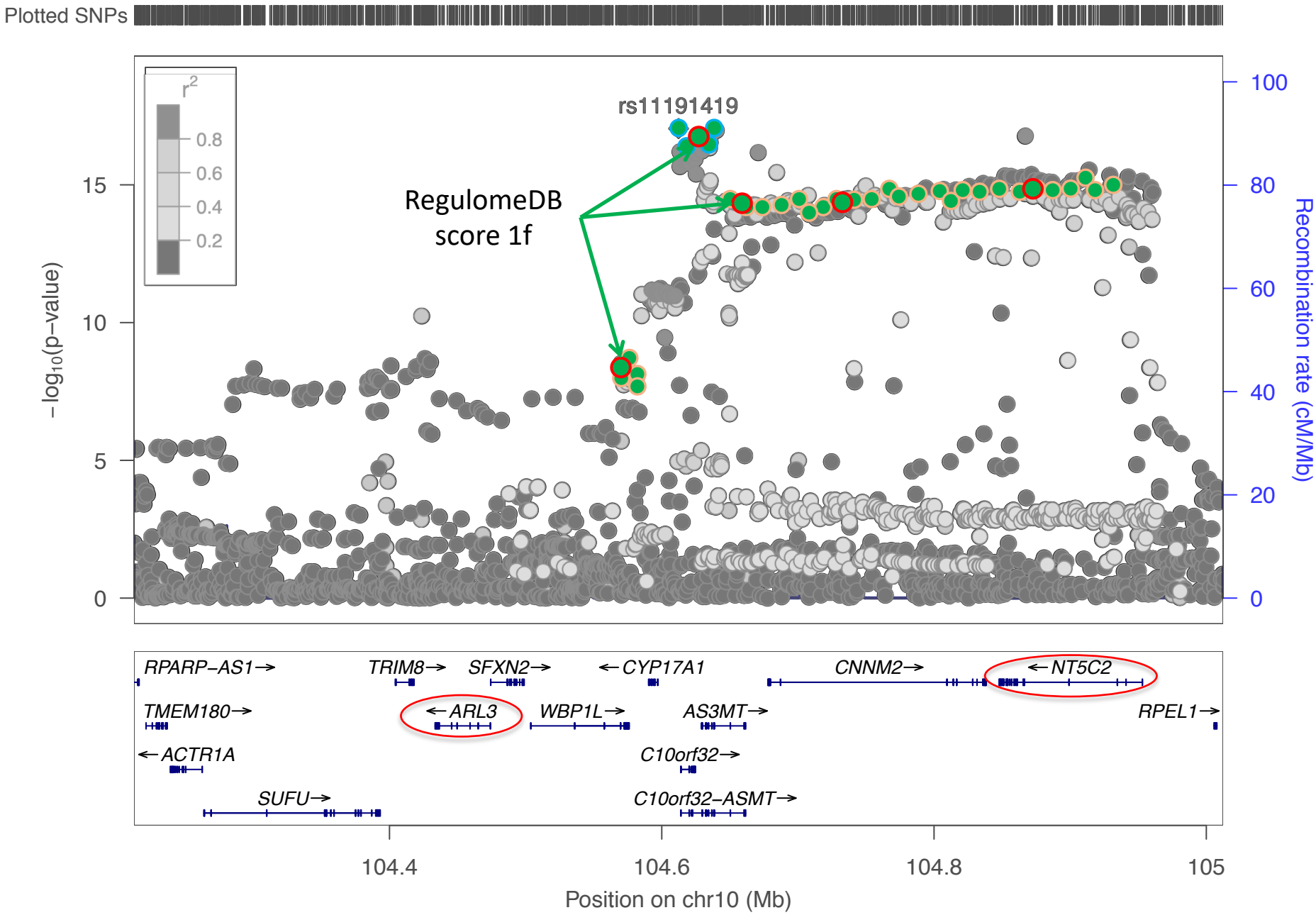
GWAS risk locus



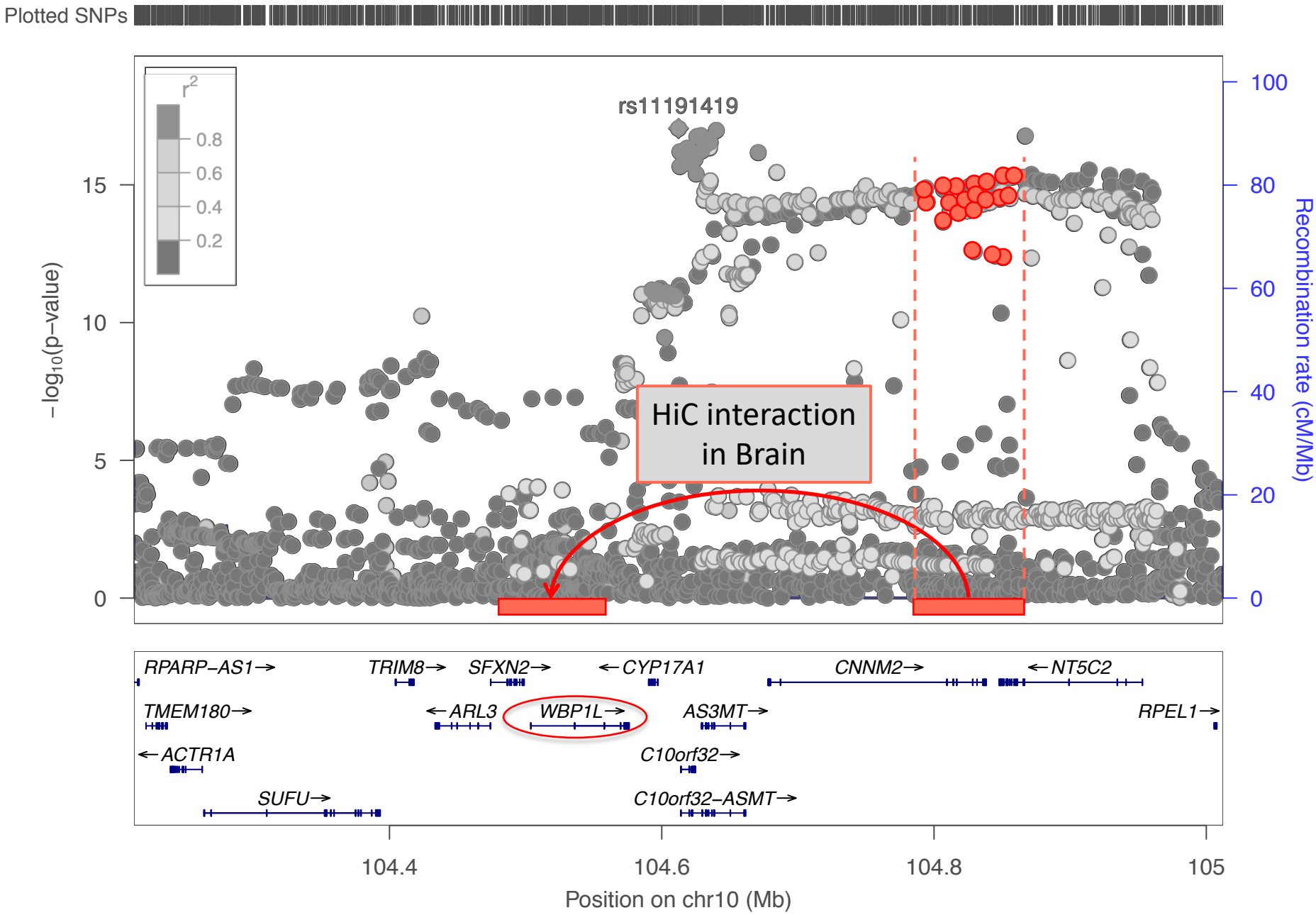
GWAS risk locus



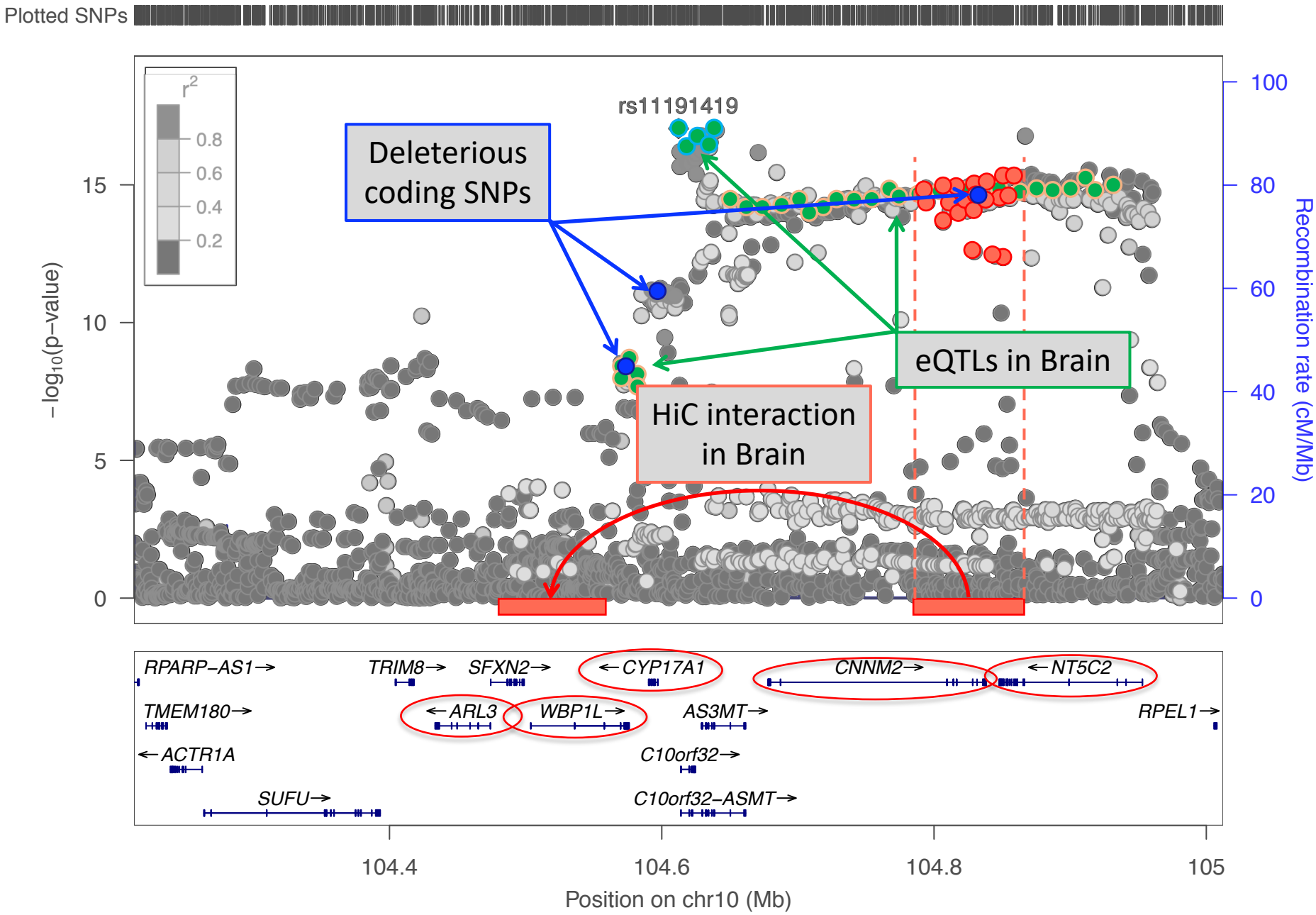
GWAS risk locus



GWAS risk locus



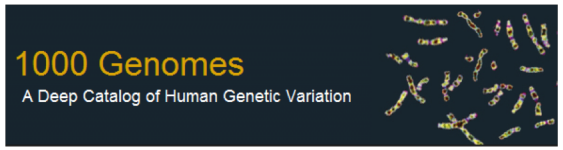
GWAS risk locus



Post-GWAS annotations and prioritization

PLINK

1. **Linkage disequilibrium (LD)**
   Identify all SNPs which are in LD of significant hits.

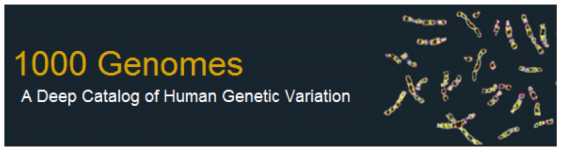


CNCR CTGLAB

Post-GWAS annotations and prioritization

PLINK

1. **Linkage disequilibrium (LD)**
   Identify all SNPs which are in LD of significant hits.



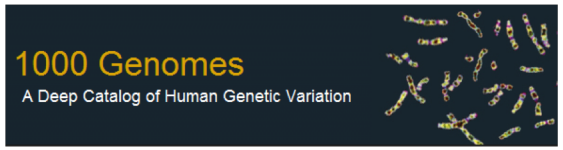
2. **Variant annotation**
   Functional consequence on genes (i.e. exonic, intronic or splicing site)

ANNOVAR

Post-GWAS annotations and prioritization

PLINK

1. **Linkage disequilibrium (LD)**
   Identify all SNPs which are in LD of significant hits.



2. **Variant annotation**
   Functional consequence on genes (i.e. exonic, intronic or splicing site)

ANNOVAR

3. **Functional annotation**
   Deleteriousness, regulatory elements and epigenetic data

CADD

GTEx Portal

BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

ENCODE

RegulomeDB

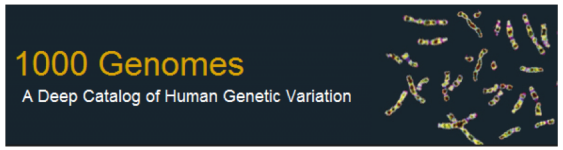
ROADMAP epigenomics PROJECT

HiC

Post-GWAS annotations and prioritization

PLINK

1. **Linkage disequilibrium (LD)**
   Identify all SNPs which are in LD of significant hits.



2. **Variant annotation**
   Functional consequence on genes (i.e. exonic, intronic or splicing site)

ANNOVAR

3. **Functional annotation**
   Deleteriousness, regulatory elements and epigenetic data

