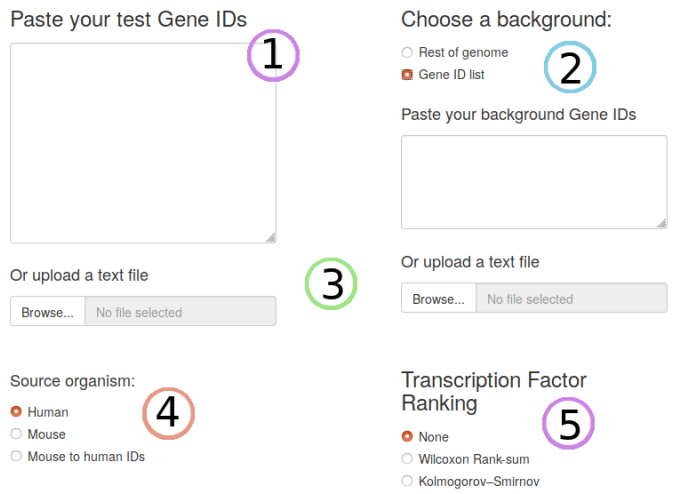
User Guide

TFEA.ChIP is a tool to estimate transcription factor enrichment in a set of differentially expressed genes using data from ChIP-Seq experiments performed in different tissues and conditions. To that end, it performs two type of analyses: association and GSEA-like.

Association analysis

Assesses whether there's a significant difference in the proportion of genes a transcription factor interacted with,in both the test gene list and the background gene list.

Input

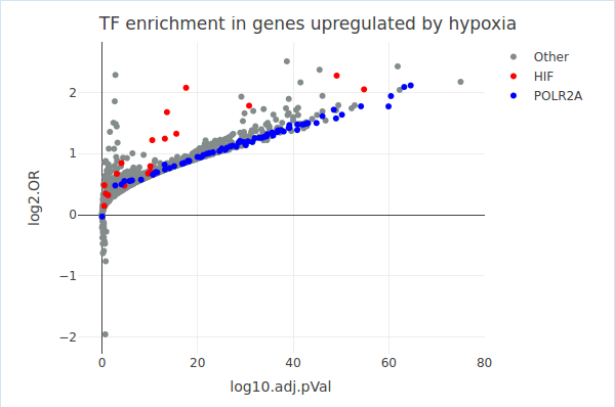


Paste your gene IDs - one ID by line - into the input boxes (-1-). TFEA.ChIP supports Ensembl gene ID, Entrez gene ID, and Gene Symbols. If you use mouse gene IDs as input, they can be translated to their human equivalent.

Choose a background gene list (-2-). Our recommendation is to use a random sample of the rest of genes expressed in your experiment, but you can also use genes that have a constant expression across the conditions of your experiment. Alternatively, you can choose to use the rest of the human genome as background. For both gene list you can paste the genes directly or upload a text file (-3-).

Select the organism (-4-): we currently have developed 6 databases based on human ChIP-Seq experiments, as well as 3 from mouse. At the moment, the human database covers a wider range of transcription factors, so it is possible to use mouse gene IDs with a human database picking the option '*Mouse to human IDs*', that will select the human genes equivalent to the mouse IDs on the input.

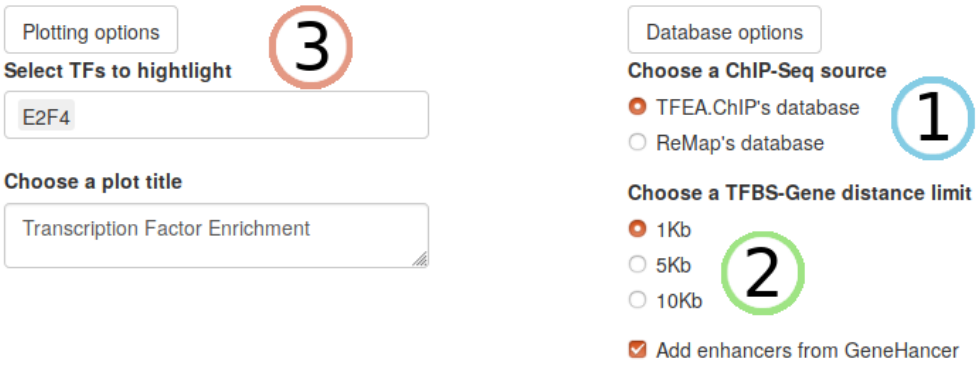
If you choose “Rest of genome” as background gene list, that includes all genes in the KnownGene databases, most of which may not be transcribed. In this case, many of the enriched transcription factors might be transcription-related - for instance, POLR2A -



