Plugin User Guide

Author: Matti Hoch

The **AirPlugins** is a MINERVA plugin explicitly developed for the "Atlas of Inflammation Resolution" (AIR) [1] to help users in exploration of information stored in the molecular interaction map as well as perform *in silico* perturbation experiments and data analysis. This guide aims to help users understand the UI and correctly interpret results. The **AirPlugins** is a plugin suite divided into individual plugins for different approaches to improving usability. In the following, we will describe each part separately.

This guide focuses on explaining UI elements in the plugins. Make sure to read the recent manuscript (currently available as a PrePrint) to understand the principles and algorithms behind the tool:

https://www.biorxiv.org/content/10.1101/2021.09.13.460023v1

Contents

At all times, while the AirPlugins are active, clicking on a phenotype on any map will visualize the influence scores of all its regulators as colored overlays on the map.

Xplore

This plugin provides tools to explore the AIR beyond the interactive submaps and perform data-independent analyses through in silico perturbation experiments. It allows the user to search for elements, find regulators and targets, perturb elements, and predict potential drug targets for selected phenotypes.

The Xplore plugin consists of four parts: Data exploration, *in silico* perturbation experiments (downstream enrichment), target prediction (upstream enrichment), and data export.

Exploring the MIM: User can access information from the MIM beyond the submaps by selecting an element on the map or manually entering the name. Tables are displayed that give an overview of all direct regulators, targeting elements, and interaction path distances to phenotypes.

Perturbing elements: In addition to user defined DCEs, elements that are knocked out (KO) can be specified. These elements are considered removed from the network, resulting in a recalculation of shortest paths and influence scores.

Highlighting interaction paths: For each element in the submaps, the shortest path associated with a phenotype can be visualized. If knocked out elements were specified, the path is automatically redirected. Figure S1 shows such a case by disrupting the path of TNF to the phenotype "apoptotic process" through the knock-out of CASP3.

Visualizing phenotype regulators: Selecting a phenotype on the map while the plugin is active, automatically highlights all regulators with a colored overlay of red for positive and blue for negative influence scores. Figure S2 shows the submap "Biosynthesis of PIMs and SPMs from AA" after selecting the phenotype "Prostaglandin synthesis". It shows how hub elements (PGH2) or key enzymes (PTGS2) have the highest scores among all regulators. Additionally, enzymes that catalyze metabolites in the pathway have negative scores due to the consumption of intermediates.

Inferring phenotypes: Users set a custom log2 FC value for elements in the MIM through a slider bar (Figure S3), defining them as DCEs. Changing the values of the DCEs automatically updates the predicted level for all phenotypes and visualizes the results as colored overlays on the map.

Inferring targets: Users select phenotypes as DCEs by defining a log2 FC value between -1 and 1 representing changes in their activity. Predicted targets, positive and negative, are visualized in a scatter plot ranked by their sensitivity and specificity and can be filtered by their molecule type.

Box 1: Main features of the Xplore plugin.

Data exploration

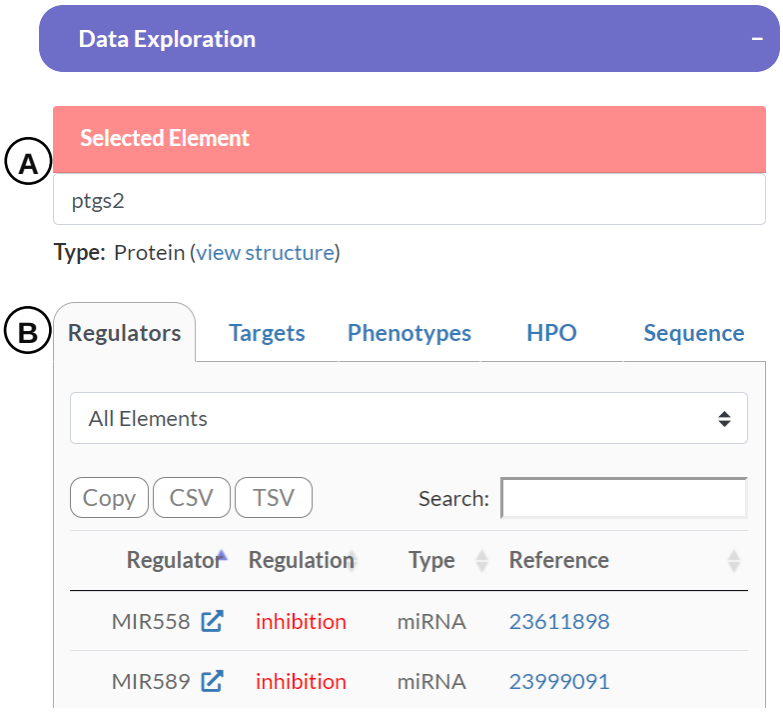

Figure 1

To get information on a specific element in the MIM, the user writes the gene symbol or metabolite name case-independent into a textbox (Figure 1A) or selects an element on the map. If the element exists, its type will be displayed below together with a link, if available, that, if clicked, opens up in a popup window an interactive 3D visualization of the element's molecular structure.

