

FUMA Practical

8th Mar. 2019

Section 1 FUMA

Part 1. Getting started

- Go to <http://fuma.ctglab.nl>

In this practice, we will use publicly available results to explore what you can do with FUMA. If you are interested, you can create your own account and try submitting a job at the end of this section (optional). Most of the details about the web application are explained in the tutorial (<http://fuma.ctglab.nl/tutorial>).

- Click “Browse Public Results” tab at the top of the page

There are multiple public results available. In this practice, pick one trait from the below that you are most interested in. The reference articles are available under the “papers” directory under /faculty/kyoko/2019/FUMA.

ID	Trait	Reference
1	Body Mass Index (BMI)	Locke et al. <i>Nature</i> . 2015
2	Crohn’s disease (CD)	Franke et al. <i>Nature Genetics</i> . 2010
3	Schizophrenia (SCZ)	Ripke et al. <i>Nature</i> . 2014
36	Neuroticism (NEU)	Nagel et al. <i>Nature Genetics</i> . 2018
37	Intelligence (IQ)	Savage et al. <i>Nature Genetics</i> . 2018

- Click on the title of the trait you selected

This will open the results page. It might take some time to load all the data points (you may need some patience).

Part 2. Explore FUMA results

In this section, you will go through the SNP2GENE results to see what kind of information you can extract from FUMA.

1) Go to the “Genome wide plot” tab from the left side bar and inspect the Manhattan and QQ plots. What is the difference between the two Manhattan plots and why is the threshold of significance different?

2) Go to the “Summary of results” tab from the left side bar.

- Is the number of risk loci the same as reported in the original paper? If not, can you explain why?
- Summarize the difference between lead SNPs, Independent significant SNPs and candidate SNPs. (If you can’t remember, have a look at <http://fuma.ctglab.nl/tutorial#riskloci>.)

3) From the bar plot for functional consequences of SNPs, find out how many exonic SNPs there are amongst the risk loci (hover your mouse over the bar plot).

4) At the bottom of the page, there is a bar plot summarizing each risk locus.

- Which locus is the largest?
- Which locus contains the largest number of SNPs and how many SNPs are in this locus?
- Can you explain why the number of mapped genes is not always the same as the number of genes that are physically located in a locus?

➤ Go to the “Results” tab from the left side bar

5) The first displayed tab under ‘Result tables’ is for ‘genomic risk loci’. Find the most significantly associated locus. By clicking the row, a locus plot will be displayed at the bottom of the page.

- What is the P-value and MAF of the most significant SNP?
- How many lead SNPs and independent significant SNPs are in this locus?
- How many SNPs are in LD with the most significant SNP?

➤ Go back to the top of the page and open the “SNPs (annotations)” tab.

6) The SNPs table contains all SNPs that are in LD with one of the independent significant SNPs, and their functional annotations.

- Which SNP has the highest CADD score, and what is the score for that SNP?
- What is the functional consequence of that SNP? If it is exonic, can you find more detailed annotation information (hint: check ANNOVAR table)?
- Which gene is most likely affected by that most deleterious SNP?

➤ Now open the “Mapped genes” table

7) This table contains all mapped genes with detailed information on how genes are mapped by FUMA.

- Is the gene that is mapped by the most deleterious SNP (from question 6) also mapped by other mapping strategies? If so, describe how the gene was mapped.

- Find a gene mapped by the most significant SNP and describe how the gene was mapped. (Sort by column “minGwasP”, if there are “NA” skip several pages to find a gene with P-value in this column.) If there are multiple genes with the same GWAS P-value, pick one.

Now let’s create a regional plot with annotations.

- Depending on the trait you chose, select the following locus from the “Genomic risk loci” table (make sure that the regional plot has been updated from the last one).

BMI: locus 48, CD: locus 15, SCZ: locus 71, NEU: locus 93, IQ: locus 14

- Scroll down to the regional plot and click “Plot” below the plot.

This will open a new tab and you will see a regional plot together with annotations

8) First, check the legends and descriptions of the plots.

- How many exonic SNPs are there?
- Which genes is affected by the most deleterious SNPs in the locus?
- Which genes are affected by eQTLs in this locus?
- Which SNP has the highest RegulomeDB score (if there are multiple, pick one) and which gene is most likely affected by this SNP (or not affecting any gene)?
- If you are asked by a biologist which genes to follow up from this locus, what would you recommend and why?

Part 3. Annotate your own GWAS (optional)

If you have time to spare (or if you are interested), you can try to submit a SNP2GENE job for a trait of interest.

