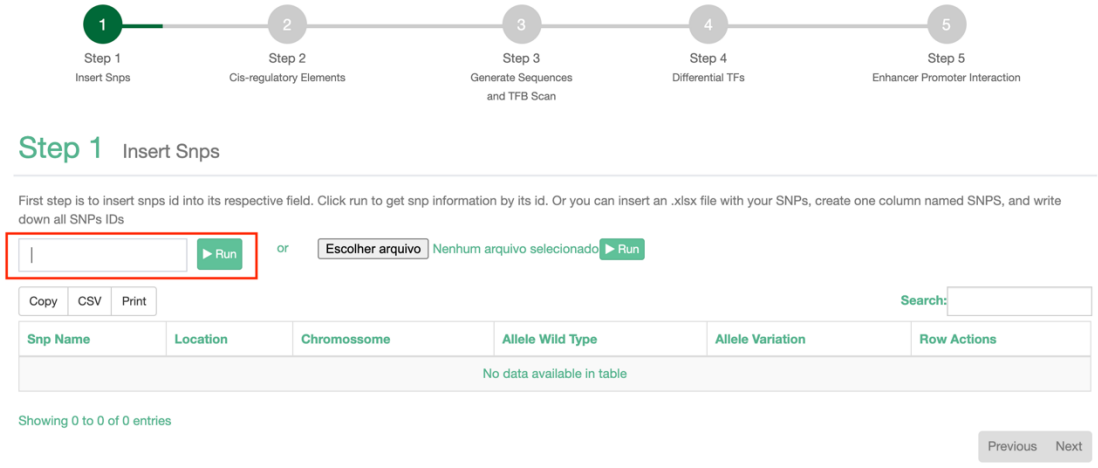
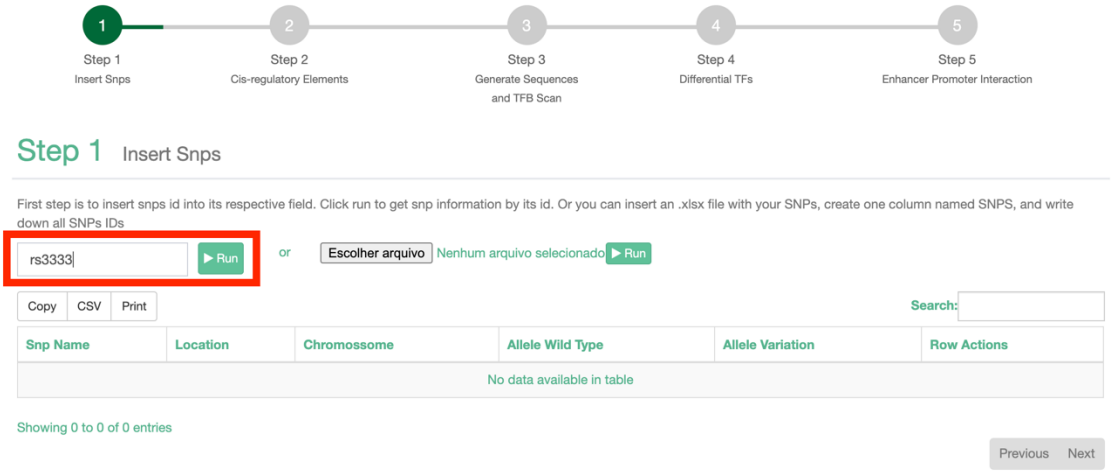
REGULOMIX TUTORIAL

After you've logged in, you will see steps that were designed to ease the process of using REGULOMIX. To use REGULOMIX, you'll have to follow each step below.

Step 1 – Insert SNPs:



Insert your SNPs manually at the blank text field inside the red box. After that, press the green "Run" button.

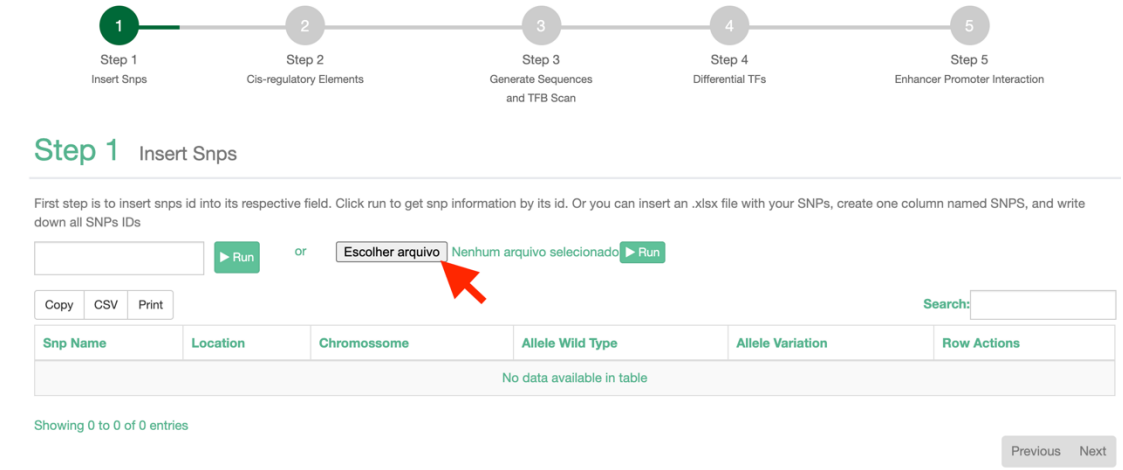


OR

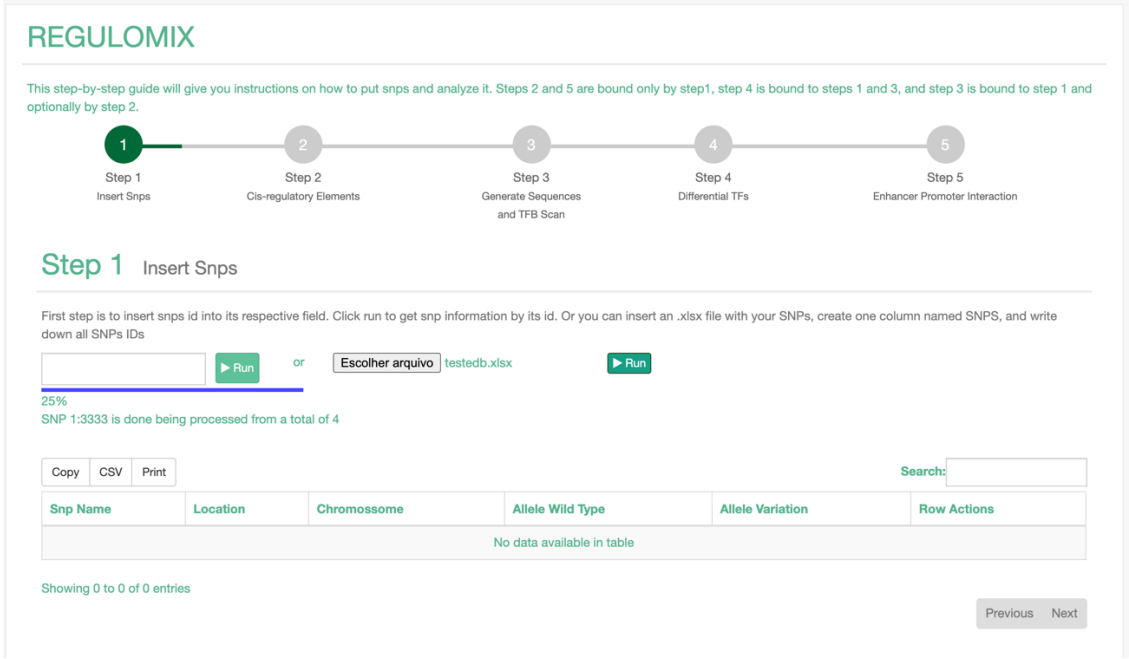
You can create a table in .xlsx to use you list of SNPs faster. To do that, you'll have to follow the example below.