Five different tabs contain all information on the selected element's properties and interactions (Figure 1B):

- *Regulations:* List of all elements that directly interact with the selected one, including their type, the type of interaction, and PubMed references. Furthermore, the user can filter the regulators by their type, e.g., showing microRNAs only.
- *Targets:* A list of all elements that the selected one directly interacts with, including their type, the type of interaction, and PubMed references.
- *Phenotypes:* List of all phenotypes that are, directly or indirectly, regulated by the selected element, the type of regulation, and the shortest path distance between them.
- *HPO:* We provide an API connection to the “human phenotype ontology” [2], showing interactions of the selected element with phenotypes and diseases from the HPO database.
- *Sequence*: If the selected element is a protein included in the UniProt database [3], this tab will display an interactive panel (by ProtVista [4]) to visualize properties of the protein's sequence, such as amino acid modifications or genetic variants.

Downstream Enrichment

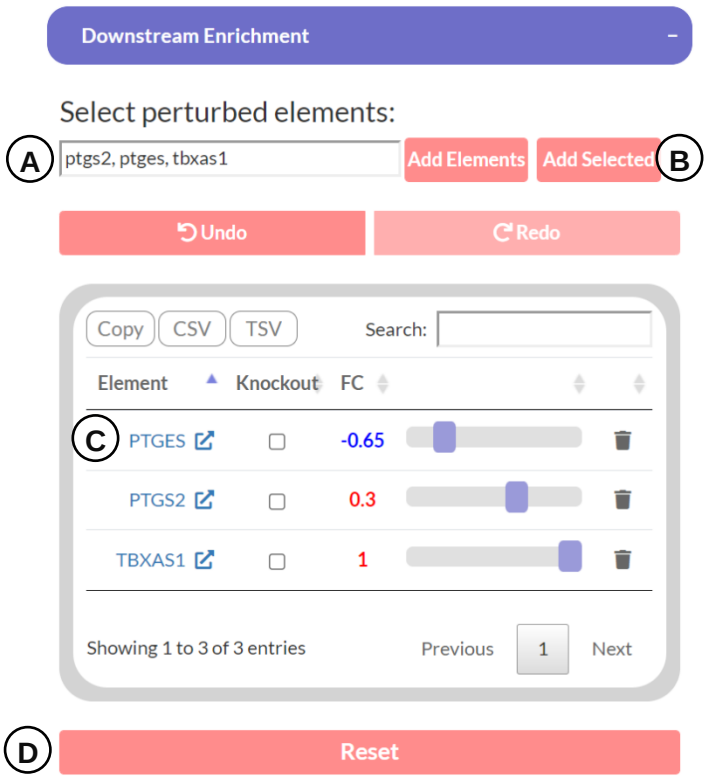

Figure 2

To perform in-silico perturbation experiments, the user must select the elements to be perturbed by either entering them comma-separated by their name (Figure 2A) or adding the selected element from the map (Figure 2B). The added elements appear in a table below (Figure 2C) in which their custom log2 fold change can be defined, or the element is knocked out entirely. The latter will result in a recalculation of influence scores and shortest paths. Be aware that this will affect other analyses as well, however, only in the AirXplore plugin. Added elements can be deleted with the trashcan button, and changes be reversed with the undo/redo buttons above the table. The reset button (Figure 2D) will set all values back to default values (zero).

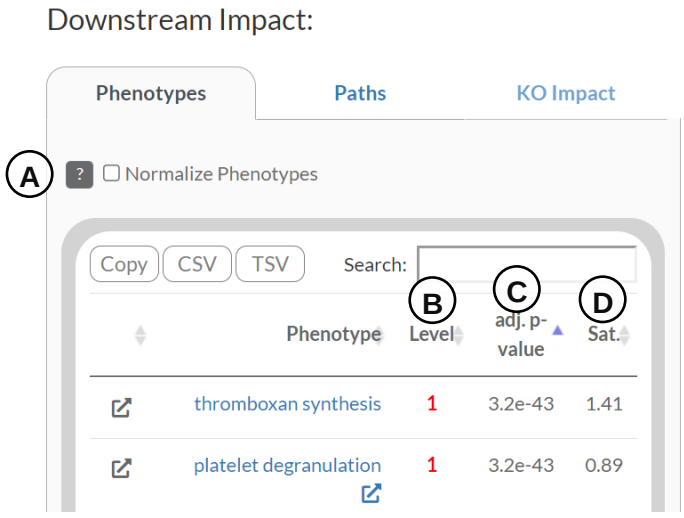

Figure 3

Every change made in the log2 FC values of perturbed elements will automatically update the result table. In the table, all phenotypes in the AIR are listed with their predicted levels (Figure 3B, values between -1 and 1 representing reduced or increased activity), p-value (Figure 3C) and saturation (Figure 3D, weighted percentage of perturbed regulators). Predicted phenotype values greater than 1 or less than -1 (e.g., when many regulators are perturbed) are set to 1 or -1.