The new feature *Transcription Factor Ranking* (-5-) can be used to summarize the results by transcription factor, testing whether ChIPs belonging to the same TF are, as a group, significantly enriched / depleted in the results of the analysis. Be aware that in the case of transcription factors whose behavior is dependent on cellular context, integrating the results of all the related ChIPs might conceal its enrichment in a particular set of experimental conditions.

All genes that are not included in the KnownGene database , such as RNA genes, will be discarded.

Database and plotting options

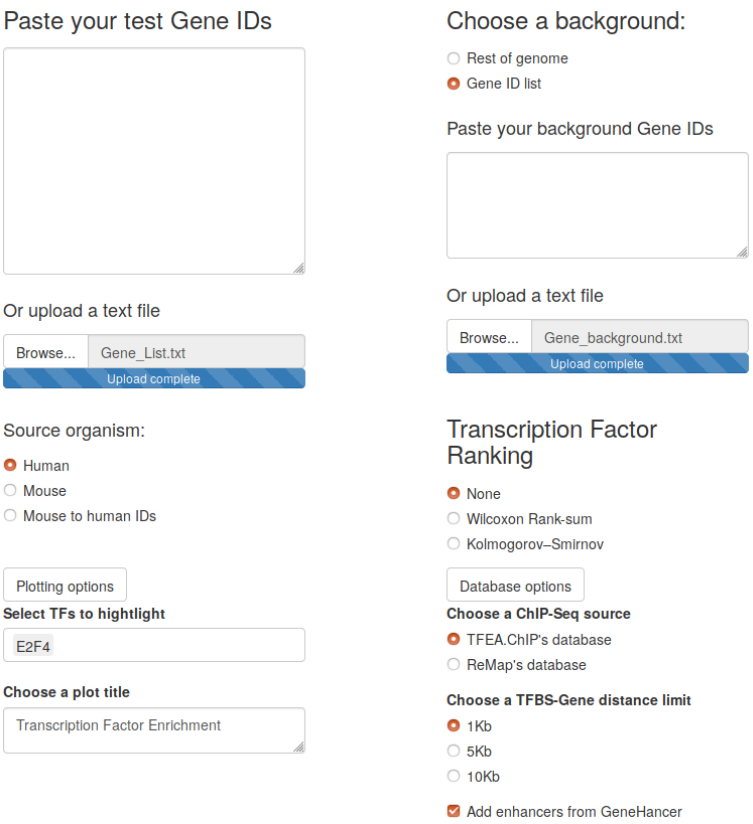


At this moment, TFEA.ChIP’s web app includes two sources of ChIP-Seq experiments for human TFs: ReMap’s 2018 selection (Jeanne Chèneby et al., 2017) and the collection originally gathered for TFEA.ChIP (-1-).