CADD

GTEx Portal

BBMRI.nl
Biobanking and BioMolecular resources Research Infrastructure The Netherlands

ENCODE

RegulomeDB

ROADMAP epigenomics PROJECT

HiC

4. **Functional analyses of genes**
   Tissue specific expression, gene set analyses

MSigDB
Molecular Signatures Database

WikiPathways
Pathways for the People

KEGG
Kyoto Encyclopedia of Genes and Genomes

GEO
Gene Expression Omnibus

FUMA: Functional Mapping and Annotation of genetic associations

Available at http://fuma.ctglab.nl

FUMA GWAS

Functional Mapping and Annotation of Genome-Wide Association Studies

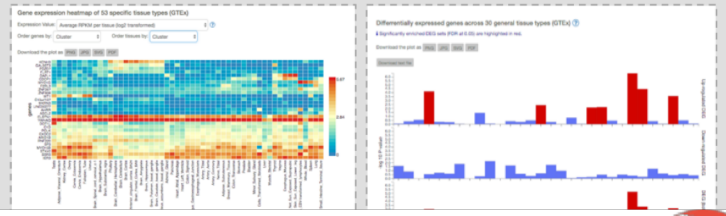
FUMA is a platform that can be used to annotate, prioritize, visualize and interpret GWAS results.
The SNP2GENE function takes GWAS summary statistics as an input, and provides extensive functional annotation for all SNPs in genomic areas identified by lead SNPs.
The GENE2FUNC function takes a list of 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 example results of FUMA for a few GWAS from Browse Examples without registration or login.

Please post any questions, suggestions and bug reports on Google Forum: FUMA GWAS users.

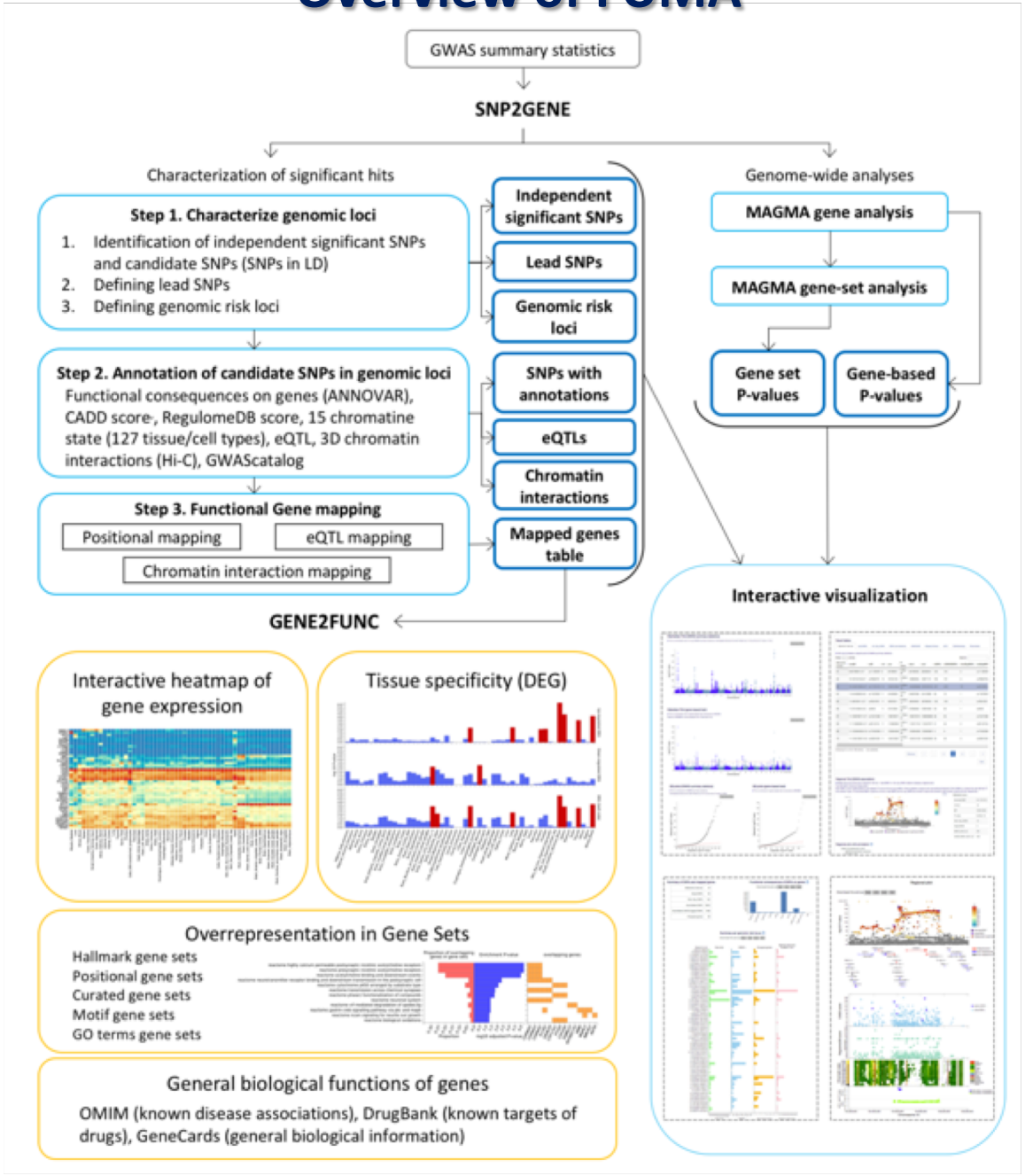
Citation:
When using FUMA, please cite the following.
K. Watanabe, E. Taskesen, A. van Bochoven and D. Posthuma. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**:1826. (2017).
https://www.nature.com/articles/s41467-017-01261-5
Depending on which results you are going to report, please also cite the original study of data sources/tools used in FUMA (references are available at Links).