Upstream Enrichment

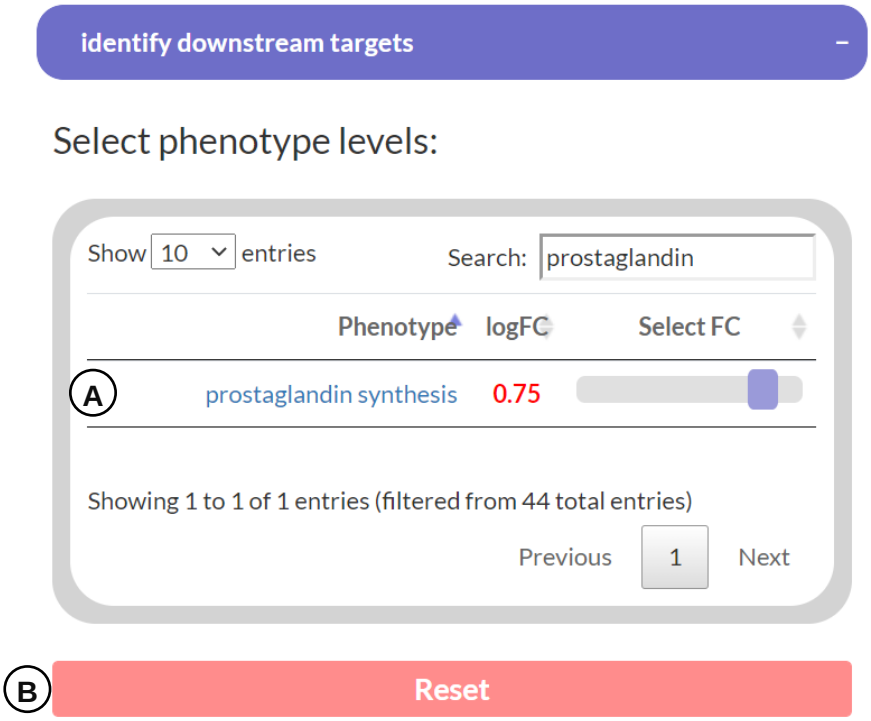

Figure 4

This tool identifies targets that regulate a set of individually perturbed phenotypes. A table lists all phenotype in the AIR and allows the user to set a custom log2 FoldChange value for each (Figure 4A). The reset button (Figure 4B) sets all values back to zero. Changing the value will automatically update the results graph below.

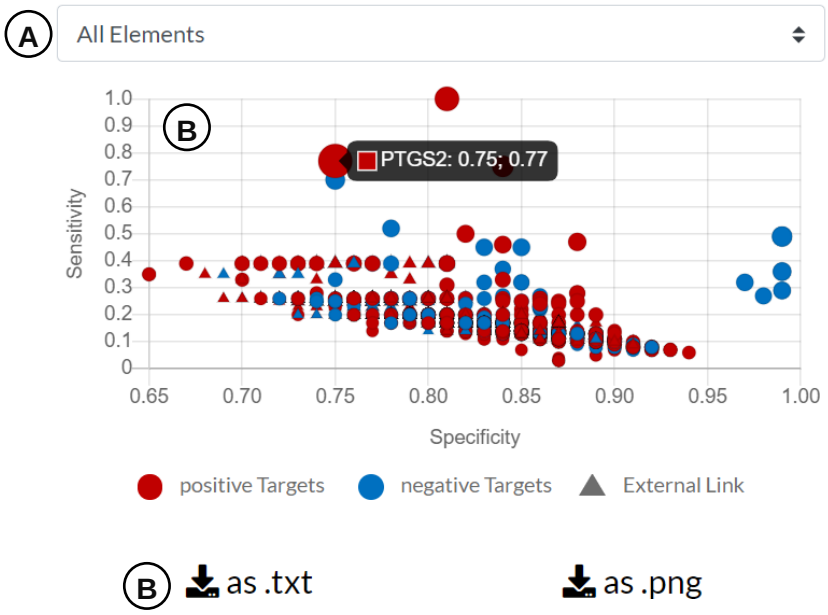

Figure 5

Predicted targets are visualized in a scatter plot (Figure 5B) by their sensitivity (ability to regulate the perturbed phenotypes in the specified way) and specificity (inability to regulate non-perturbed phenotypes). The targets are divided by their color based on their type: *positive* if the element induces the user-defined phenotypic changes, and *negative* if it induces the exact opposite changes. The data points on the plot directly link to the position of the individual element on the submap. Elements marked as '*external links*' are not included on the maps and will link to their entry on public databases. The targets can be filtered by their molecule type (Figure 5A), e.g., only showing proteins or metabolites. Results can be downloaded as .txt files in a tab-separated format or as a graph image (Figure 5C).

