Training Exercise: Post-GWAS Analysis Using FUMA

KidneyGenAfrica Training Programme

Jean-Tristan Brandenburg

January 2026

Outline

What is Post-GWAS Analysis?

Post-GWAS analysis refers to analyses performed **after a genome-wide association study (GWAS)** to translate statistical associations into biological insight.

Goals:

- Interpret GWAS signals beyond single SNP associations
- Identify relevant variants, genes, and pathways
- Link genetic signals to biological mechanisms

Key Inputs for Post-GWAS

Post-GWAS analyses typically rely on:

- GWAS summary statistics
- Linkage disequilibrium (LD) information
- Genotype or reference panel data
- External functional and biological databases

Common Post-GWAS Analyses

- Quality assessment: λ, QQ plots, Manhattan plots
- Identification of significant SNPs and loci
- Fine-mapping to prioritize causal variants
- Functional annotation and visualization
- Replication of known associations

Gene-Level and Pathway Analyses

- Gene-based association tests
- Gene-set and pathway enrichment analyses
- Variant-to-gene mapping using:
  - eQTLs
  - Chromatin interaction data

More Advanced Analyses

- Polygenic Risk Scores (PRS)
- Mendelian Randomization
- Multi-omics integration:
  - Transcriptomics
  - Epigenomics
  - Proteomics

Technical Strategies for Post-GWAS

The choice of tools depends on:

- Research question
- Available data
- Computational resources

Manual Analysis

Advantages

- High flexibility
- Full transparency

Limitations

- Time-consuming
- Error-prone

Common Tools for Manual Analysis

- **PLINK** – filtering, clumping
- **GCTA** – conditional analysis, heritability
- **LocusZoom** – regional plots
- **FINEMAP, SuSiE, CAVIAR** – fine-mapping
- **ANNOVAR, VEP** – variant annotation
- **MAGMA** – gene and gene-set analysis
- **R / Python** – custom analyses

Pipelines and Workflows

- Automated and reproducible
- Scalable (HPC / cloud)
- Require setup and maintenance

Example:

- H3AGWAS pipeline

Web-Based Platforms

- Minimal computational burden
- Predefined workflows
- Reduced flexibility

Example:

- FUMA

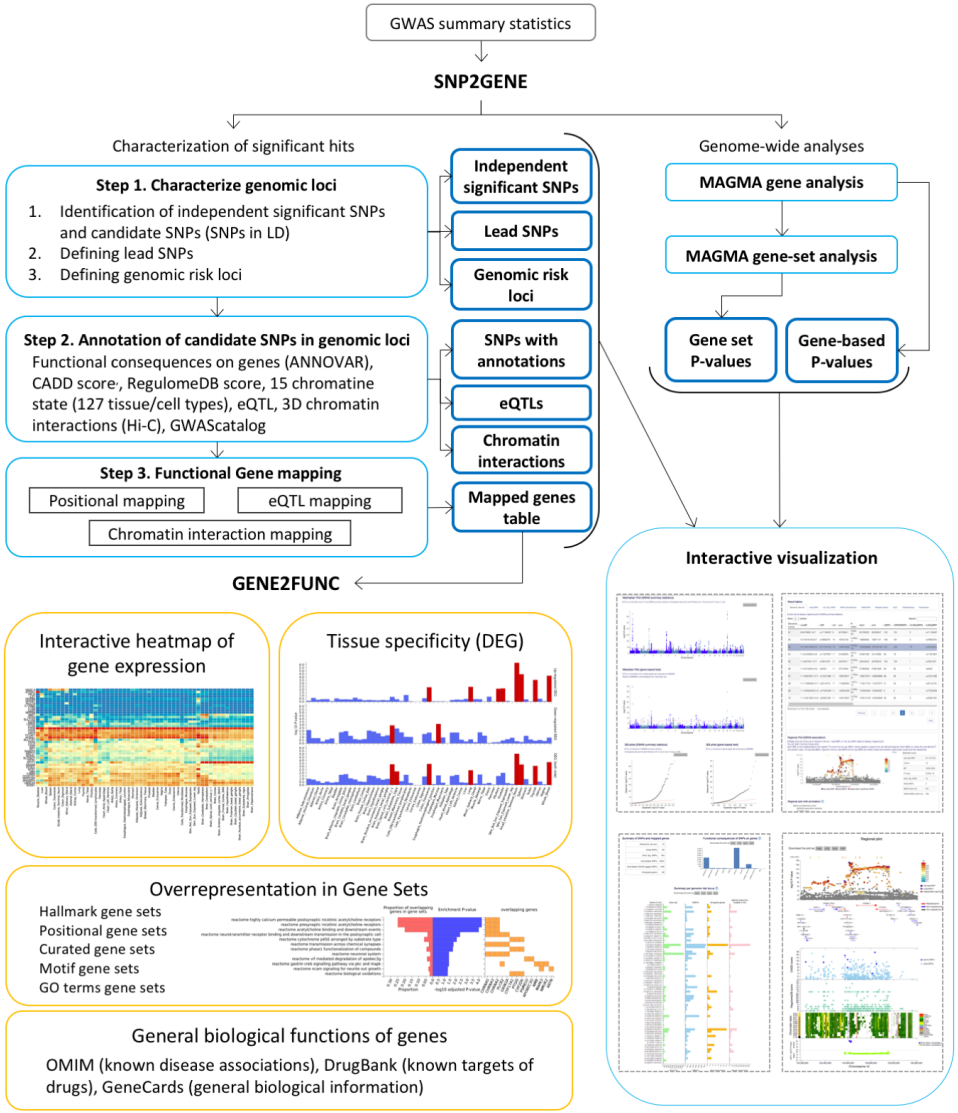
What is FUMA?