SNP2GENE

GENE2FUNC

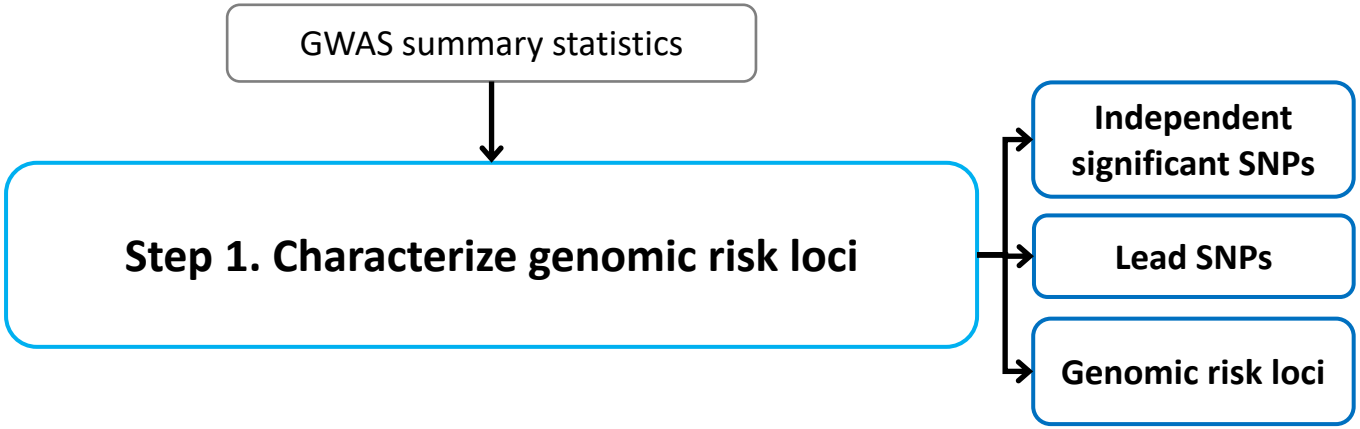


Overview of FUMA

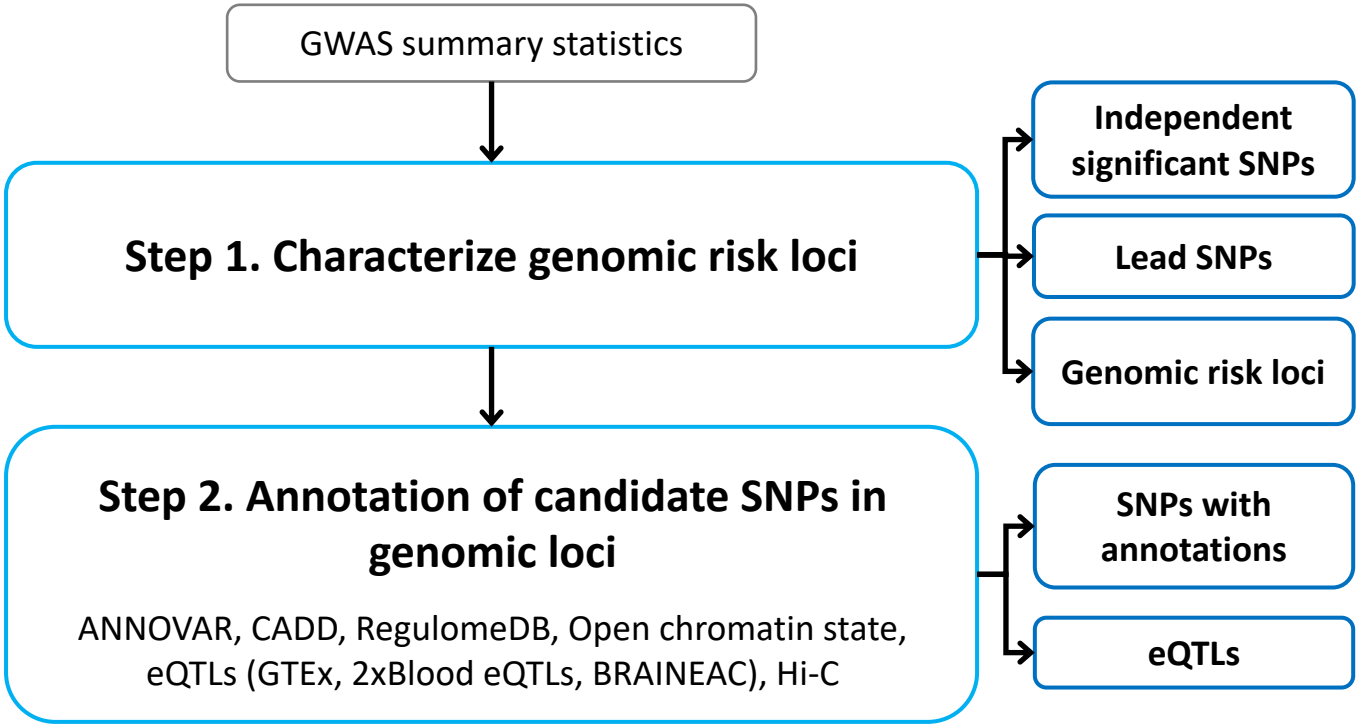


CNCR CTGLAB

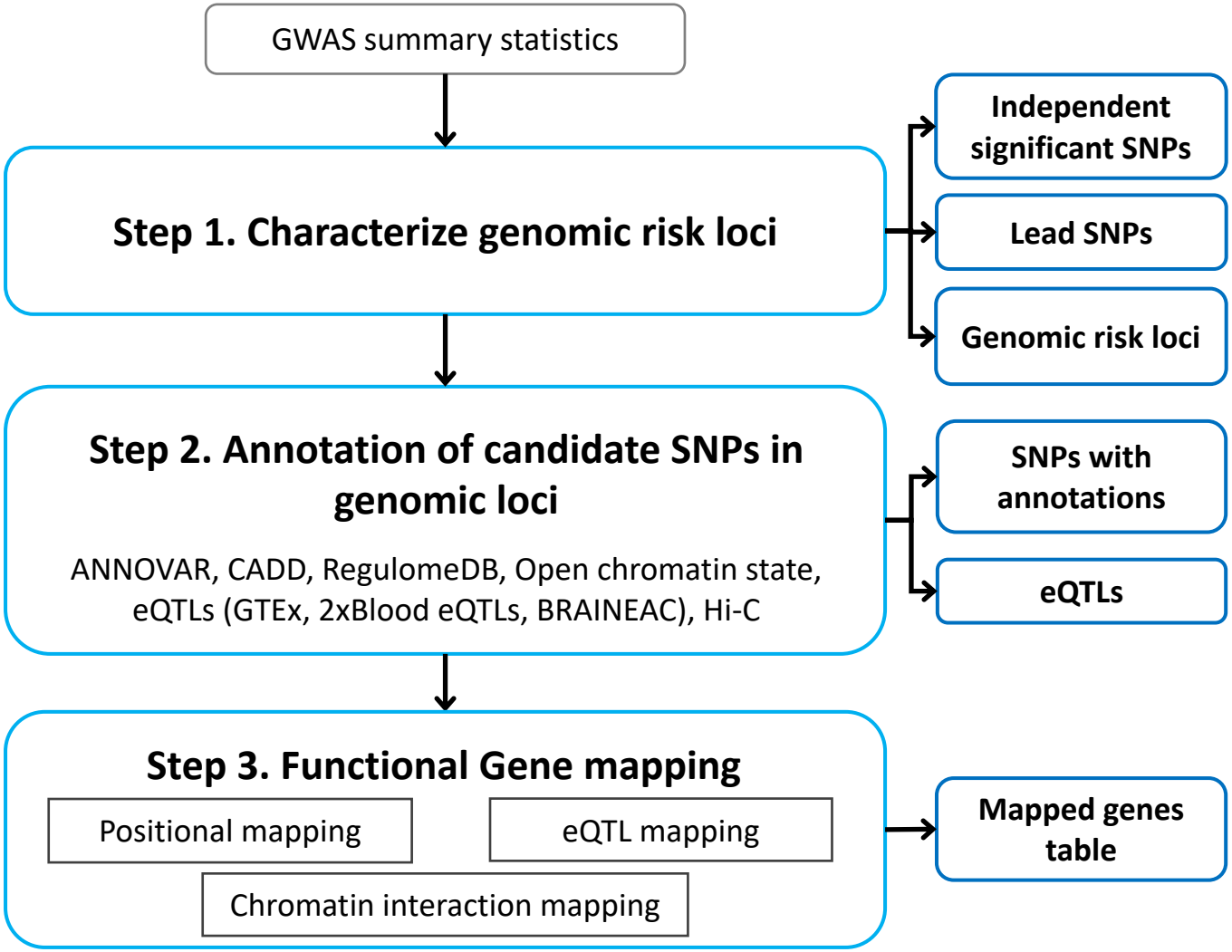
SNP2GENE


GWAS summary statistics
Step 1. Characterize genomic risk loci
Independent significant SNPs
Lead SNPs
Genomic risk loci