Exporting Data

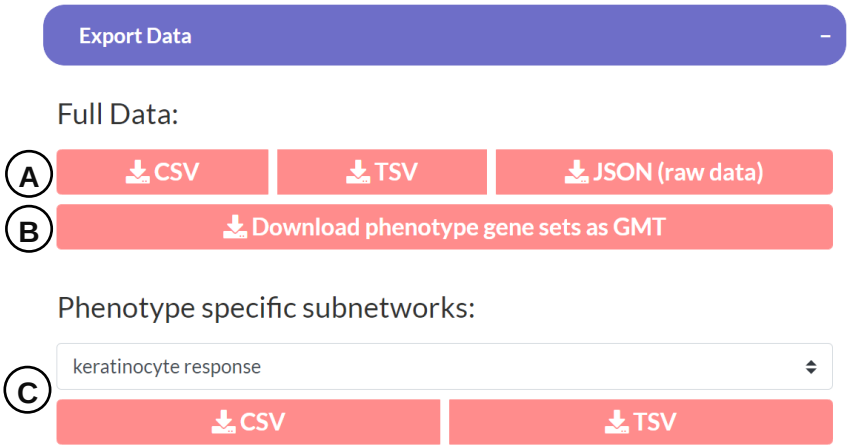

Figure 6

All information on the AIR is publicly available. We provide options to download the data, or parts of it, in multiple formats:

(A) the whole data as comma-separated values files (CSV), tabular separated values files (TSV), or the raw data as JSON
(B) gene sets for each phenotype in GMT format, including gene ontology (GO) IDs if available.
(C) phenotype-specific subnetworks extracted of the original data that display regulatory pathways for each phenotype.

(B) and (C) are generated from interactions in the MIM between elements with influence scores on the phenotype other than zero.

Omics

This plugin allows users to upload context-specific -omics data files and to analyze them in the context of acute inflammation.

Selecting a data file

In the default UI, users can choose local files that contain the data either in tab- (tsv) or comma-separated (csv) format (Figure 7A). The files should contain one column with the probe's identifier, either the name or a database ID (Figure 7D), and log2fold change values (FC) of each sample in other columns. The file can also contain p-values as another column following the FC for each sample. Figure 8 gives an example for a tsv file without p-values (A) and a csv file with p-values (B). Since the probes in the data files can be of any molecule type, they are referred to hereafter as "differentially modified elements" (DCEs).

After selecting a file, the plugin automatically selects the specifications, such as file type (Figure 7B), the column containing probe identifiers (Figure 7C), and whether p-values are included (Figure 7E). Furthermore, the mapping type can be specified if the file contains multiple copies of the same probe, e.g., multiple transcripts of the same gene (Figure 7F). Finally, Figure 7G selects the type of data (differential or non-differential). Currently, differential analysis is the way to go. However, features to analyze non-differential data, i.e., normalized read counts instead of FC values, are in development. If entries are not readable (e.g., are non-numerical), the user is notified through a pop-up that these values were replaced with a 0 (for FC values) or a 1 (for p-values). Empty lines or lines with an inconsistent number of columns will be skipped.