	A	B	C	D	E	F
1	SNPS					
2	rs3333					
3	rs356168					
4						
5						
6						
7						
8						
9						
10						

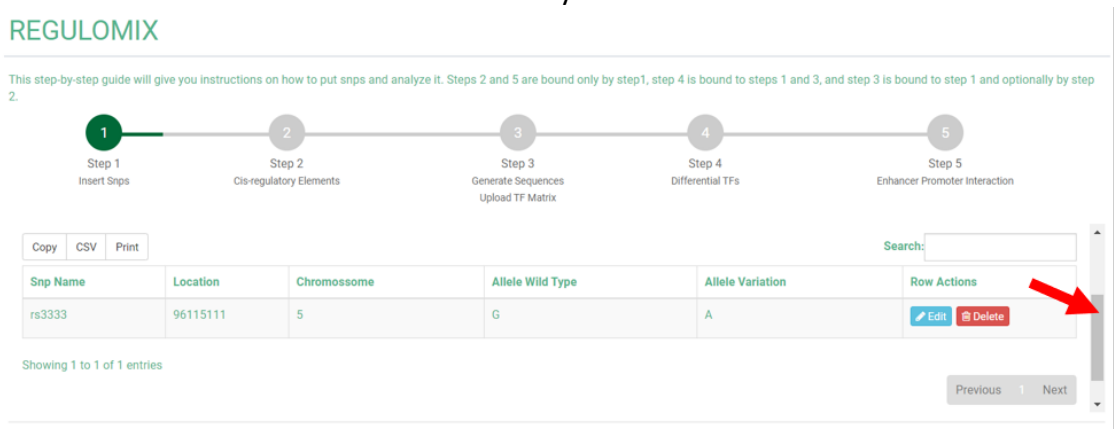
Example of an excel sheet containing SNP IDs. For the algorithm to work, create a file in .xlsx with only one column with header named SNPS and write in each row one SNP ID.



Choose your SNP chart file to be uploaded by clicking the "Choose File" button and select you own .xlsx file. After that, press the green “Run” button.



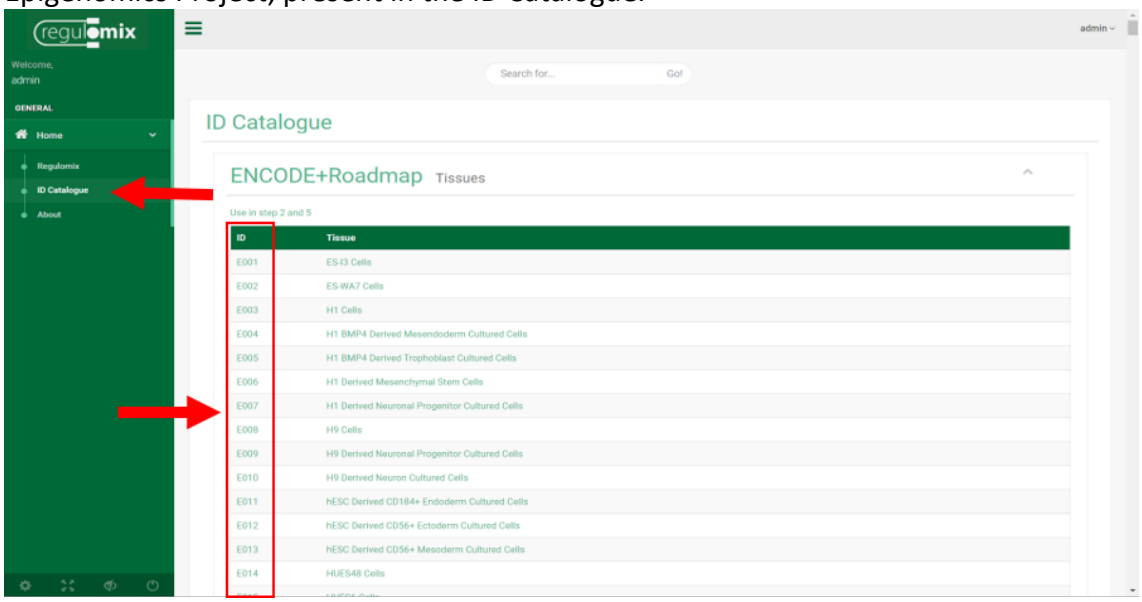
A progress bar will appear showing the percentage of SNPs that already had been processed, the SNP id of the SNP that had just been processed and its index.
The progress bar will not be shown if the user close the browser tab, but the progress will continue and the user will be notified by email when it's done.



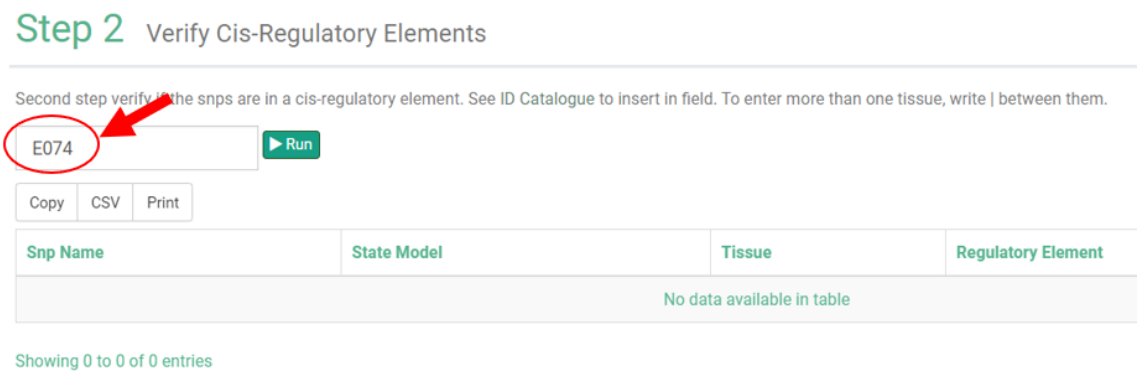
Use the Scroll Down Bar (red arrow) to see the information about the SNP inserted at the chart below.

Step 2 – Cis-regulatory Elements:

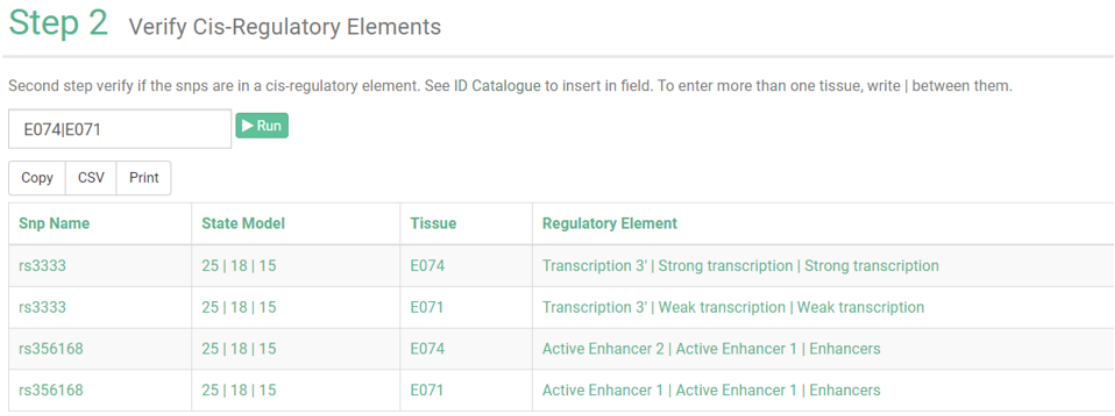
REGULOMIX uses, in STEP 2, ROADMAP Epigenomics Project data and servers. To verify Cis-regulatory elements, you'll have to enter the ID of the tissue from ROADMAP Epigenomics Project, present in the ID Catalogue.



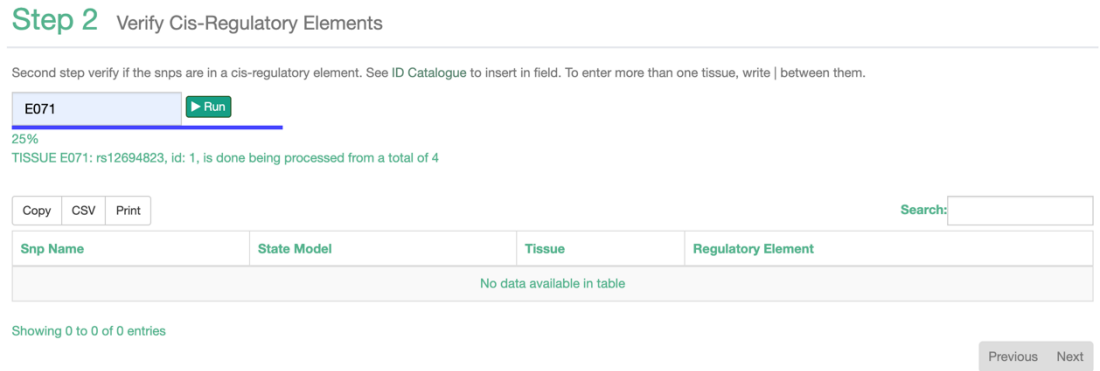
Tissues ID are in the ID Catalogue (1). To access the ID Catalogue (1), click on "ID Catalogue" at the left green panel. The tissues ID (2) will be at the first column.