SNP2GENE

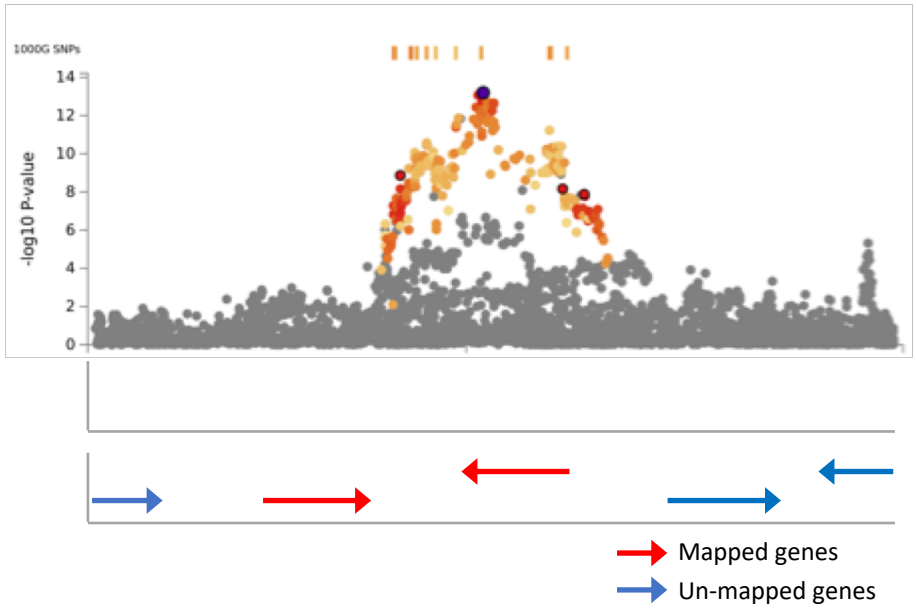

GWAS summary statistics
Step 1. Characterize genomic risk loci
Independent significant SNPs
Lead SNPs
Genomic risk loci
Step 2. Annotation of candidate SNPs in genomic loci
ANNOVAR, CADD, RegulomeDB, Open chromatin state, eQTLs (GTEx, 2xBlood eQTLs, BRAINEAC), Hi-C
SNPs with annotations
eQTLs

SNP2GENE


GWAS summary statistics
Step 1. Characterize genomic risk loci
Independent significant SNPs
Lead SNPs
Genomic risk loci
Step 2. Annotation of candidate SNPs in genomic loci
ANNOVAR, CADD, RegulomeDB, Open chromatin state, eQTLs (GTEx, 2xBlood eQTLs, BRAINEAC), Hi-C
SNPs with annotations
eQTLs
Step 3. Functional Gene mapping
Positional mapping
eQTL mapping
Chromatin interaction mapping
Mapped genes table