1) First let’s find some GWAS summary statistics. There is a web-based database for GWAS summary statistics (GWAS ATLAS) where you can search publicly available GWAS.

- Go to <http://atlas.ctglab.nl>
- Go to “Browse GWAS” tab at the top of the website

From the search box, try to find a trait you are interested in. You can also filter by trait Domains. If there are multiple GWAS for your trait, sort by N (from large to small) and select the one with the largest sample size.

If you find one, click the trait on the table, which will open a new tab with the summary of the GWAS results. From a table at the left side of the page, you can find a link to the original GWAS summary statistics. Download the file.

2) Before submitting to FUMA, check the header of the summary statistics. If the header is not automatically detected, you need to manually provide column names. See <http://fuma.ctglab.nl/tutorial#prepare-input-files> for more details. To check header, you can use UNIX command line:

➤ `gzip -cd <file name>.gz | head`

3) Now you are ready to submit!! There are lots of parameters you can set. See <http://fuma.ctglab.nl/tutorial#parameters> for more details, or use the default for now. Once you decide the parameters, hit the “Submit” button. You will need to have created an account to do this.

The SNP2GENE job normally takes between 15 mins to 1 hour depending on the number of significant hits. When the server is busy, your job will be queued. In this case, you might have to wait for a while.

4) Once the job is done, you will receive an email from FUMA. You can login to your account and access your new results from the ‘job list’. The job might not complete within the practical, in that case skip this section and whenever you have time, explore your results. If you have any questions, check tutorials (<http://fuma.ctglab.nl/tutorial>) or the FUMA paper (<https://www-nature-com.vu-nl.idm.oclc.org/articles/s41467-017-01261-5>). There is also a Google Forum at <https://groups.google.com/forum/#!forum/fuma-gwas-users>. And you are always welcome to contact me (k.watanabe@vu.nl)!!

Section 2 GWAS ATLAS (optional)

In this practical, you will have a tour of the GWAS ATLAS website.

- Go to <http://atlas.ctglab.nl>

The atlas contains summaries of GWAS results and includes several analyses such as MAGMA gene-based and gene-set analyses, SNP heritability and genetic correlation with LD score regression. Details can be found in the documentation (<http://atlas.ctglab.nl/documentation>). If you have any question, please check the FAQ (<http://atlas.ctglab.nl/faq>).

Part 1. Explore the database

- Go to “Browse GWAS” tab from the top of the page.

1) Search a trait you are interested in from the search box at the right top of the table. You can also filter GWASs by domain/sample size/published year. Then click on the row of your trait in the table. This will open a new tab with summary results of the GWAS. If you don't have a particular trait you are interested in, let's type “intelligence” and order by sample size (click column “N” twice so it will be descending order). Take the GWAS with the largest sample size as an example.

- What is the domain of the trait?
- What is the number of associated loci?
- What is SNP heritability estimate?
- Is there any significantly associated gene set?
- Which trait has the highest (significant) genetic correlation (excluding GWASs of the same trait)?

Part 2. Compare multiple GWAS

- Go to “Multiple GWAS comparison” tab from the top of the page.

2) Let's select 5-20 GWASs that you want to compare. You can search GWAS by trait name or filter by trait domain/sample size/published year. Check the box of the GWAS in the table.

If you don't know what to select, let's use an example selection; limit the published year from 2017 and minimum sample size 100000 (hit enter to update the table). On the first page (10 GWAS), click the button “Select all displayed GWAS”, then go to the next page and click the button “Select all displayed GWAS” again. This will result in selecting 20 GWAS. Once you select the GWASs, click “Compare selected GWAS” at the bottom of the table.

- Which trait has the highest SNP heritability?
- Can you find a cluster of traits with strong genetic correlation?
- Which genomic locus is associated with the largest number of GWAS? By default, the plot is based on the number of associated domains, please change the Y-axis option to “number of associated GWAS”.
- Which genes is associated with the largest number of GWAS?
- Which GWAS shows the most significant association with that gene?

Part 3. Create PheWAS plot

- Go to “PheWAS” tab from the top of the page.

3) Create a PheWAS plot of a gene or SNP that you are interested in. For a gene, you can provide either the gene symbol or a gene ID. For a SNP, you can provide either the rsID or the SNP coordinate (chr:position on hg19). If you don't have any favorite gene or SNP, let's use “DRD2” gene as an example.

- Which GWAS shows the most significant association?
- Are there significant associations with GWAS from multiple domains?