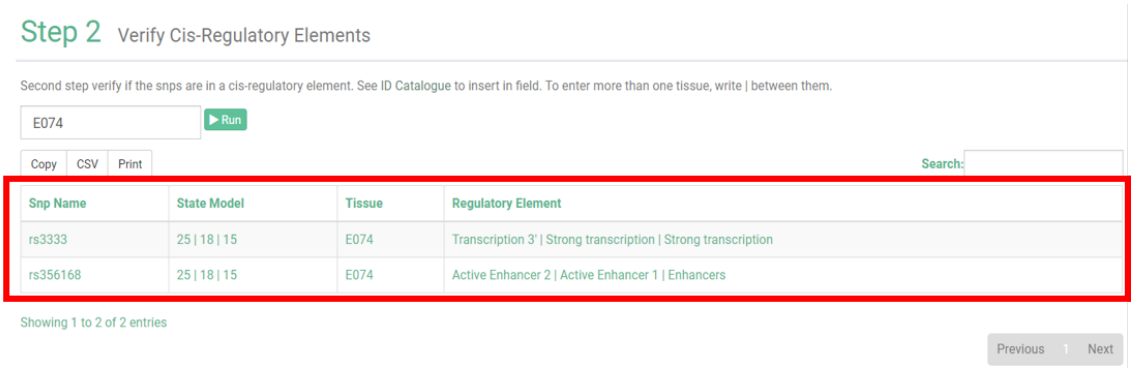
At Step 2, the first thing you'll need to do is enter the tissue ID at the textbox. You can enter multiple tissues by using the symbol "|" between them!



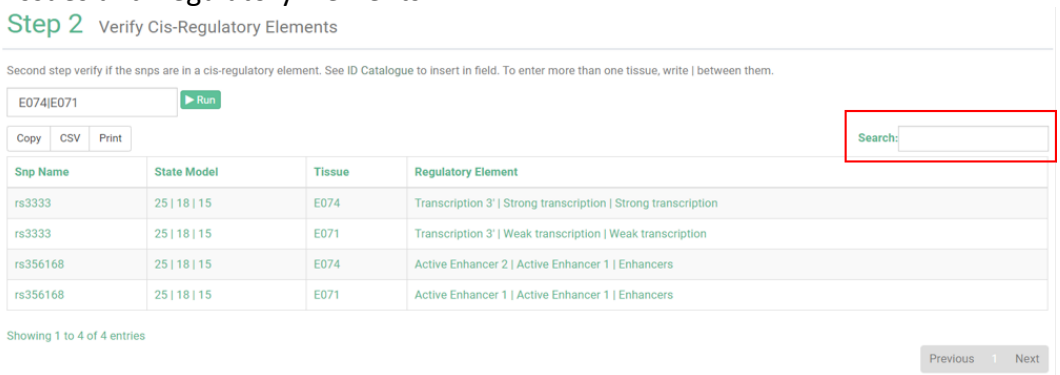
After entering the tissue ID, press the green "Run" button to the right of the tissue ID textbox. It will appear that it is processing your request with the same progress bar as the step 1. It is going to show the tissue ID, the SNP that just finished being processed and its index.

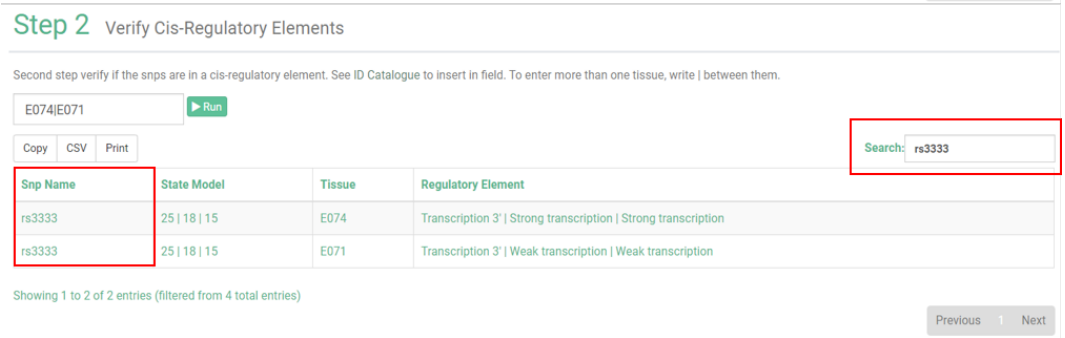


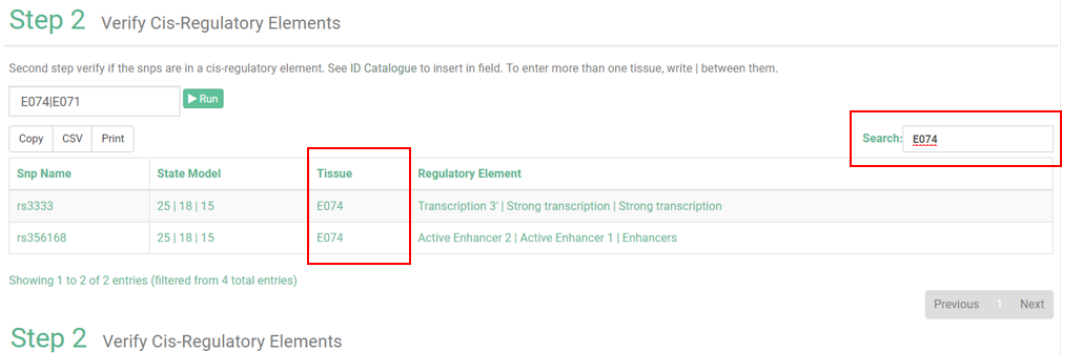
After REGULOMIX process your SNPs inside tissue data from ROADMAP Epigenomics, the information will be displayed at the table below.

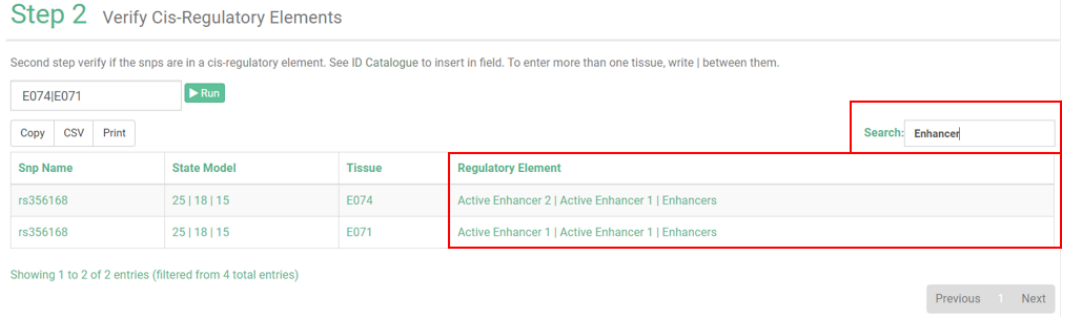


You can use the “Search” textbox to look for any value in the table such as: SNPs, Tissues and Regulatory Elements.