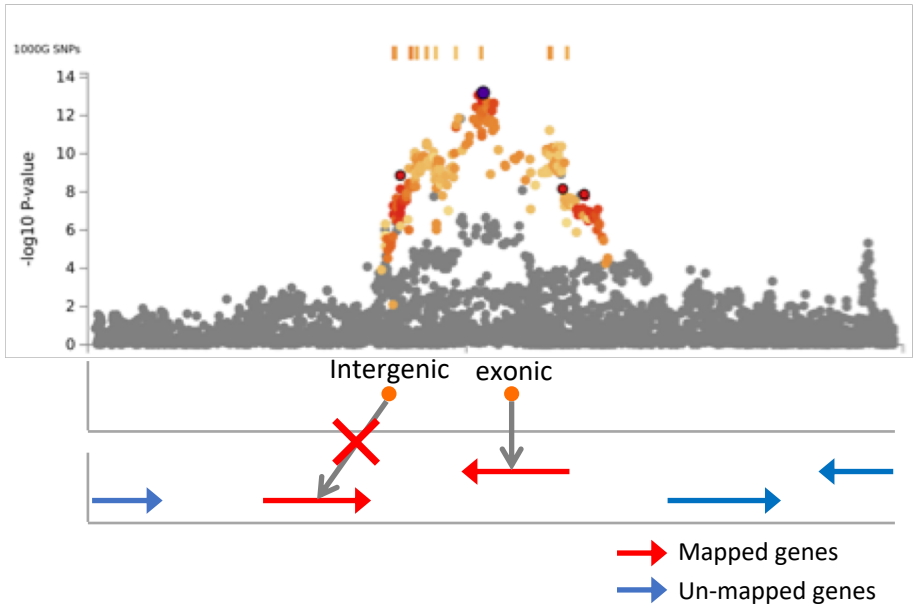
Gene mapping

Positional mapping



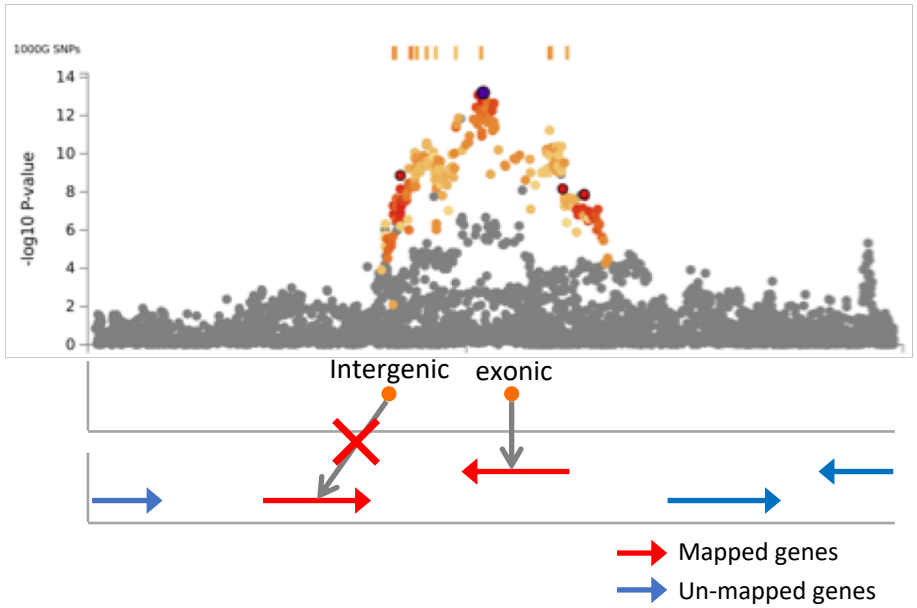
Gene mapping

Positional mapping

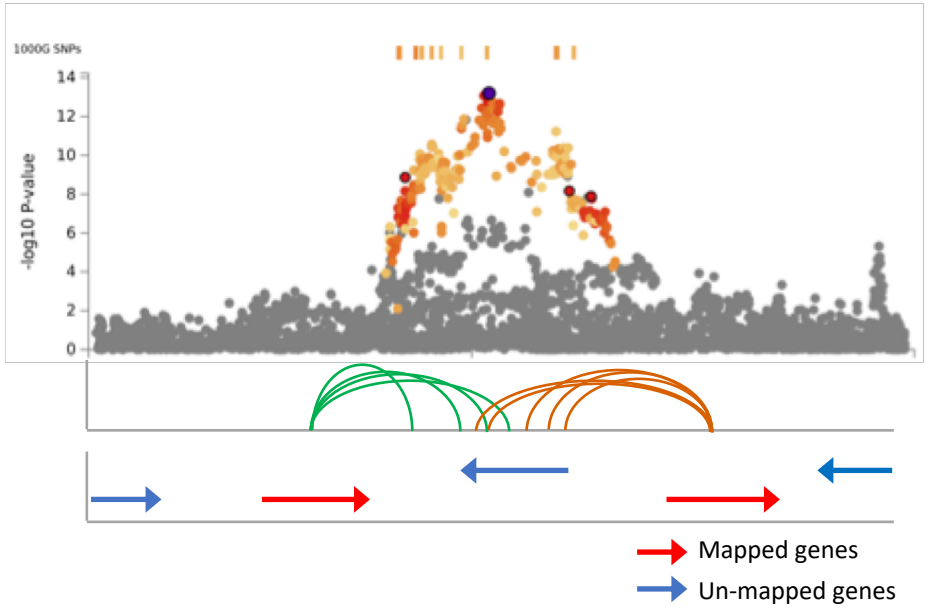


Gene mapping

Positional mapping



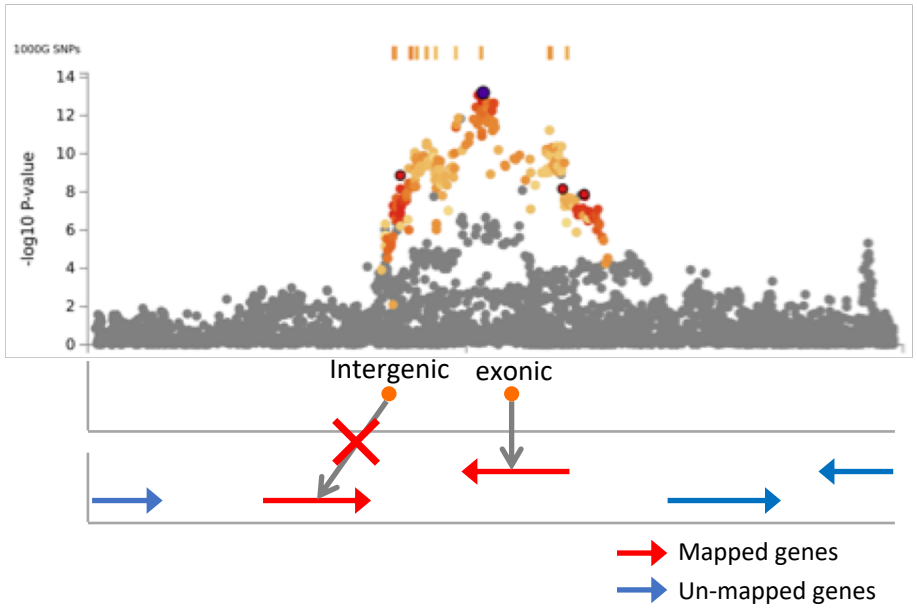
eQTL mapping



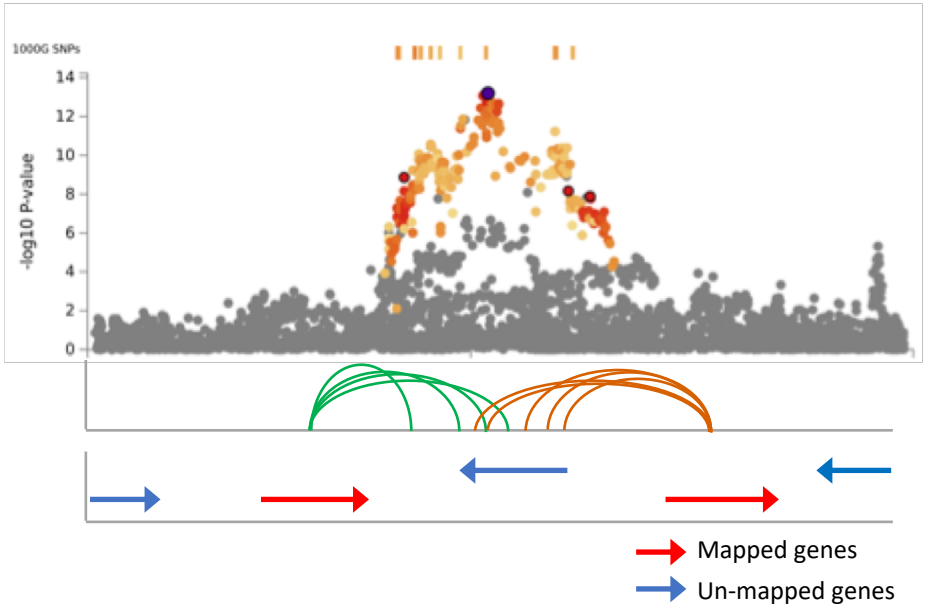
CNCR CTGLAB

Gene mapping

Positional mapping



eQTL mapping



Further filtering of SNPs based on functional annotations prior to gene mapping

CADD score (>12.37 is considered as highly deleterious)

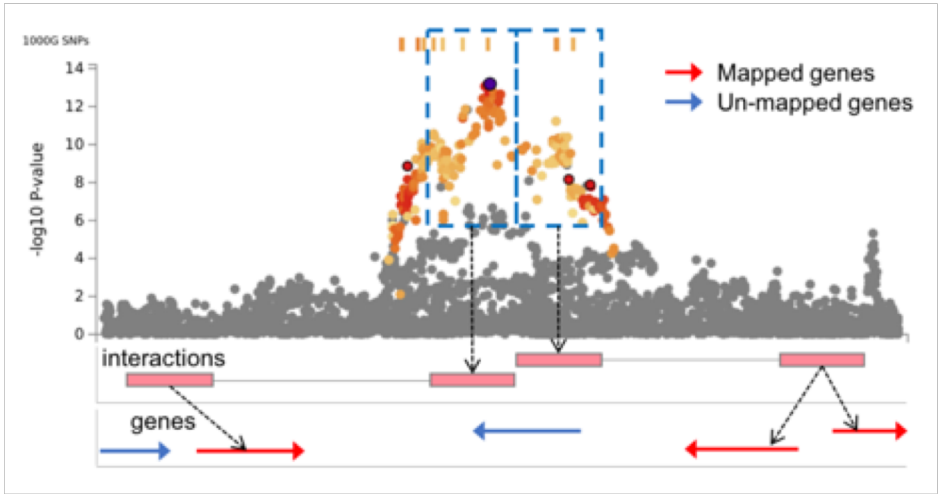
RegulomeDB score (categorical score from 1a to 7)

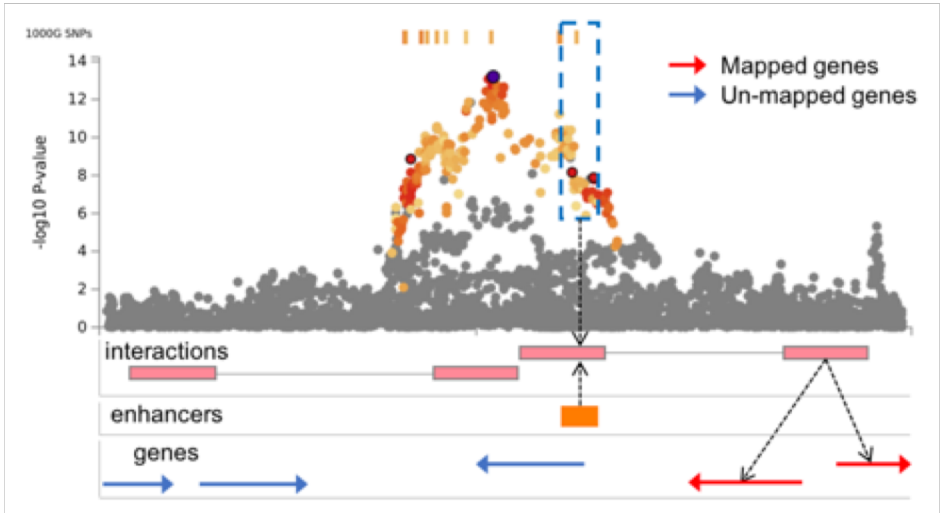
15-core chromatin status (1-7 states are considered as open, tissue specific)