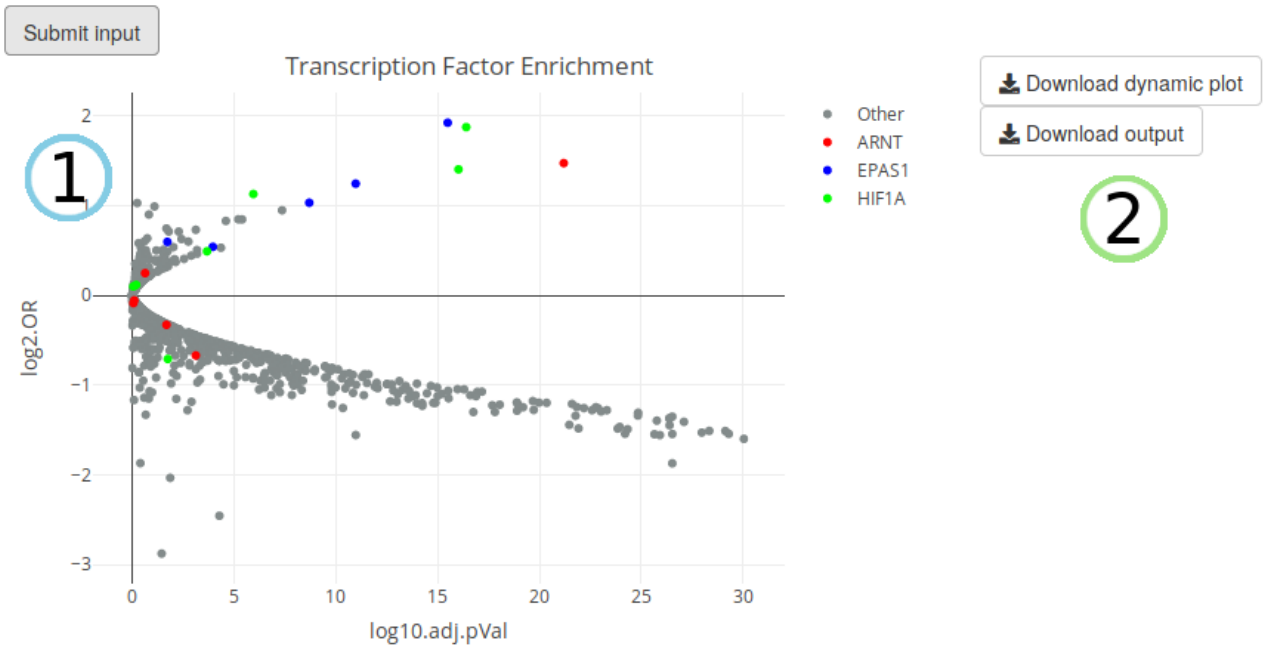
Choose the maximum distance allowed between TFBS and genes (-2-). You have the option to include TFBS inside distant regulatory elements identified in “*The accessible chromatin landscape of the human genome*“ (Thurman et al., 2012).

To represent the results, choose the transcription factors you’re interested in, in order to plot them in different colors (-3-). An interactive plot will be generated below. Hovering the mouse over a point will display more information about the ChIP-Seq experiment in question (Cell type, accession ID, or treatment).

Once all the fields are filled, the page will look like this:



Click the “Submit input” button to start the analysis. In this example we’ve used genes that are up-regulated after 8h of hypoxia as a test gene list, and a random selection of the rest of the genes in the experiment as background.



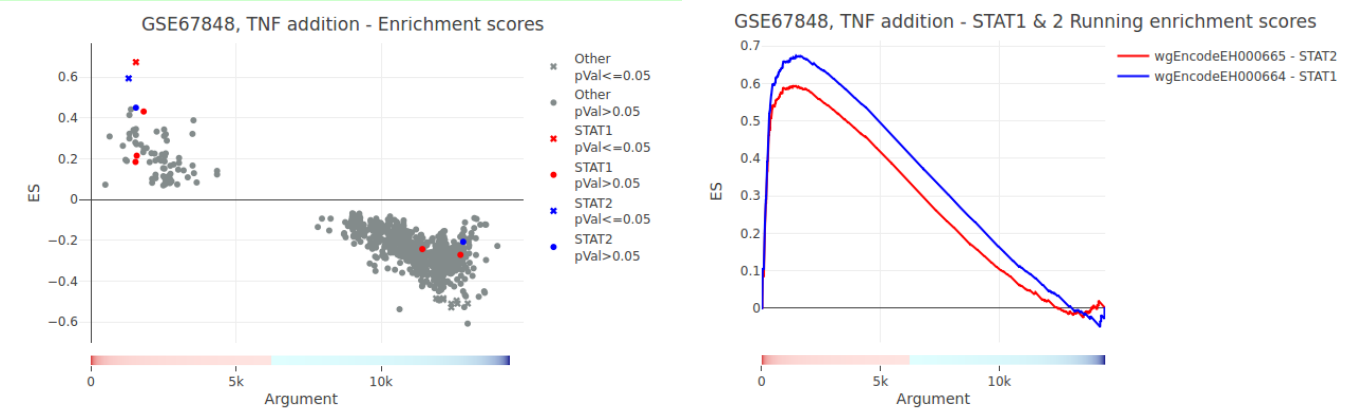
The plot will represent Log2(Odds Ratio) versus Log10(adjusted p-value) for every ChIP-Seq experiment (-1-).