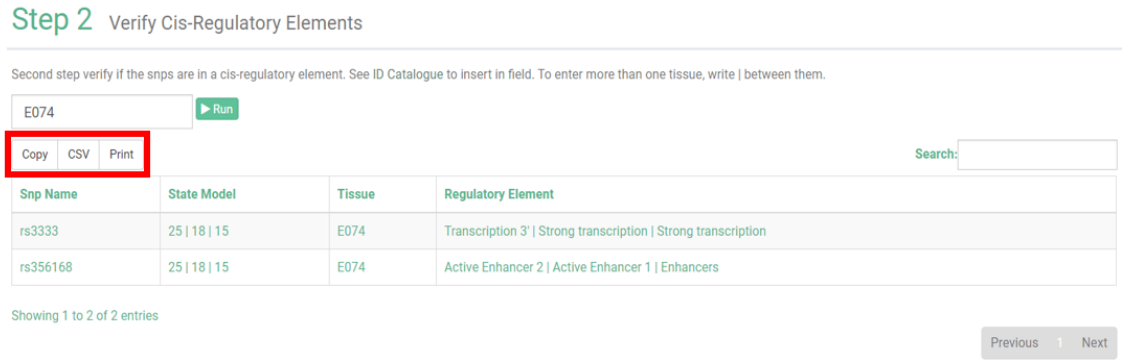




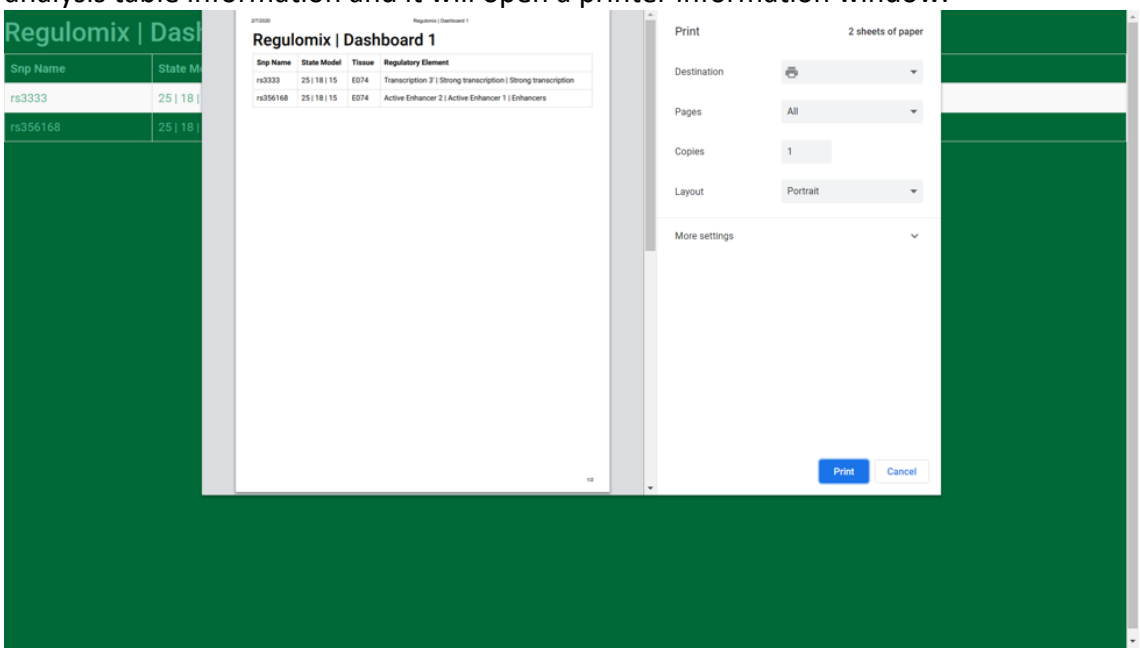




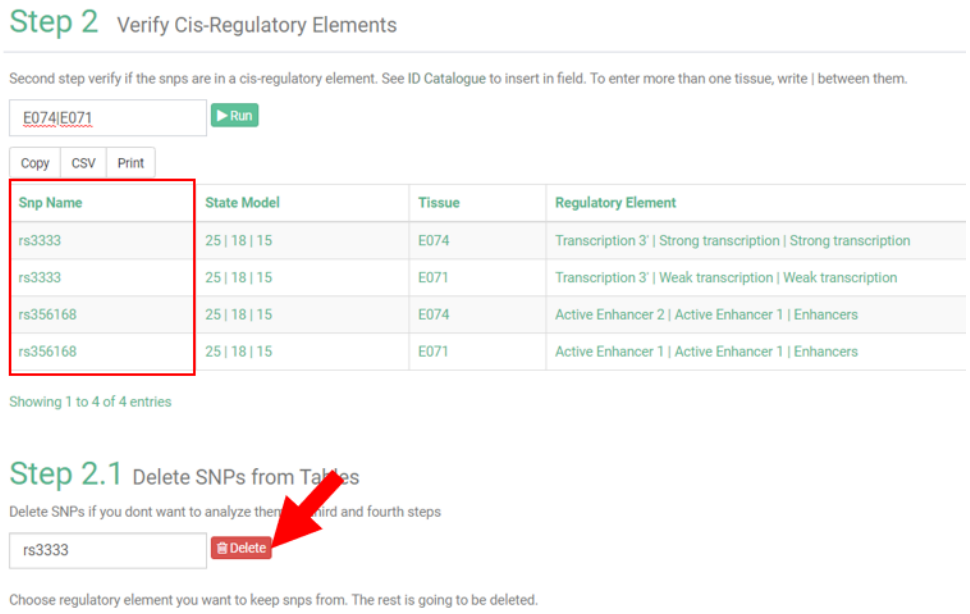
It is possible to “Copy”, “Print” or download a “.csv” file of your data from STEP 2 using the buttons below “Tissue” text field.</p>



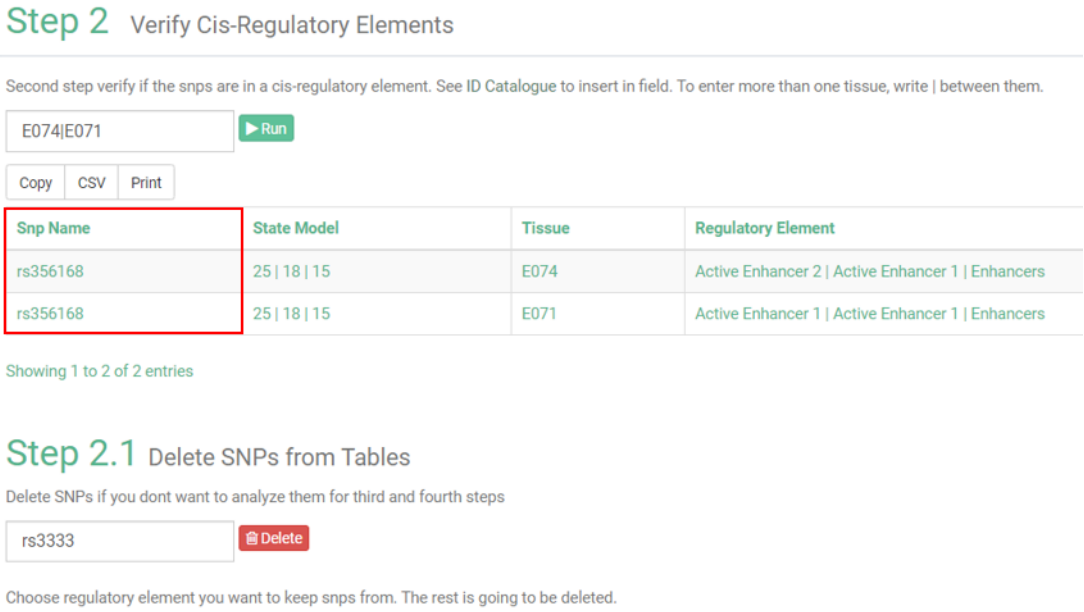
Pressing “Copy” will copy the Cis-regulatory Elements analysis data to your Clipboard. You can paste it in whatever program you want to save the information.
Pressing “CSV” will download a .csv file with the information of the Cis-regulatory Elements analysis.
Pressing “Print” will open a new tab with only the Cis-regulatory Elements analysis table information and it will open a printer information window.