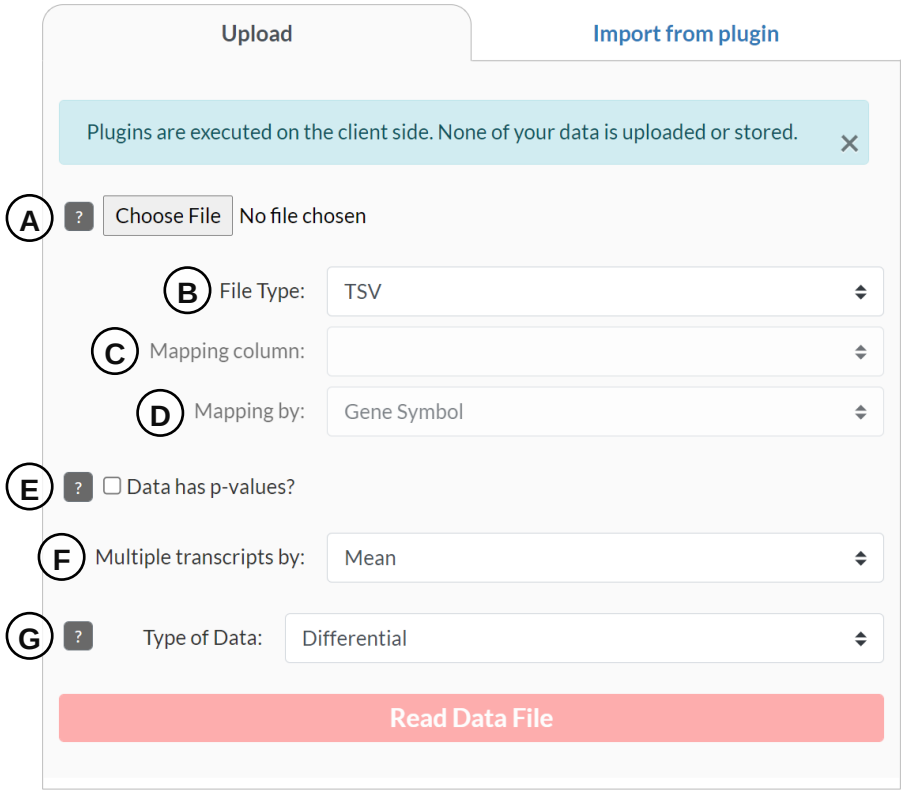

Figure 7

(A)

gene	DSSd02	DSSd04	DSSd06	DSSd07	DSSd08	DSSd10	DSSd12	DSSd14					
0610007P14Rik	0.210246420274739	0.228157458847654	0.487217230740923	0.296866387523099	0.367879615751189	0.581557804289795	0.675252683415104	0.492951935762434					
0610009B22Rik	-0.114110214568408	0.0530964464608665	0.211510199526561	0.265641512087683	0.343216340617059	0.576245043857571	0.502006959520067	0.406488815896629					
0610009L18Rik	0.186447000353112	0.14170631236766	-0.177679233153853	-0.521323765512055	0.411976130331396	0.0160183807741337	0.170355692818444	0.212125330587446					
0610009O20Rik	0.0589755663561134	0.0371876123194772	-0.0443389240603329	-0.228283652930036	-0.0688109280745007	-0.219487114748834	-0.16569424160708	-0.0117481008590319					
0610010F05Rik	0.167688991137998	0.00128062351805297	0.409265234293219	0.402573543237004	0.417177230037588	0.643004484482591	0.383685979326851	0.415092837879018					
0610010K14Rik	0.136847542138415	0.273421565126885	0.173327548463693	-0.0302118906586328	-0.0298320194029359	0.729313530568813	0.244985734922296	0.0706133987506003					
0610030E20Rik	-0.516684170867145	-0.567406675570997	-0.371684897808254	-0.212901621562827	-0.182997021217456	-0.876403271676305	-0.409021847568805	-0.169725995403962					
0610037L13Rik	0.101324164365937	0.192919816412273	0.358000319431534	0.194842663294701	0.157550589946717	0.348171449014905	0.226238261674266	0.421959185214603					

(B)

```
gene,DSSd02,DSSd02_pvalue,DSSd04,DSSd04_pvalue,DSSd06,DSSd06_pvalue,DSSd07,DSSd07_pvalue,DSSd08,DSSd08_pvalue,DSSd10,DSSd10_pvalue,DSSd12,DSSd12_pvalue,DSSd14,DSSd14_pvalue
0610007P14Rik,0.210246420274739,0.270696053630314,0.228157458847654,0.131828102724511,0.487217230740923,0.000110617404143746,0.296866387523099,0.037828340839234,0.367879615751511
0610009B22Rik,-0.114110214568408,0.798856567199683,0.0530964464608665,0.866662952370649,0.211510199526561,0.42223179858883,0.265641512087683,0.209143791992511,0.343216340617059
0610009L18Rik,0.186447000353112,0.824748264219118,0.14170631236766,0.824646517859631,-0.177679233153853,0.797541845497837,-0.521323765512055,0.302380971940065,0.411976130331396
0610009O20Rik,0.0589755663561134,0.891454835815489,0.0371876123194772,0.860113244989666,-0.0443389240603329,0.881972887457659,-0.228283652930036,0.0885277184095822,-0.0688109280
0610010F05Rik,0.167688991137998,0.703756416717136,0.00128062351805297,0.997765311955666,0.409265234293219,0.215963232859941,0.402573543237004,0.195469997720934,0.417177230037580
0610010K14Rik,0.136847542138415,0.712322890853117,0.273421565126885,0.118413233666358,0.173327548463693,0.462471469195913,-0.0302118906586328,0.901854783732355,-0.0298320194029
0610030E20Rik,-0.516684170867145,0.00583538206882734,-0.567406675570997,0.00044556168039192,-0.371684897808254,0.0865397419309454,-0.212901621562827,0.27980783798321,-0.1829970
```

Figure 8

Importing results from other plugins

Suppose analyses have previously been performed using the Variant and MassSpec plugin. In that case, their results can be directly imported into the Omics plugin, e.g., to identify the effects of genetic variants (Variant) or changes in metabolite quantities (MassSpec) on phenotypes.