You can download the interactive plot -in HTML format-, or a zip file with all the analysis results, including the transcription factor ranking if selected. (-2-). The results table includes: ChIP-Seq accession, cell type, treatment, transcription factor tested, raw and adjusted p-values, odds ratio, and euclidean distance from log2(OR)/log10(p-value) to the origin.

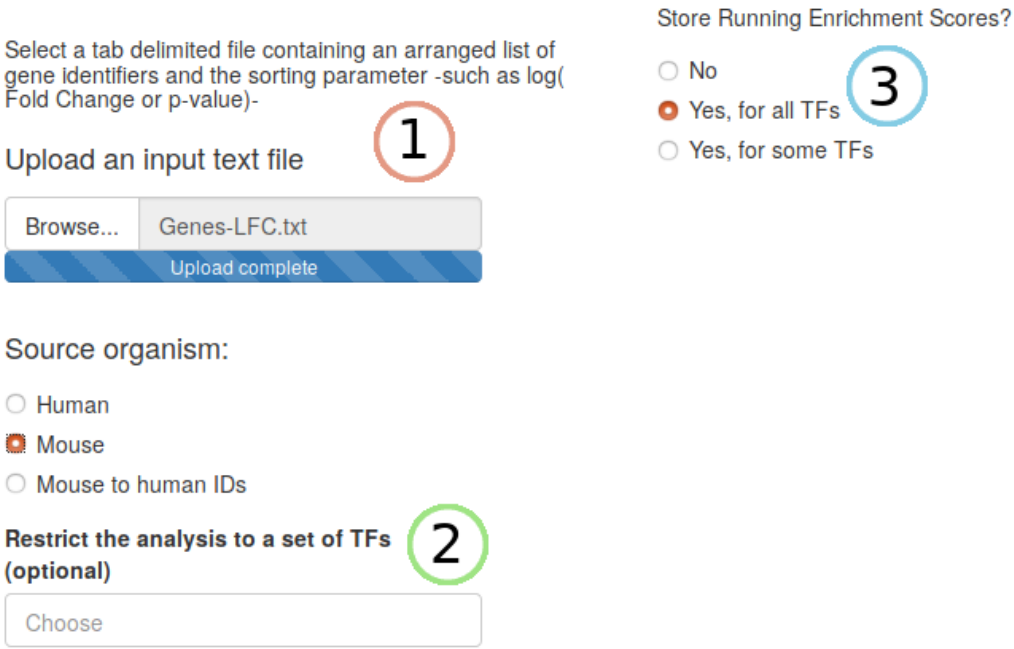
GSEA-like analysis

With this analysis it's possible to estimate if there's a significant concentration of a TF's targets in a list of genes -if the genes are together in a section of the list, or dispersed across it-.

Here is an example of the results generated:



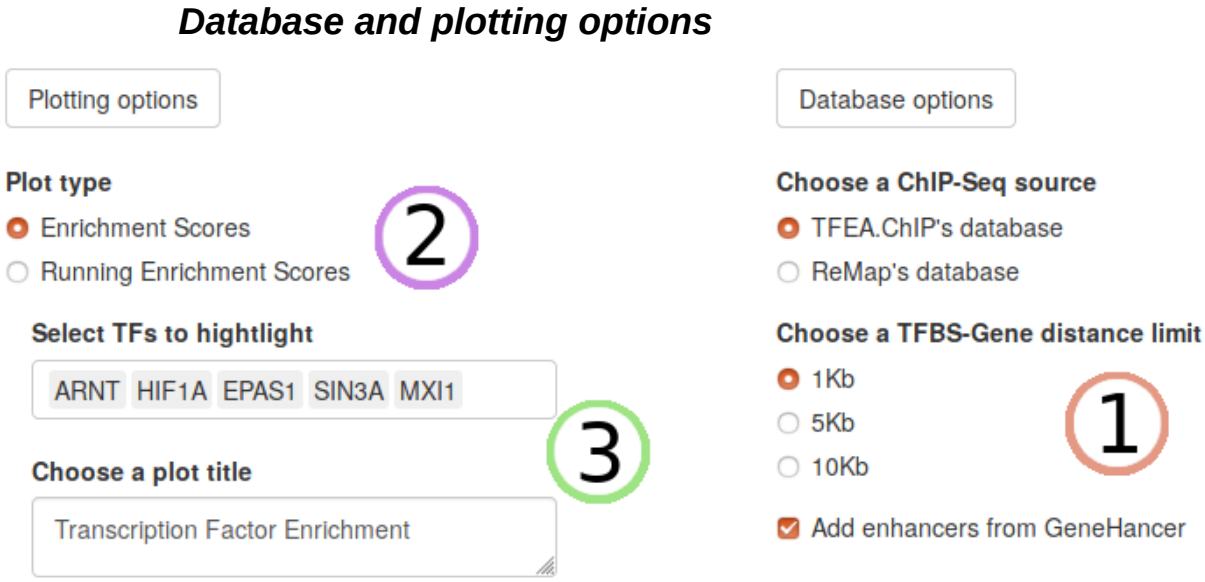
Input



In this case the input required is a text file with two columns separated by tabulation, containing gene IDs in the first column and log2(Fold Change) values in the second (-1-). We recommend that you exclude genes with a log2(FoldChange) of ±Inf or 0.

This analysis takes more time to compute using all the ChIP-Seqs included in our database, so you can select specific transcription factors (-2-) of your interest instead of using all to reduce computing time.

You can also choose whether to store the Running Enrichment Scores (-3-) - the score for every ChIP-Seq experiment at each point of the gene list- for all or some transcription factors.

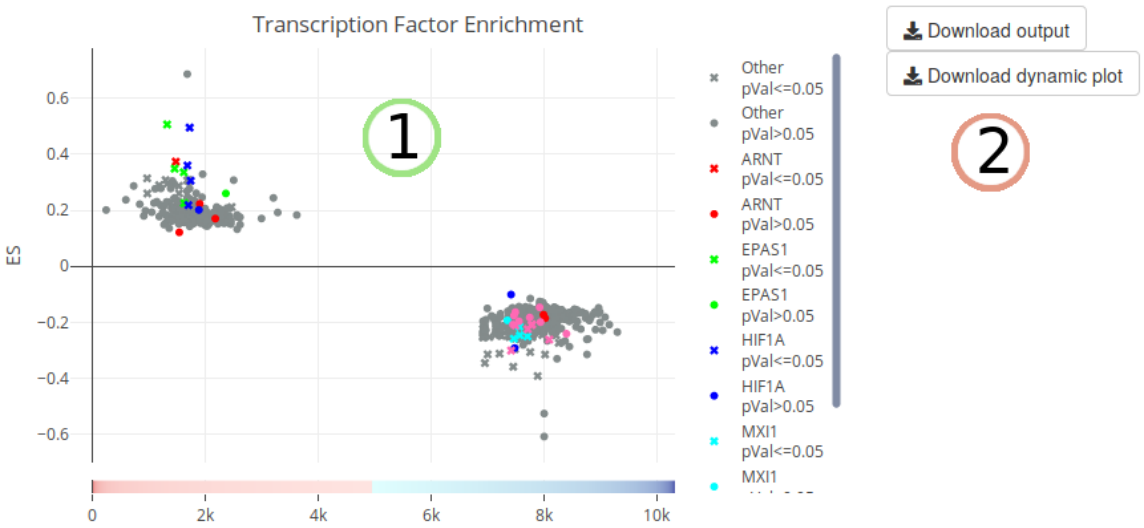


As in association analysis, we can specify the ChIP-seq source, maximum distance allowed between TFBS and genes and whether to include enhancers. (-1-).

As for graphic options, we can choose to plot either Enrichment Scores (-2-) - maximum score for each ChIP-Seq experiment-, or Running Enrichment Scores - the score of each ChIP-Seq at each element of the gene list-.

To plot Enrichment Scores, as with association analysis, just select which transcription factors to highlight (-3-).