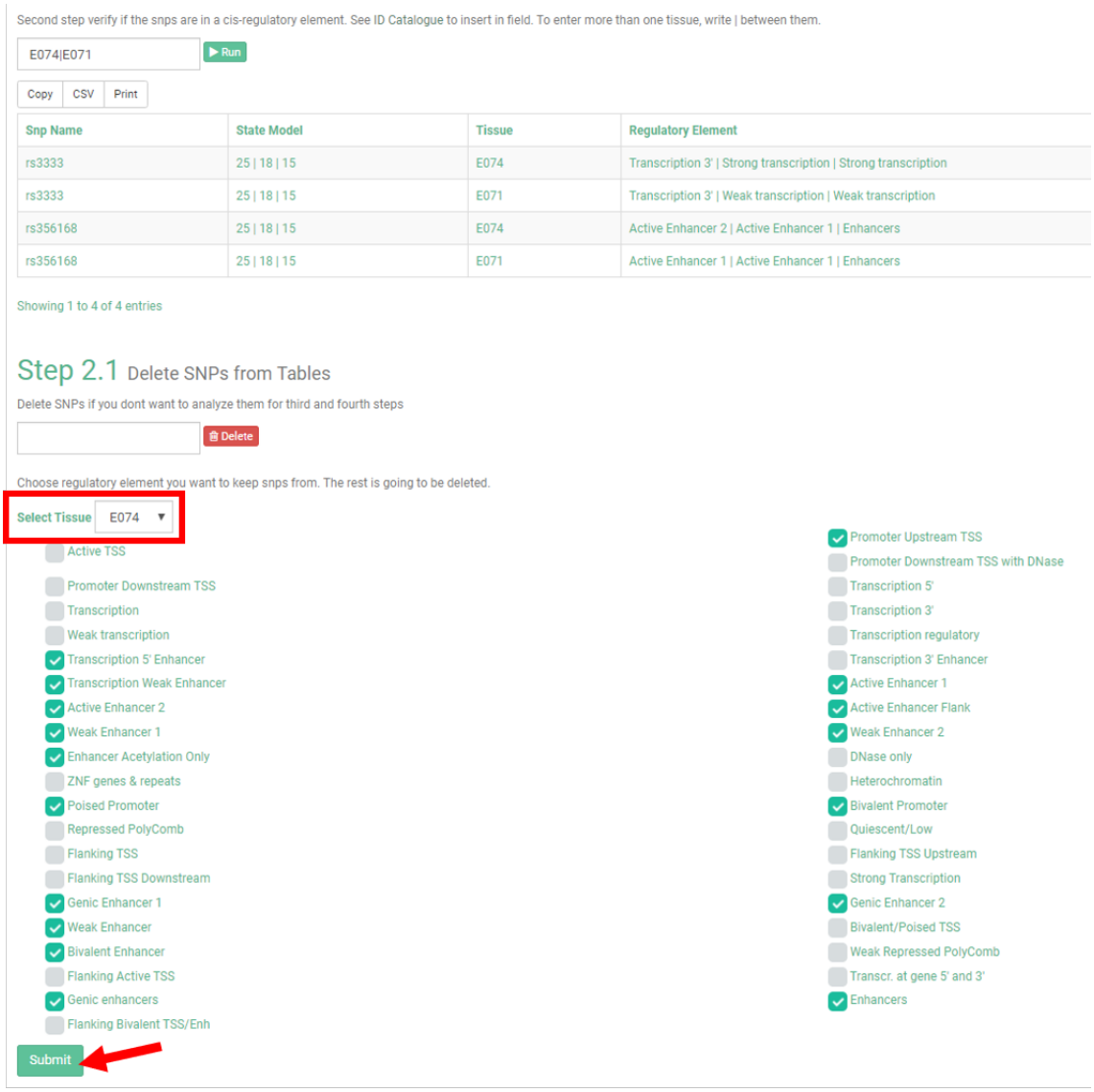
If you want to exclude SNPs from next steps, you can write the name of that SNP inside the Delete SNPs textbox and press “Delete”.



After pressing delete, you'll notice that the SNP will be deleted from the table above.



If you want to continue to STEP 3, you should select one of the tissues analyzed and any of the respectively regulatory element you want, then press "Submit".



Step 3 – Generate Sequences and TFB Scan:

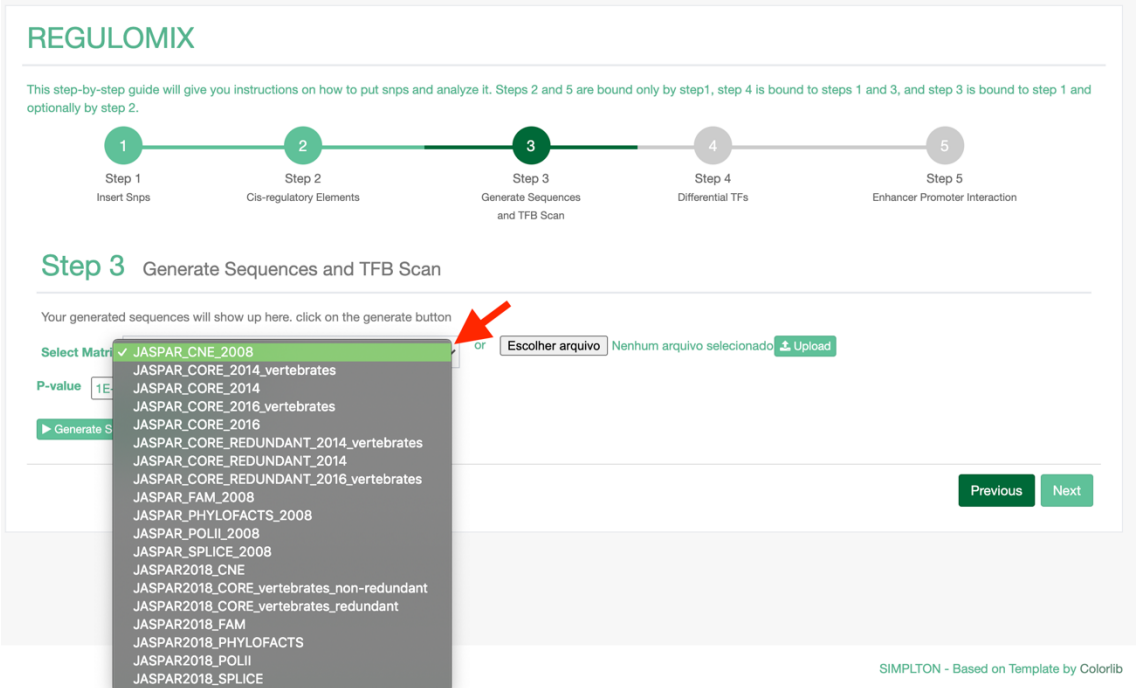
In Step 3, REGULOMIX is going to create sequences of 63 nucleotides long with each SNP filtered from Step 2 in different positions within the sequence. These sequences will be used with JASPAR matrixes to identify which Transcription Factors to bind in each SNP filtered.

To select one of the matrixes from JASPAR already in our server, click on the drop-down list box to choose. Matrix information can be found at Jaspar website, inside "About" section: http://jaspar.genereg.net/about/

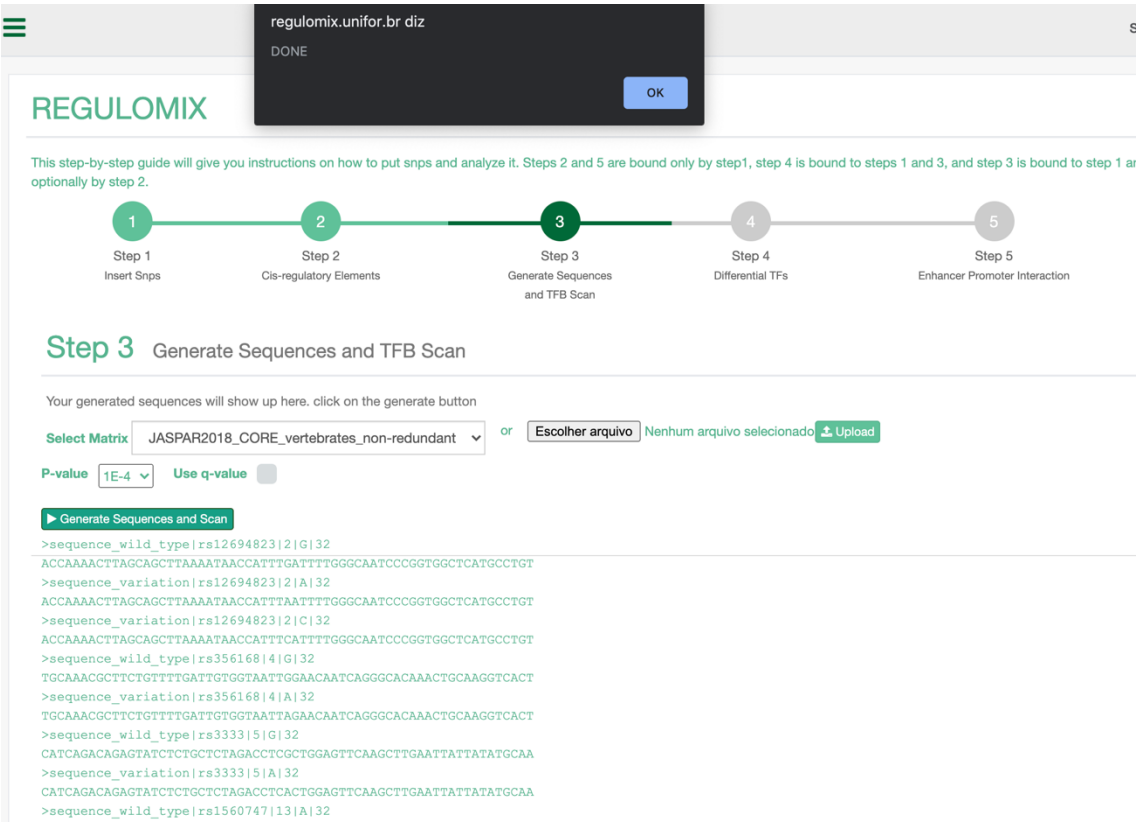
You can also upload matrixes in meme format to be used in REGULOMIX, that uses FIMO in the background, by clicking "Choose File". After you choose your file, you should click "Upload".