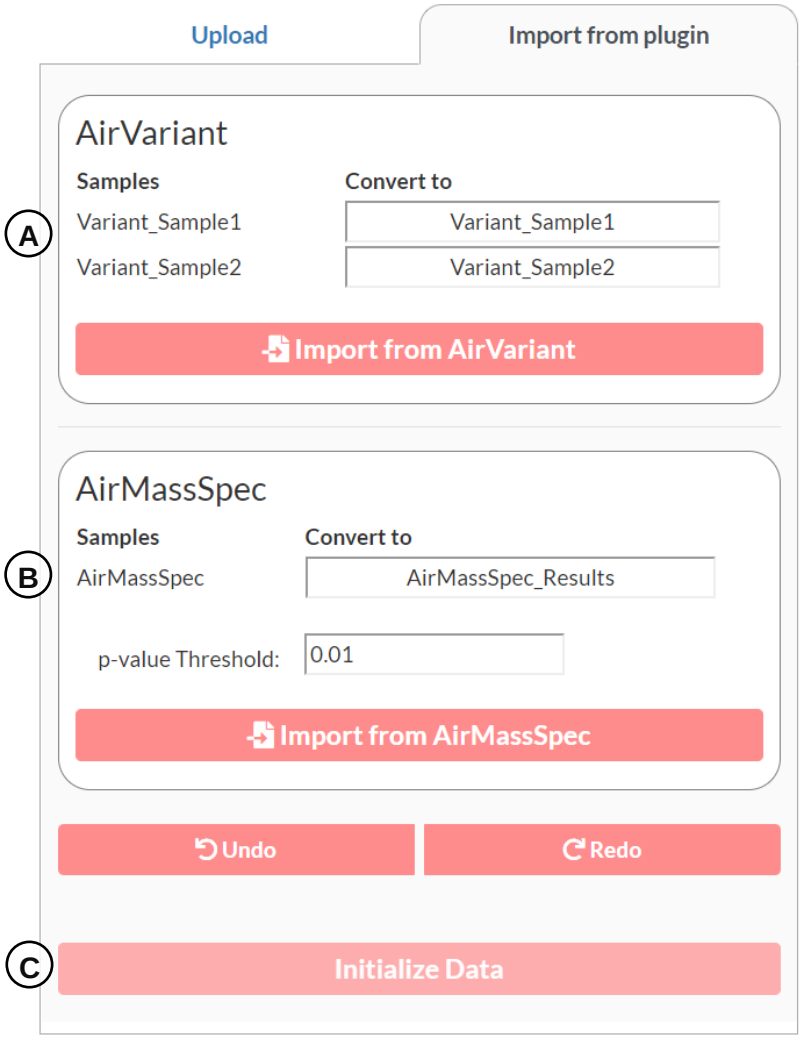

Figure 9

A custom sample name can be specified for each file used in the Variant (Figure 9A) and the MassSpec analysis (Figure 9B), which will be used in further analyses. If sample names overlap, their results will be merged.

Data analysis

The results panel will be available after the data has been initialized. If another data set has been initialized before, the user can decide whether to replace the data entirely or merge both sets based on the sample names. Two primary analyses can be performed: Estimating the effects of differential data on downstream phenotypes or predicting upstream targets.