*Parameters can be set for each mapping separately

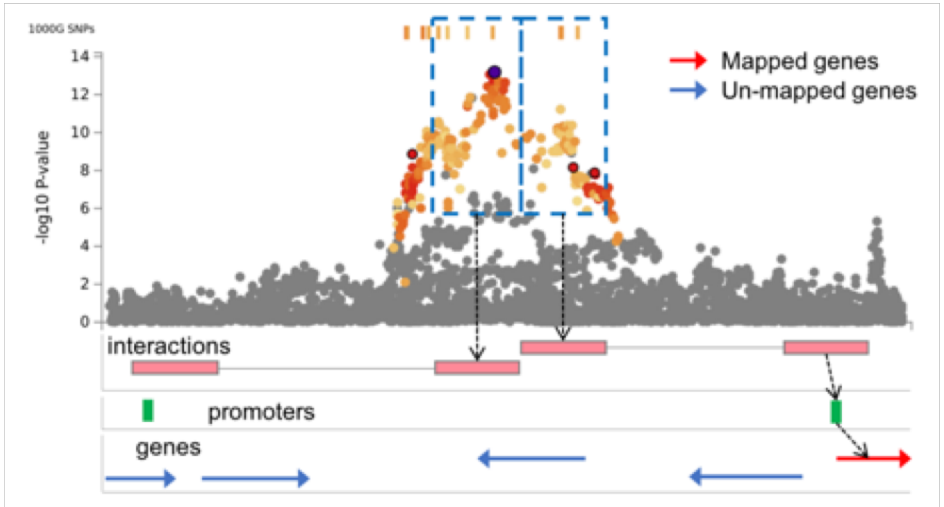
Gene mapping

Chromatin interaction mapping



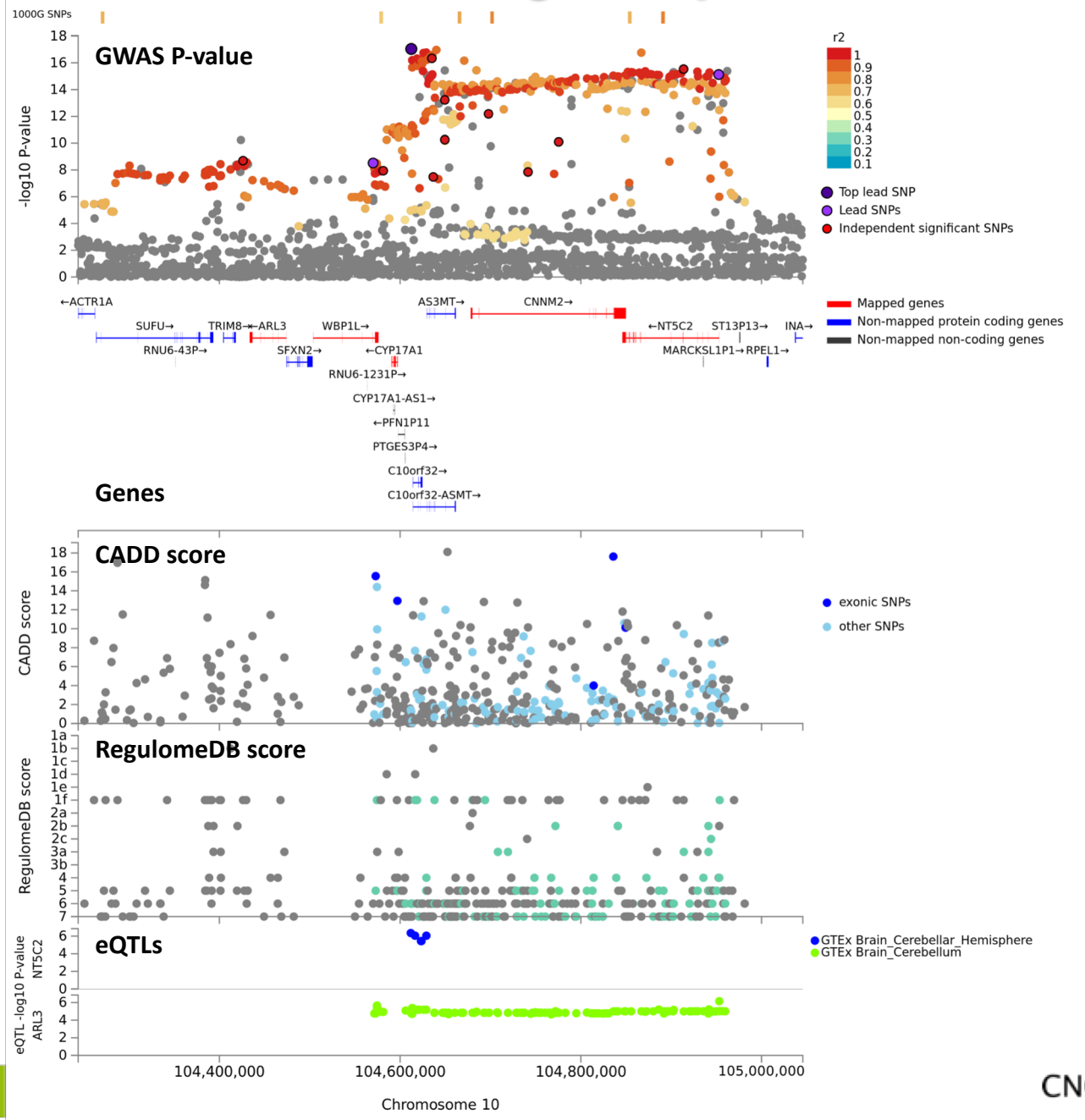


Map SNPs in a genomic region interacting with promoter region of genes (250bp up- and 500bp downstream of TSS) in selected tissue types. In addition to functional annotations, SNPs can be further filtered on such that overlap with predicted enhancers from Roadmap. Mapping can be limit to genes whose promoter regions are overlapping with predicted promoters from Roadmap.