You can select a different value for p-value parameter in FIMO or you can select q-value to use "q-value" as FIMO threshold parameter

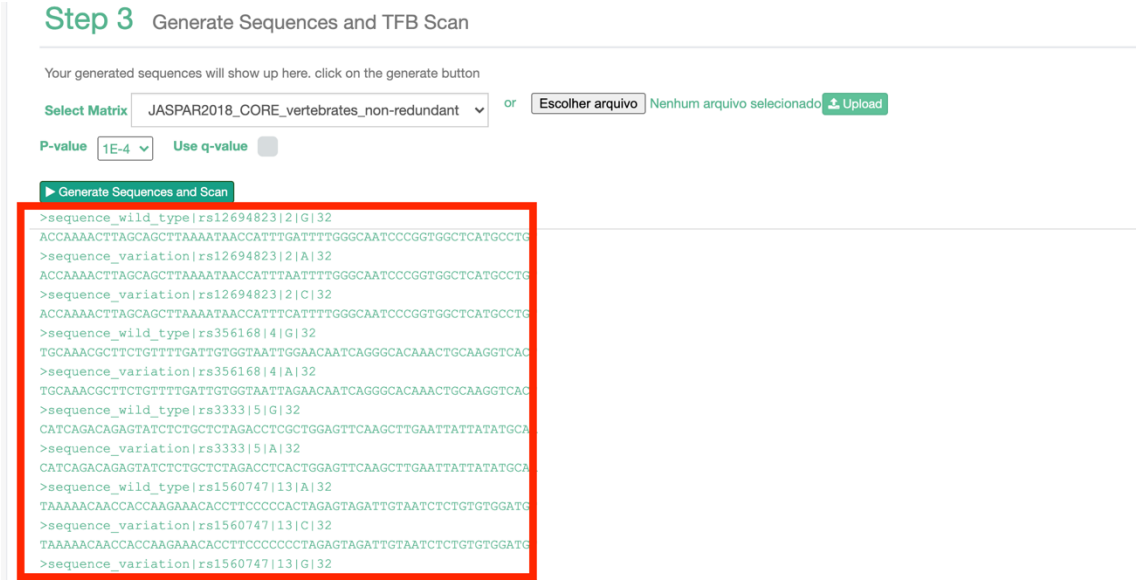
Click “Generate and Scan” to request to Generate Sequences and process sequences in FIMO.



When the generation of sequences is over, it will appear a popup message saying it’s done.



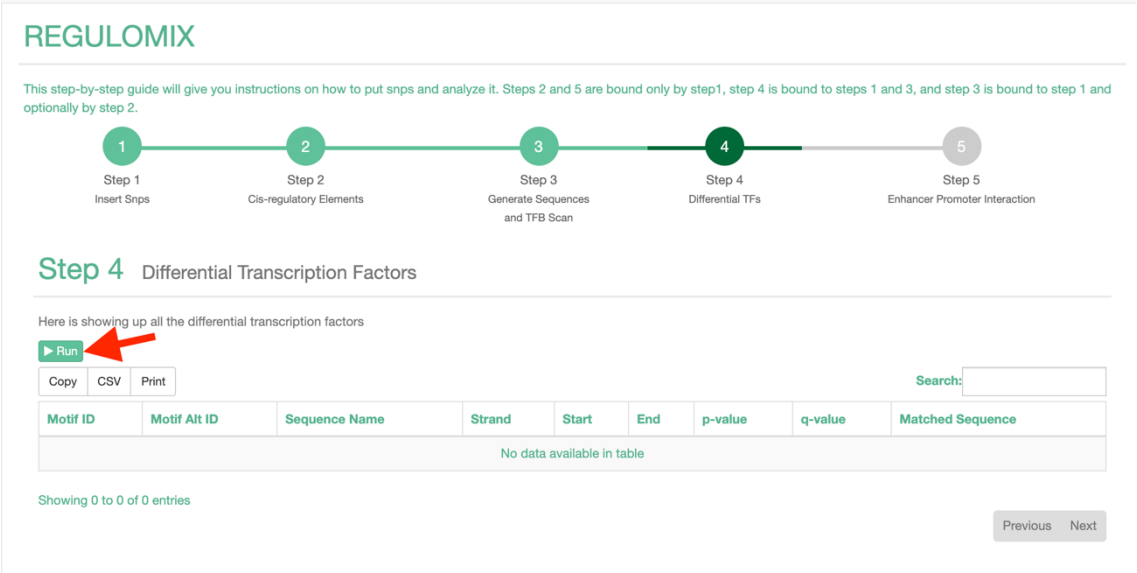
You will able to see the sequences generate below the “Generate and Scan” Button.



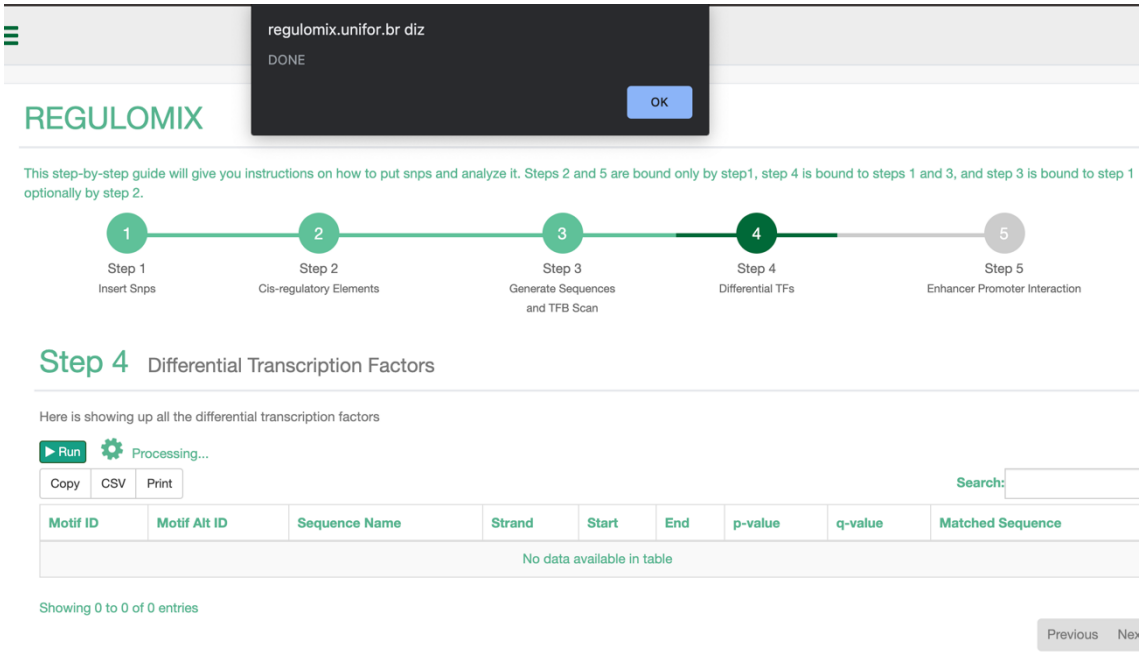
You can press “Next” to continue to Step 4.

Step 4 – Differential Transcription Factors:

In Step 4, you’ll be able to request analyses of which Transcription Factors bind to each sequence generated in Step 3. Therefore, it’s possible to manually search for Differential Transcription Factors.
To request the Differential Transcription Factors analyses, press the green “Run” button.



When the Differential Transcription Factor Analyses is over, it will appear a popup message saying it’s done. After the analysis is over, it will appear a table below the “Run” button to visualize each TF binding for each sequence.



The meaning of each column in step 4:

Motif ID: Transcription Factor/Motif ID.