Phenotype Inference

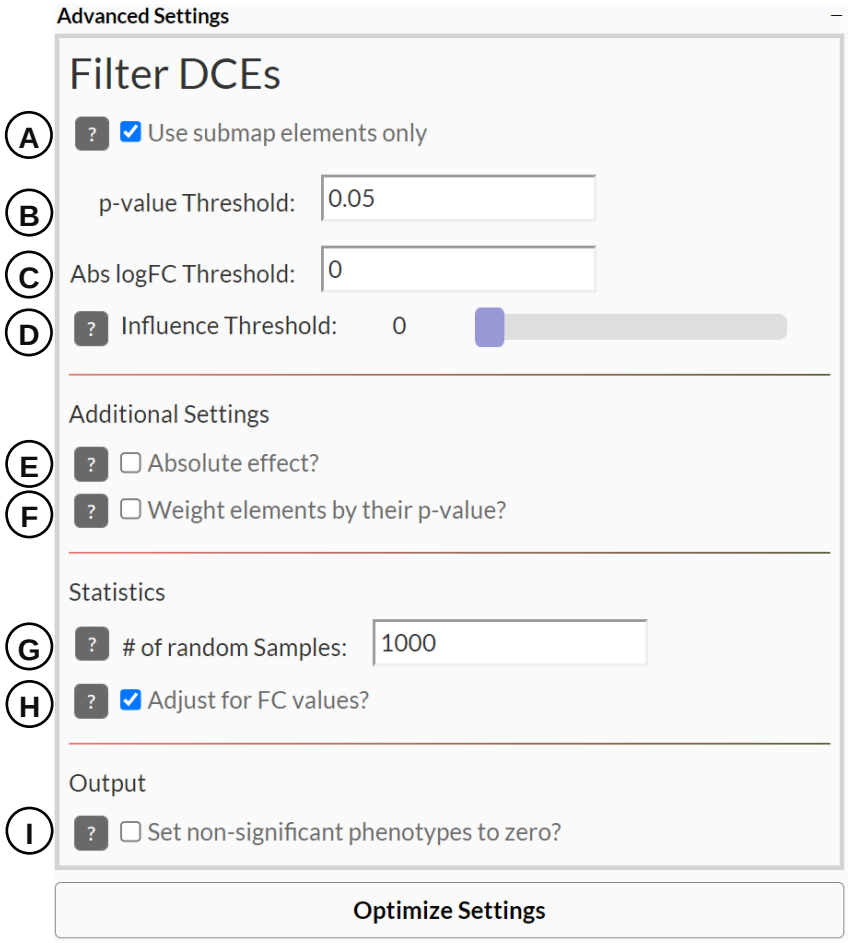

Figure 10

The AIR contains detailed information about the molecular pathways involved in the regulation of phenotypes. Because these pathways are directional, i.e., contain information about the regulation - positive or negative - we expect to predict changes in phenotype activity based on the influence of their regulatory elements and context-specific omics data mapped on the AIR. For example, if positive regulators of a particular phenotype show an increase in their expression, i.e., have a positive logFC value, we assume that the activity of that particular phenotype has also increased and vice versa. Thus, the algorithm combines the FC value of the elements with the network topology features to statistically evaluate their aggregated impact on the phenotype.

The phenotype estimation is the default active window after initializing the data. Usually, the analyses can be started directly by clicking the button (Figure 10J), with settings optimized for large datasets with many DCEs. In the advanced settings drop-down menu, different settings can be specified to fit parameters of the analysis, which will be explained in the following table:

(A) If checked, only elements from submaps are considered. Since the information from submaps consists of manually curated paths, this can increase accuracy, but only if many DCEs are available.

(B) p-value threshold for probes to be considered as DCEs. Only relevant if the dataset contains p-values.

(C) Log2 Fold change threshold (as absolute value) for probes to be considered as DCEs.

(D) Influence threshold (between 0 and 1) for regulators of phenotype to be considered for the analysis. Increasing the threshold may result in higher accuracy if many significant probes are available by removing the “background noise”.

(E) If checked, the absolute value of fold change will be considered for the phenotype assessment.

(F) If checked, p-values are included as a weighting factor for phenotype regulators. Each DCE’s (if below the threshold) impact on the phenotype will be multiplied by $\frac{\text{pvalue}_{threshold} - \text{pvalue}}{\text{pvalue}_{threshold}}$

(G) Number n of random samples to be generated for statistical evaluation. The impact on the calculation time is proportional to n.

(H) If checked, the statistical threshold (parameter k) is adjusted to the highest FC in every permutated set. If checked, false negatives are reduced in cases where the permuted FC values are higher than the FC values of the original sets by preventing nonphysiological FC values from biasing the results. However, as a result of this adjustment, sets with DCEs that have per se high FC values lose statistical power.

(I) If checked, enriched phenotypes with a p-value > 0.05 will be set to 0 and excluded from the normalization as well as the overlays.

After clicking of “Estimate Phenotypes” and the calculations are complete, the results are presented in three parts:

Table (main output)