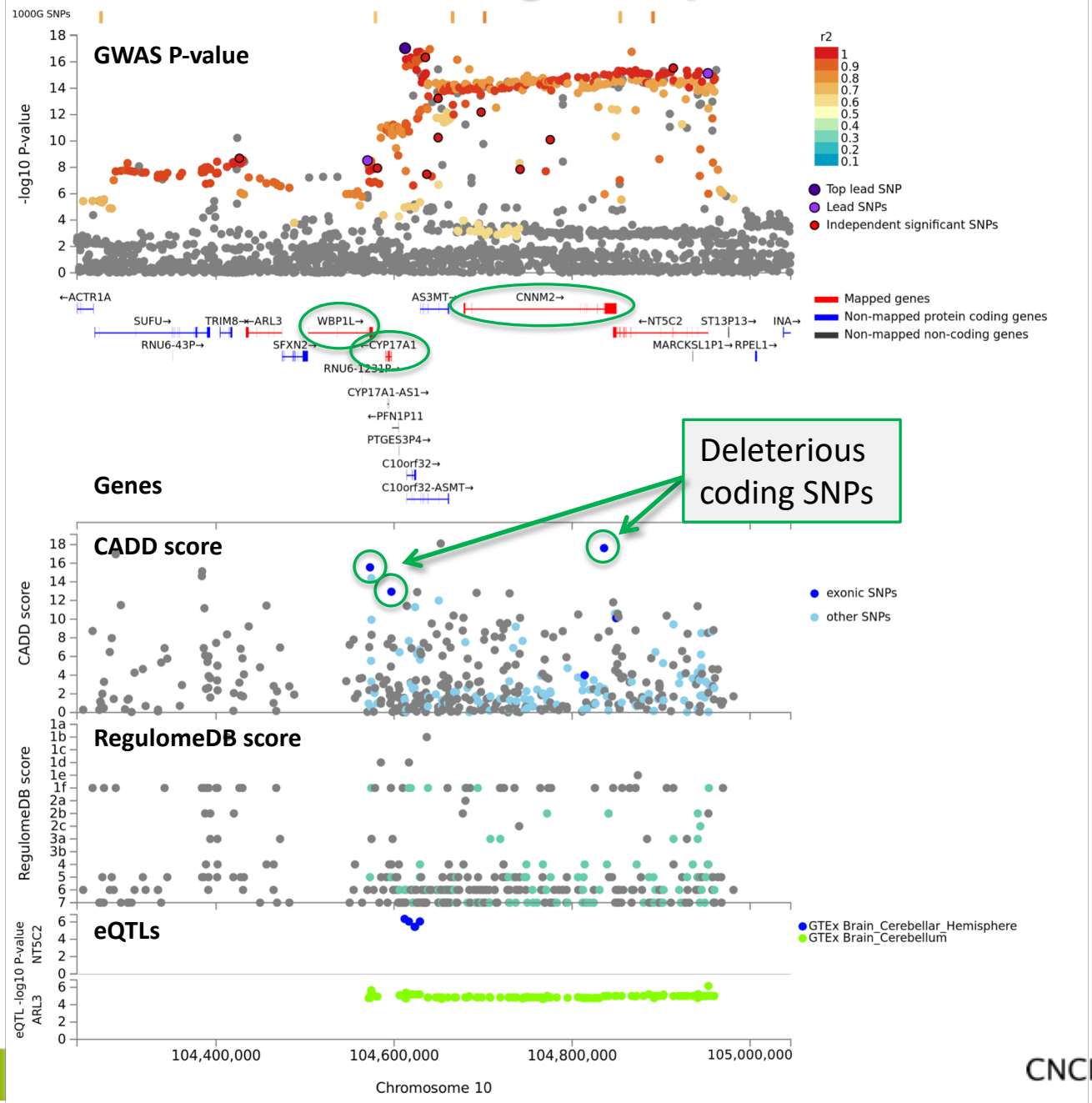


CNCR CTGLAB

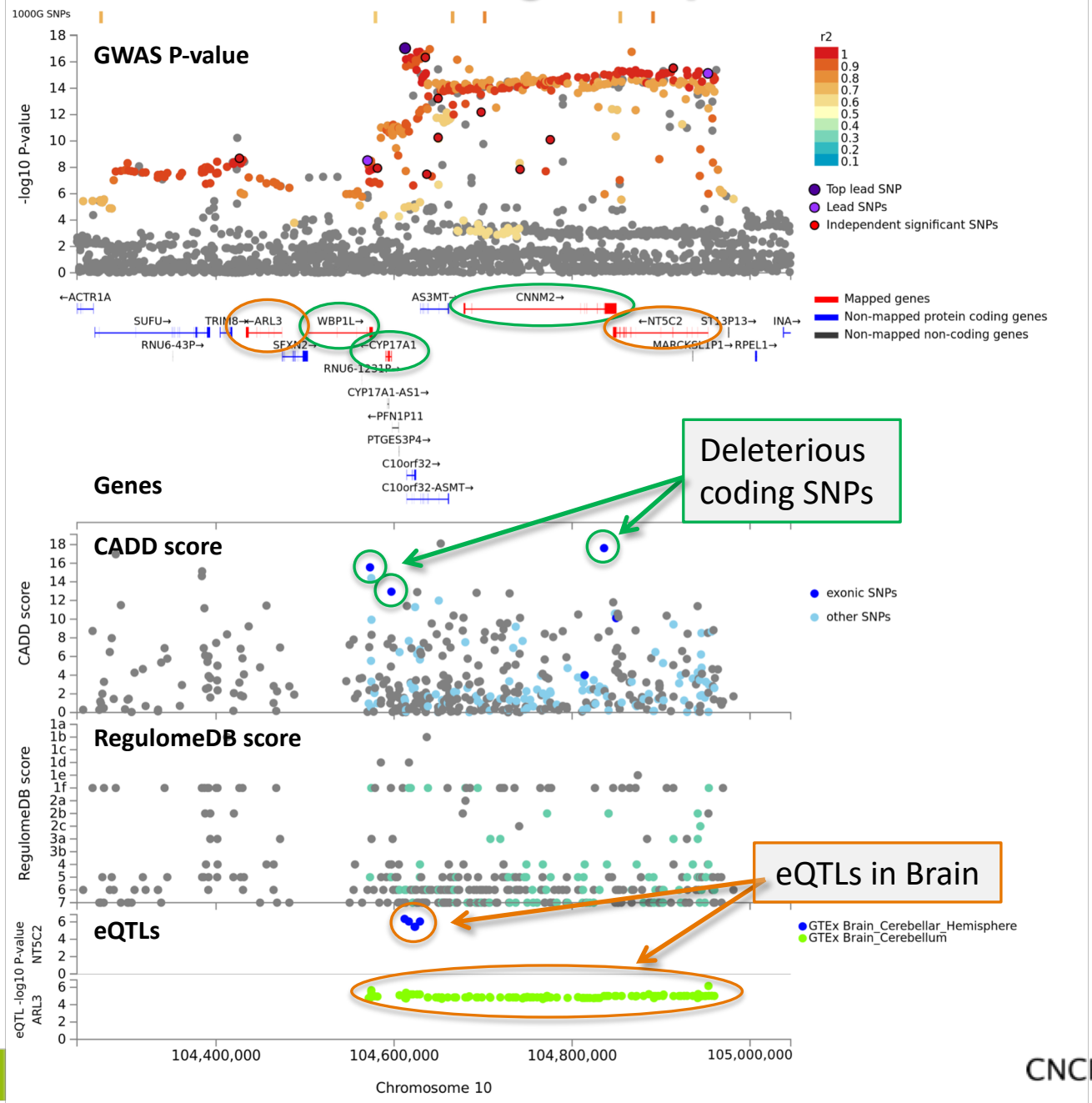
FUMA: regional plot



FUMA: regional plot

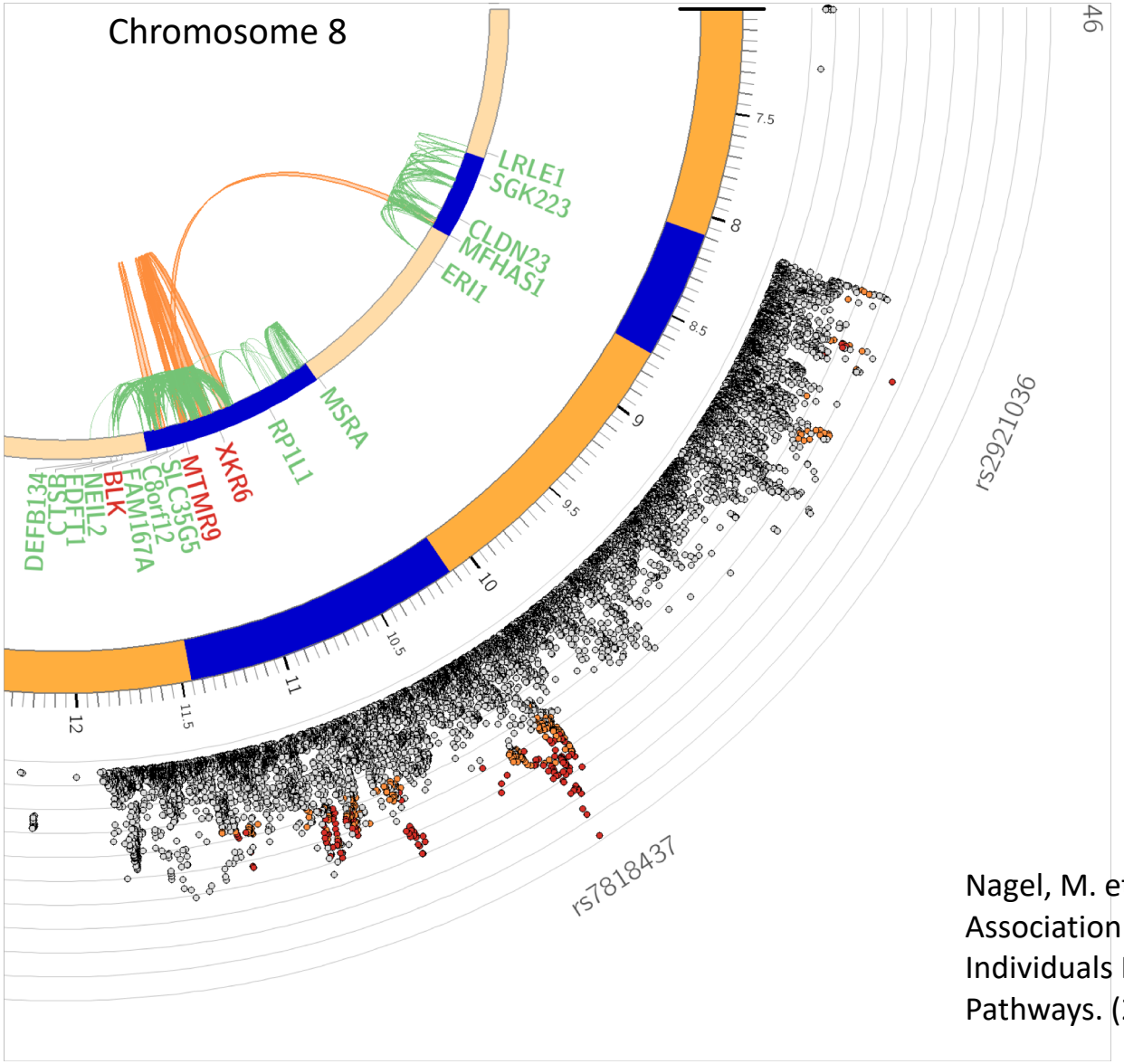


FUMA: regional plot



CNCR CTGLAB

FUMA: circos plot

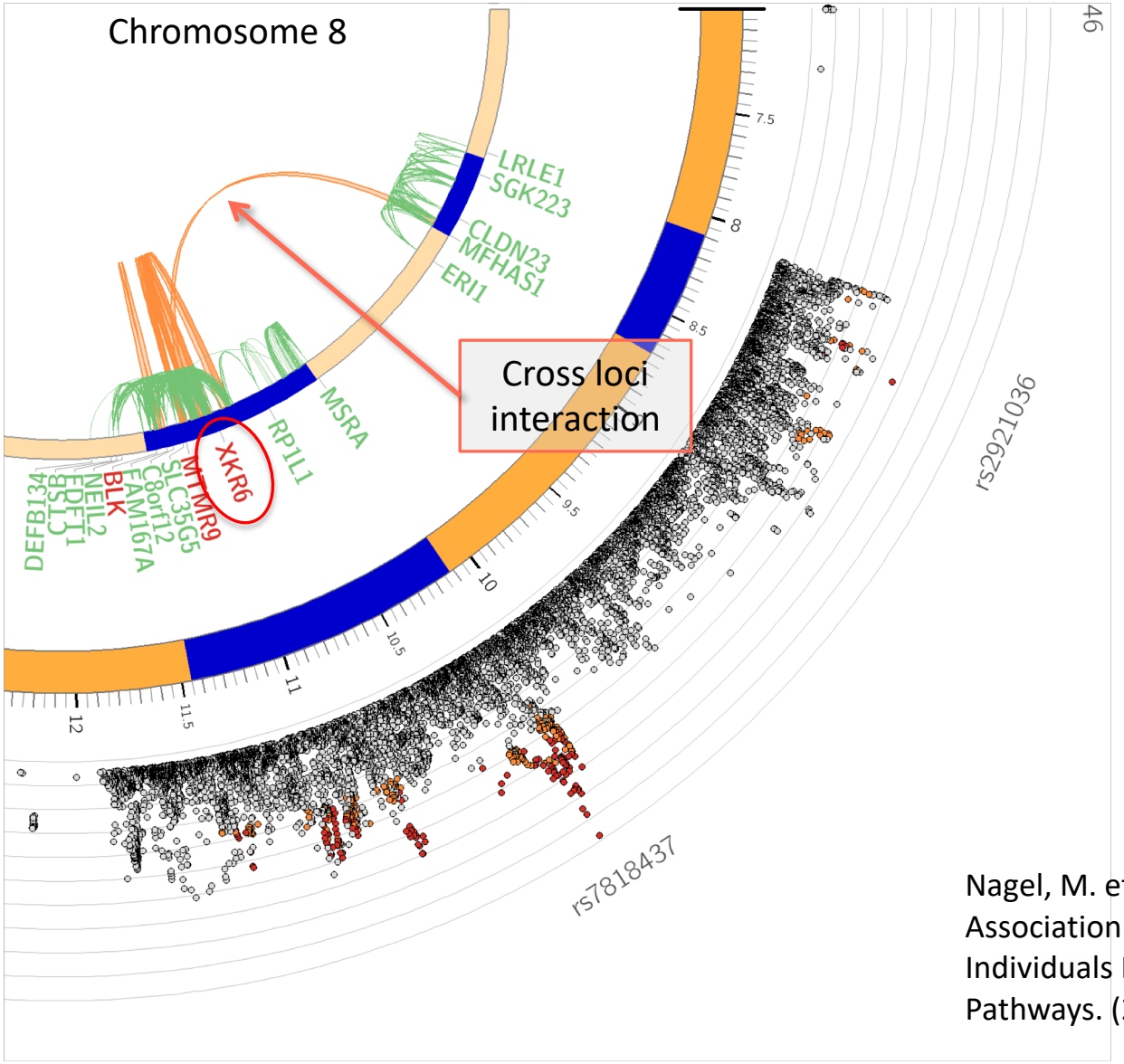


Chromatin interactions can map SNPs to distal genes.