Motif Alt ID: Name of the Transcription Factor.

Sequence Name:

This column is divided in 5 parts:

- Part 1: Sequence type: It can be “sequence_wild_type” for the sequence that contains a Wild-type SNP allele, or “sequence_variation” for the sequence that contains a non wild-type/variation SNP.
- Part 2: SNP ID: The RefSNP number.
- Part 3: SNP chromosome.
- Part 4: SNP Allele: The Allele of the SNP contained in the sequence that the TF binds to.
- Part 5: Position of the SNP inside the sequence in nucleotide.

Strand: Show the Positive-sense (5’-to-3’) or the Negative sense (3’-to-5’) of the sequence.

Start: The beginning of where the transcription factor binds in the sequence, in nucleotide.

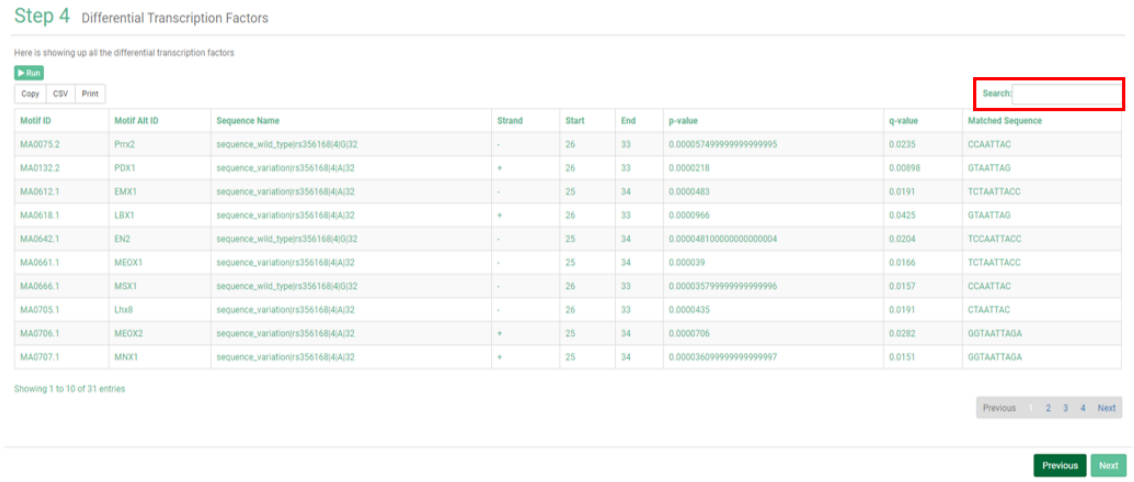
End: The end of where the transcription factor binds in the sequence, in nucleotide.

p-value: Shows the p-value for the binding.

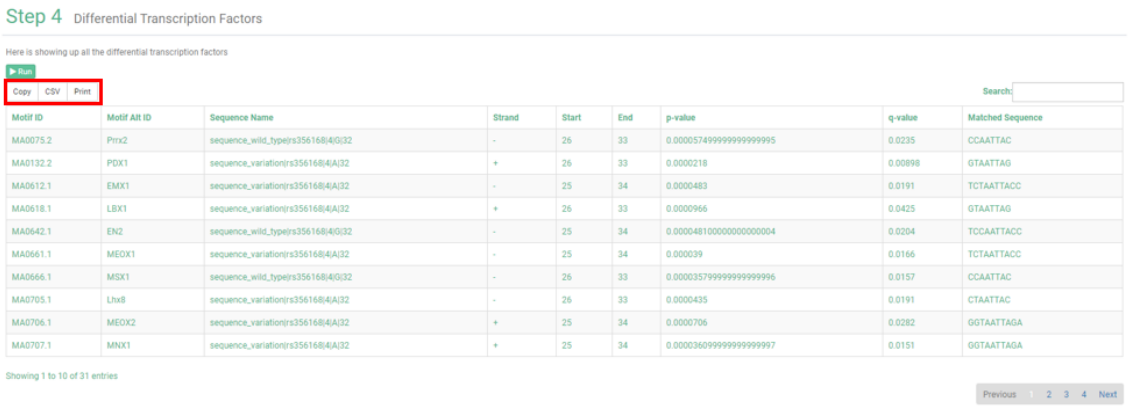
q-value: Shows the q-value for the binding.

Matched Sequence: Shows what part of the sequence is binding to the Transcription Factor.

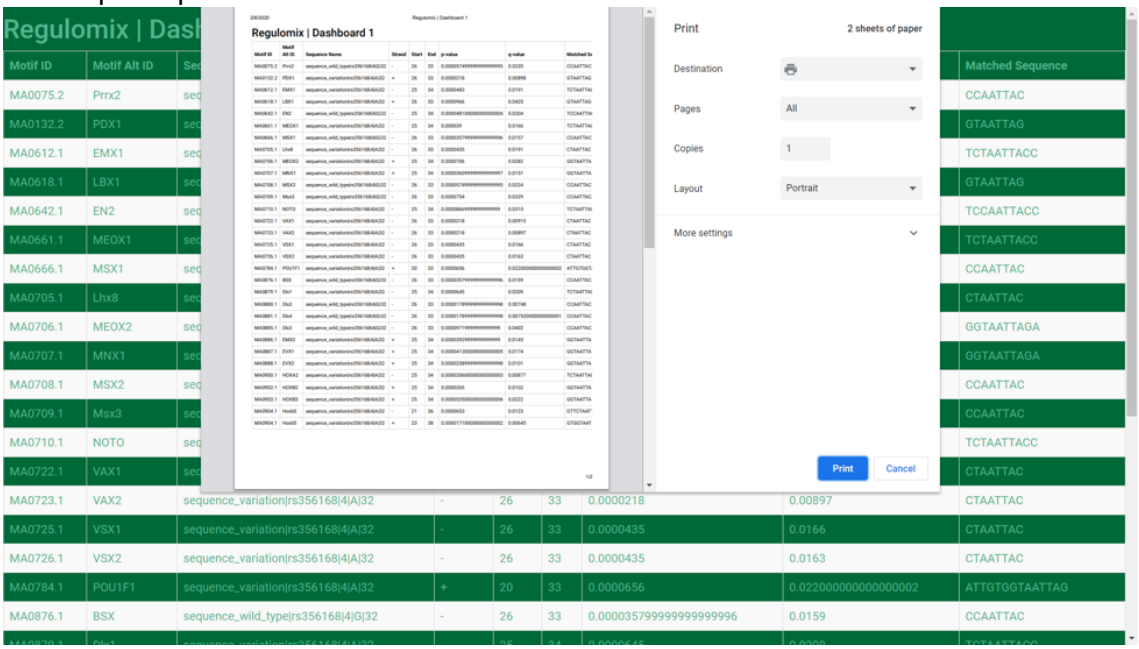
You can use the “Search” textbox to look for any value in the table such as: Motif ID, Motif Alt ID, Sequence Name, Strand, Start, End, p-value, q-value and Matched Sequence.



It is possible to “Copy”, “Print” or download a “.csv” file of your data from STEP 4 using the buttons below “Run” button.

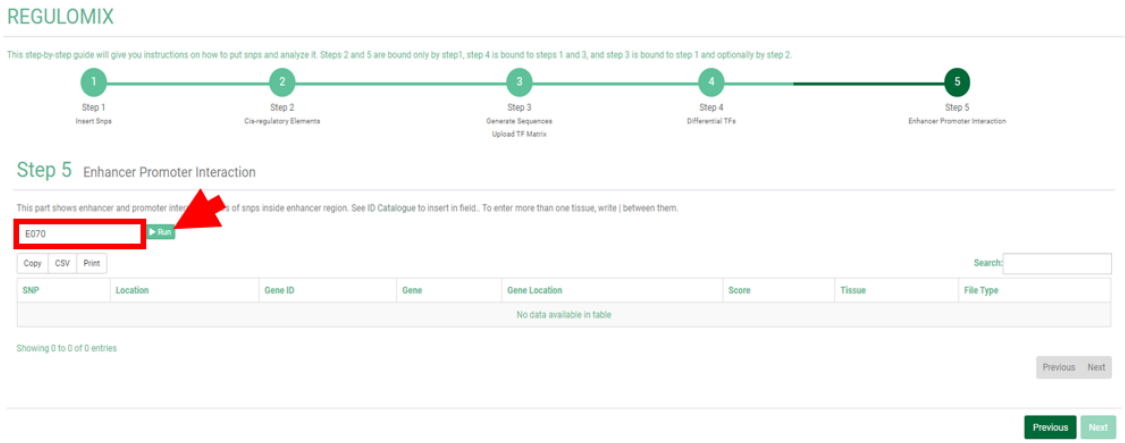


Pressing “Copy” will copy the Step 4 analysis data to your Clipboard. You can paste it in whatever program you want to save the information.
Pressing “CSV” will download a .csv file with the information of the Step 4 analysis.
Pressing “Print” will open a new tab with only the Step 4 analysis table information and it will open a printer information window.