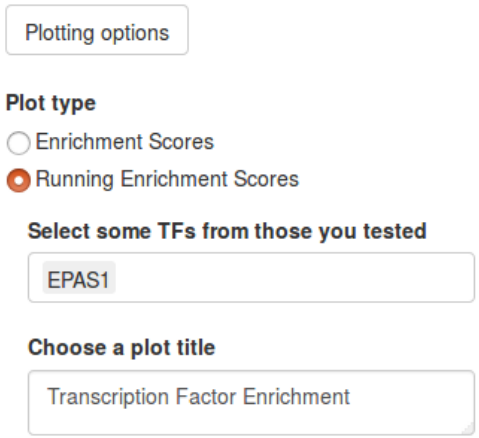
Following the former example, we will use a dataset with gene IDs and their log2(Fold Change) after being exposed to hypoxia:



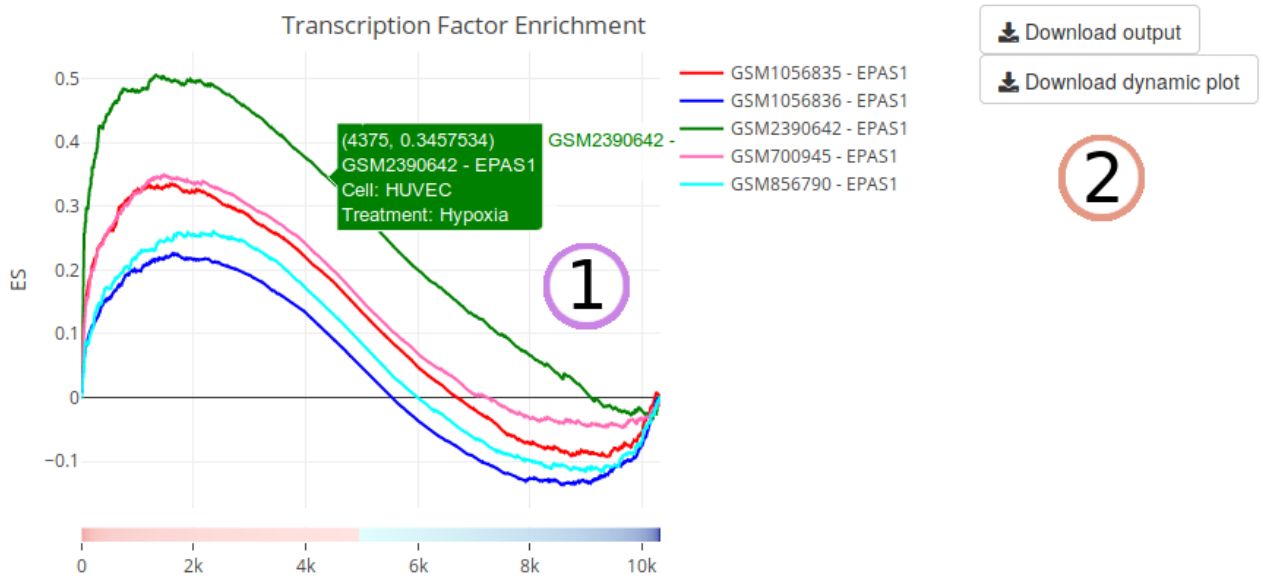
The plot will represent the Enrichment Scores versus the Argument -the position on the gene list- (-1-).

You can download either the plot, or a zip file containing the results table in CSV format, as well as on RData format to later use it on R.

If you stored the Running Enrichment Scores for any transcription factor, check “Running Enrichment Scores” in Plot type, and then select the transcription factor whose RES you want to plot.



Here is an example of the RES for EPAS1 ChIP-Seqs in our hypoxia dataset:



You can see the information displayed when hovering with your mouse over a line (-1-). In this case, you can also download the Running Enrichment Scores you stored in CSV format, with a ChIP-Seq experiment in each column, and their scores step by step in each row (-2-).

More

If you want to integrate TFEA.ChIP to your analysis pipeline, personalize the TFBS database with your data, or create your own from scratch, please check our R package on Bioconductor.

Our databases and the latest version of TFEA.ChIP for R are available for download at our GitHub repository.