FUMA (Functional Mapping and Annotation of GWAS) is a web-based platform for:

- Functional annotation of GWAS results
- Gene prioritization
- Visualization and biological interpretation

Tutorial: `https://fuma.ctglab.nl/tutorial`

What is FUMA?


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
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)
Interactive visualization

FUMA Modules

- **SNP2GENE**
  - SNP selection
  - Locus definition
  - Functional annotation
  - MAGMA analysis
- **GENE2FUNC**
  - Expression and pathway analysis
- **Cell Type**
  - Cell-type enrichment using MAGMA

Independent Significant SNPs

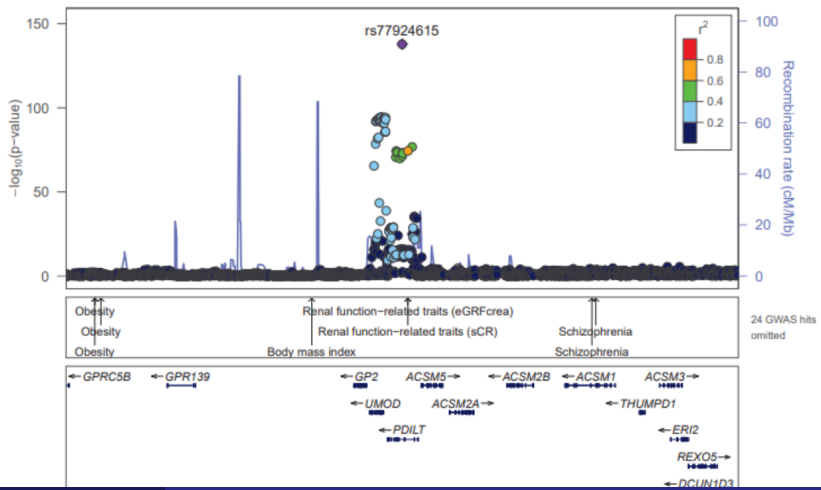
Independent significant SNPs:

- $P \leq 5 \times 10^{-8}$
- Independent at $r^2 \leq 0.6$

They serve as anchors for downstream analyses.

FUMA Lead SNPs and Putative Causal Variants Based on Clumping

- Each clump represents a putatively independent genomic locus defined by SNPs in linkage disequilibrium (LD).
- FUMA prioritizes putative causal SNPs using lead SNPs and SNPs in LD, combined with functional annotations, without performing statistical fine-mapping.
- This approach provides a simple, LD-based approximation of a credible set (less precise).
- The importance of the LD panel and its parameters used in plink (LD and max p-value)



Lead SNPs

Lead SNPs:

- Subset of independent significant SNPs
- Independent at a stricter threshold ($r^2 < 0.1$)
- Represent distinct association signals

Genomic Risk Loci

Genomic risk loci are defined by:

1. Grouping SNPs in LD ($r^2 \geq 0.1$)
2. Merging nearby loci (default 250 kb)
3. Selecting the SNP with the smallest P-value

Candidate SNPs

Candidate SNPs:

- In LD with independent significant SNPs
- Optional threshold: $P \leq 0.05$
- Fully annotated and used for gene mapping

Functional Annotation Outputs

- Annotated candidate SNPs
- ANNOVAR functional consequences
- Mapped genes
- eQTL mappings
- GWAS Catalog overlap

Gene-Based Analysis with MAGMA

- SNP-wise mean model
- SNPs assigned to genes by position
- Gene-level association statistics
- Gene-set and gene-property analyses

Visualization in FUMA

- SNP-level Manhattan plots
- Gene-based Manhattan plots
- Regional association plots

Alternative tool:

- LocusZoom v2: `http://locuszoom.org/`

Interpretation and Limitations

- LD-based loci are approximations
- Lead SNPs are not necessarily causal
- Multiple signals may be merged into one locus

Bayesian fine-mapping provides more precise causal inference.

Dataset Description

- ∼67,000 African American individuals
- Meta-analysis of 8 cohorts
- Generated using METAL

Data: `https://susztaklab.com/GWAS2M/Download.php`

FUMA Parameters Used

- Reference panel: 1000 Genomes (African populations)
- LD threshold: $r^2 > 0.6$
- MAF > 0.01
- eQTL mapping enabled
- MAGMA enabled

Results: `https://fuma.ctglab.nl/browse/697303`

Exercise Questions (1)

- How many genomic risk loci are identified?
- How are lead SNPs supported by neighboring variants?
- What differences do you observe across loci?

Exercise Questions (2)

- Which loci overlap with GWAS Catalog?
- Which loci appear novel?
- What are the limitations of GWAS Catalog overlap?
- Which locus and SNP showed the most significant association signal with eGFR?

Exercise Questions (3)

- Which SNPs are potentially pathogenic?
- Which genes are significant in MAGMA?
- How do gene-based and SNP-based results differ?

African-Ancestry Considerations

- Is 5×10^{-8} appropriate?
- Impact of higher diversity and shorter LD
- Trans-ethnic significance thresholds
- Alternative validation strategies