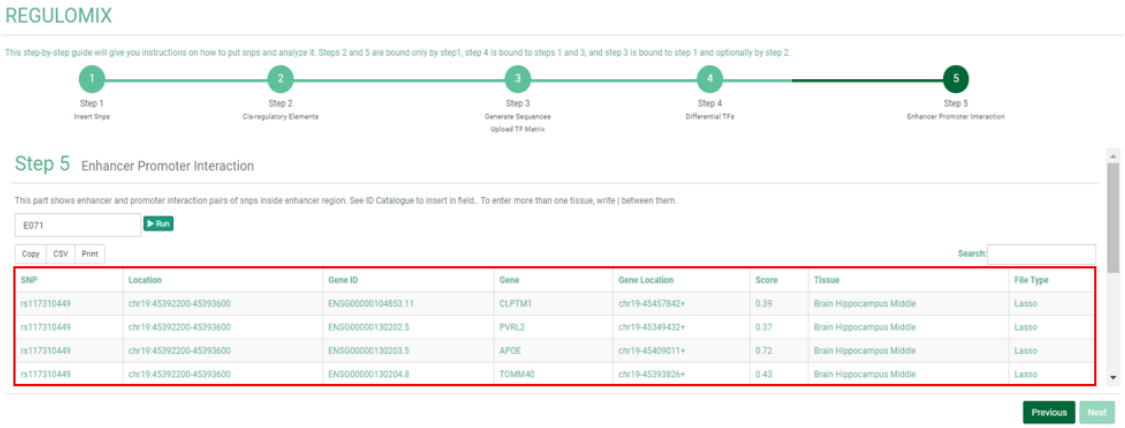


Step 5 – Enhancer Promoter Interaction:

In Step 5, you'll be able to request enhancer-promoter interaction (EPI). REGULOMIX uses JEME, joint effect of multiple enhancers (Cao, Qin et al., 2016), in Step 5 to identify EPI. The prediction uses a learned random forest model with cross-validation for predicting enhancer targets. For more information about how JEME works, go to doi:10.1038/ng.3950.
Use the "ROADMAP Epigenomics tissue ID" of the tissue you want to use for the analysis in the textbox. You can enter multiple tissues by using the symbol "|" between them.



After the REGULOMIX is done, a table will appear below the tissue textbox.

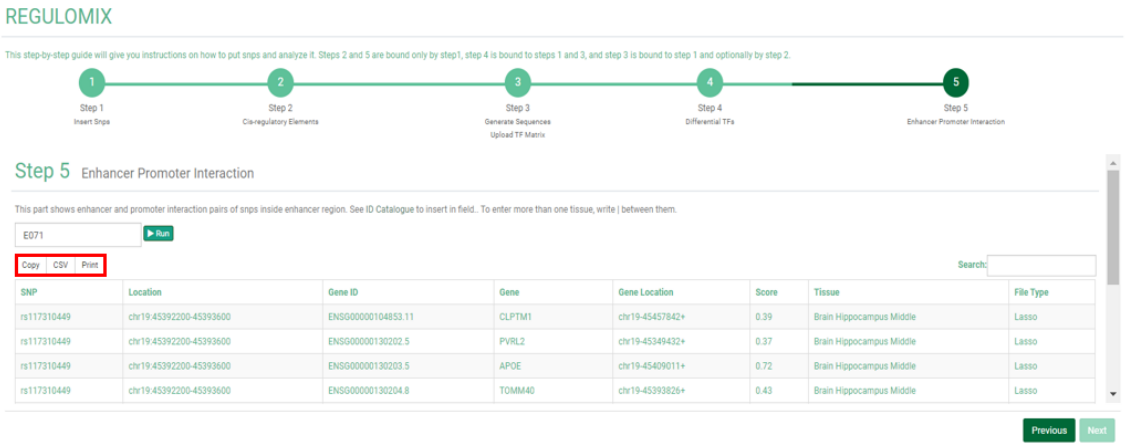


The meaning of each column in step 5:
SNP: SNP ID/RefSNP number of the SNP contained in the sequence being analyzed.
Location: Location of the SNP contained in the sequence being analyzed.
Gene ID: Ensembl Gene ID of the Gene predicted to interact with the sequence being analyzed.
Gene: The gene that potentially interacts with the sequence being analyzed.
Gene Location: Localization of the Gene predicted to interact with the sequence being analyzed.

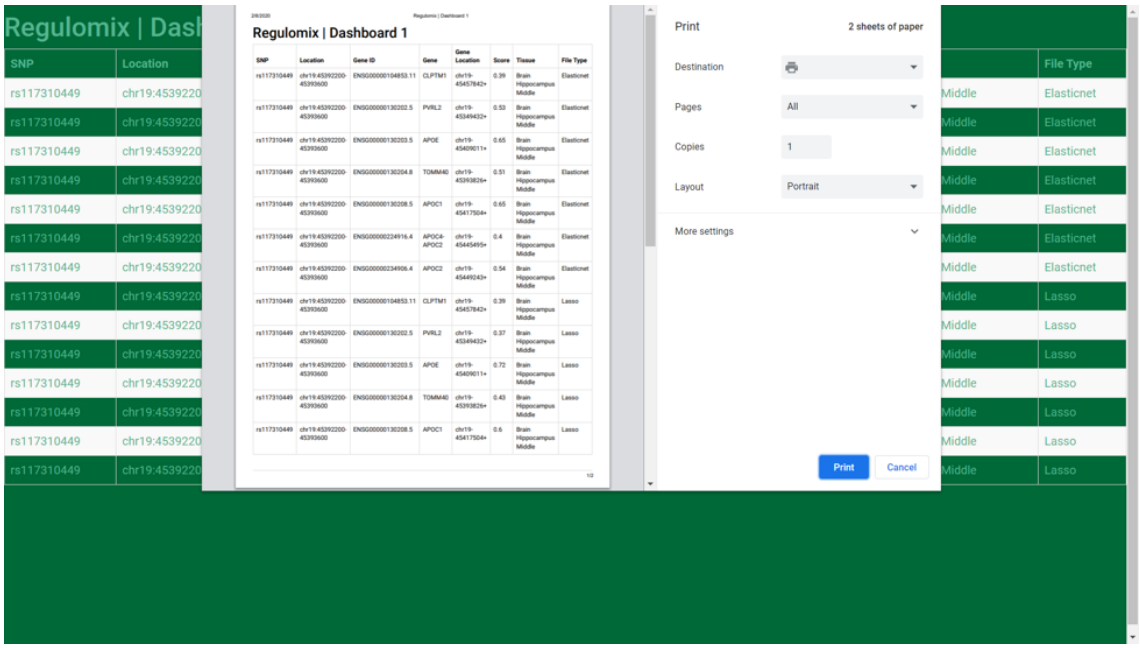
Score: Enhancer-target Interaction Scores. REGULOMIX uses JEME, joint effect of multiple enhancers (Cao et al., 2016), in Step 5 to identify EPI. The prediction uses a learned random forest model with cross-validation for predicting enhancer targets. For more information about how JEME works, go to doi:10.1038/ng.3950.
Tissue: Chosen by user.Tissue where EPI predictions were made.
File Type: Statistic regression method used in the prediction of EPI.

REGULOMIX uses JEME, joint effect of multiple enhancers (Cao et al., 2016), in Step 5 to identify EPI. JEME uses Elasticnet regression or Lasso regression for it's results.

For more information about how JEME works, go to doi:10.1038/ng.3950.



It is possible to "Copy", "Print" or download a ".csv" file of your data from STEP 5 using the buttons below "Run" button.



Pressing "Copy" will copy the Step 5 analysis data to your Clipboard. You can paste it in whatever program you want to save the information.
Pressing "CSV" will download a .csv file with the information of the Step 5 analysis.

Pressing “Print” will open a new tab with only the Step 5 analysis table information and it will open a printer information window.