Nagel, M. et al. Meta-Analysis of Genome-wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways. (2018) *Nat. Genet*.

CNCR CTGLAB

FUMA: circos plot



Chromatin interactions can map SNPs to distal genes.

XKR6 is mapped by within and cross loci interactions, also mapped by eQTLs.

Nagel, M. et al. Meta-Analysis of Genome-wide Association Studies for Neuroticism in 449,484 Individuals Identifies Novel Genetic Loci and Pathways. (2018) *Nat. Genet*.

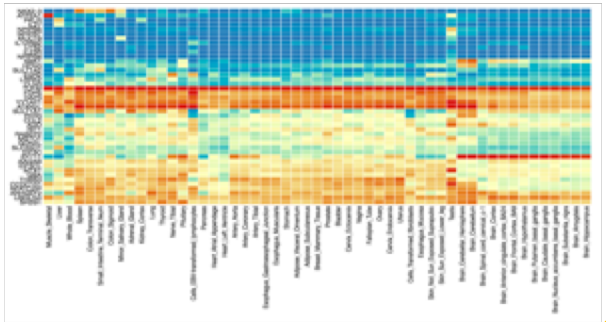
CNCR CTGLAB

SNP2GENE: FAQ

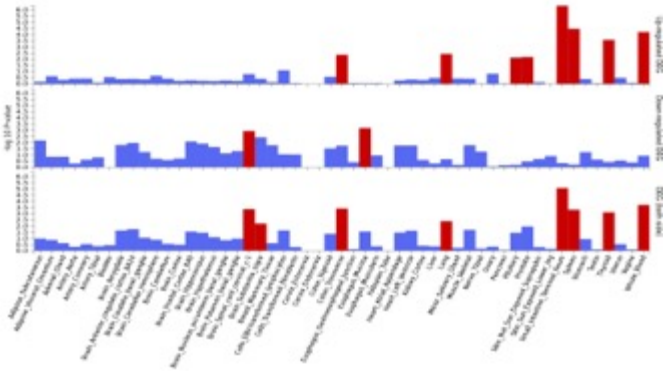
- ✓ FUMA provides annotations for SNPs which are in LD with one of the independent significant SNPs which can help you prioritize likely interesting (functional) SNPs and genes in a locus but does not pinpoint the actual causal SNPs/genes
- ✓ GWAS P-value is used only to identify indep. sig./lead SNPs and candidate SNPs but not used to weight/score SNPs or genes
- ✓ For eQTL mapping, significant eQTLs are annotated to SNPs in loci but they are not colocalized (meaning there might be overlap of SNPs and eQTLs due to LD)
- ✓ MAGMA process is independent from main SNP2GENE process, gene mapping does not use MAGMA gene-based results
- ✓ Input SNPs need to be present in the reference panel FUMA uses. If you find missing SNPs, it's most likely because those SNPs does not exist in the selected reference panel or mismatch of alleles

GENE2FUNC

Interactive heatmap of gene expression

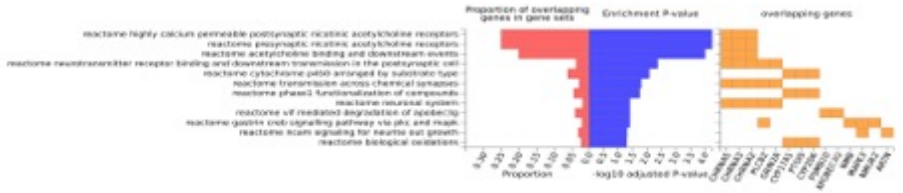


Tissue specificity (DEG)



Overrepresentation in Gene Sets

Curated gene sets
Motif gene sets
GO terms gene sets
(Total 20 categories)

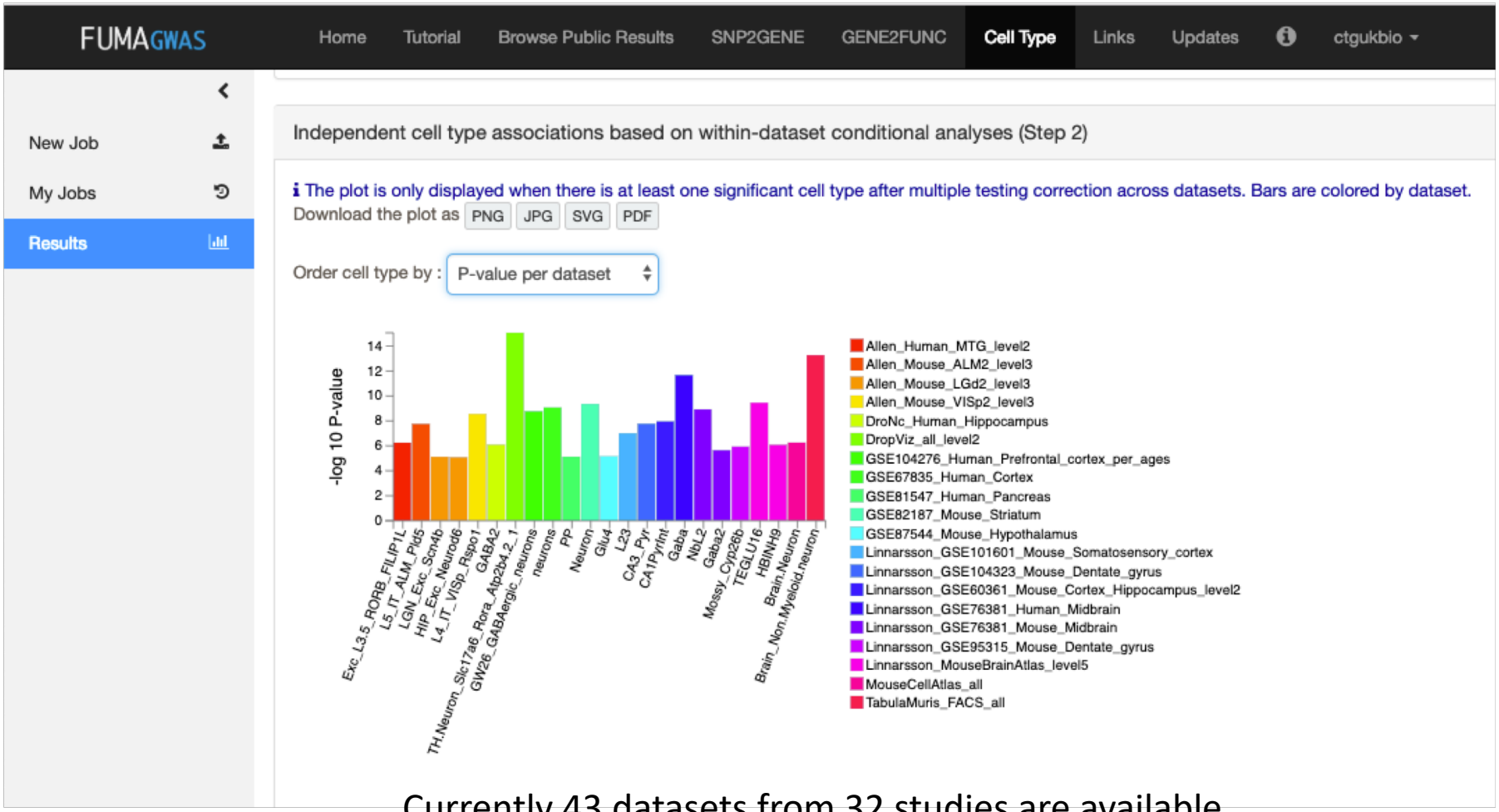


General biological functions of genes

OMIM (known disease associations), DrugBank (known targets of drugs),
GeneCards (general biological information)

Cell type specificity analysis

scRNA-seq dataset + GWAS summary statistics + MAGMA regression model → Association of specific cell type



Currently 43 datasets from 32 studies are available

CNCR CTGLAB

Technical information on FUMA

- ✓ You are required to register to submit a job
- ✓ Normally, SNP2GENE process takes 15-45 mins (depending on the number of hits and parameters, might take >1 hour)
- ✓ There are 5 parallel processes on FUMA server, your job might queue when the server is busy (average 45 jobs per day)
- ✓ Your input file(s) will be deleted from the server once the process is done
- ✓ When you get error and don't know how to solve
  Google group: https://groups.google.com/forum/#!forum/fuma-gwas-users
  Send me email: k.watanabe@vu.nl
  *Please do not delete your job when you contact about an error of a specific job!

Practical

Copy and open /faculty/kyoko/2019/FUMA/fuma_practical.pdf

The practical is based on the online web application. Please use **Chrome** if possible (some javascript functions do not work well with IE).

Section 1: FUMA (http://fuma.ctglab.nl)

- ✓ Explore FUMA web application with example GWAS results
- ✓ Detail look into SNP2GENE results
- ✓ Prioritize genes based on SNP annotations for an example locus

***You are more than welcome to submit job during the workshop but the queue is currently shutdown, it will re-start after the session.**

Section 2: GWAS ATLAS (http://atlas.ctglab.nl)

- ✓ Quick tour of the website
- ✓ Comparison of multiple GWASs
- ✓ Create PheWAS plot

CNCR CTGLAB