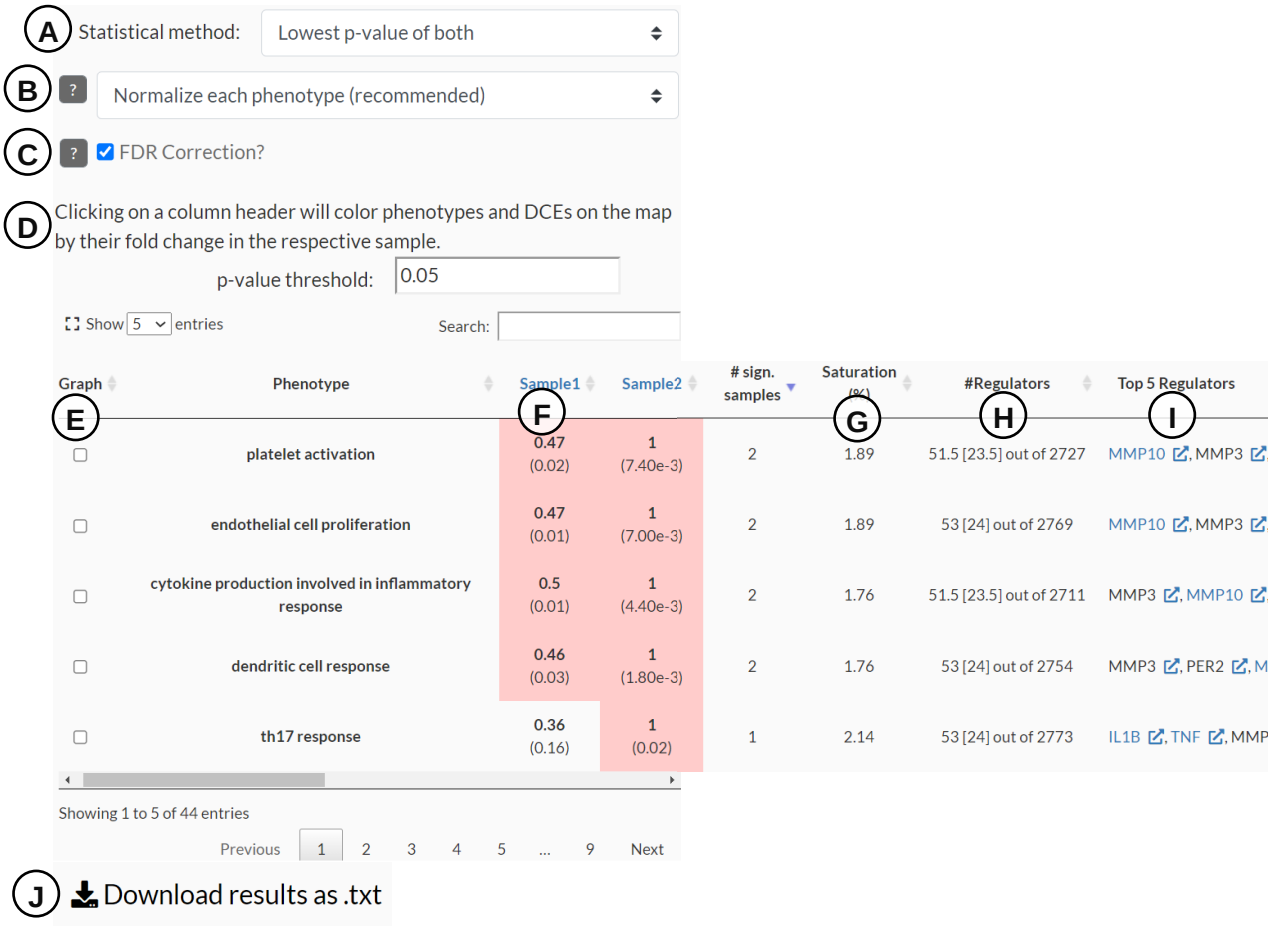

Figure 11

All results are directly displayed in a table. Clicking on a single value itself will pop up a new graph, containing information on the influence and FC values of all regulators:

(A) A drop-down menu to select the statistical method. For the differences between the methods and their interpretation, see the plugin manuscript.

(B) A drop-down menu to select the type of normalization applied to the phenotype levels. We recommend normalizing the results for each phenotype individually because values among different phenotypes cannot be directly associated with a different activity. However, if analyzing the absolute effect (C), no normalization may be the best way to go.

(C) If checked, p-values will be adjusted for multiple testing using the Benjamini-Hochberg method

(D) Clicking on the column header (F), i.e., the sample name will instantly color the phenotypes with their estimated levels (blue for negative, red for positive values) on the submaps if the p-value is lower than the supplied value.

(E) Checkbox to show the phenotype levels in a graph below (Figure 13) for up to 10 phenotypes.

(F) A column for each sample containing the estimated value (between -1 and 1) and the p-value for each phenotype in parentheses. Clicking on a value will show a scatter plot in a popup window with detailed information on the phenotype’s

regulators, their fold change in the data, and their influence on the phenotype as well as information on statistics (Figure 12).

Figure 12

(G) The percentage of phenotypes regulators considered as DCEs in all samples weighted by their influence score.

(H) The number of regulators that are DCEs in as mean [+ std. dev.] among all samples

(I) Top5 regulators among all samples with the most impactful combined fold change in the data and influence on the phenotype

(J) All results can be downloaded as a tab-separated text file containing estimated levels and p-values of all phenotypes.

Results Graph

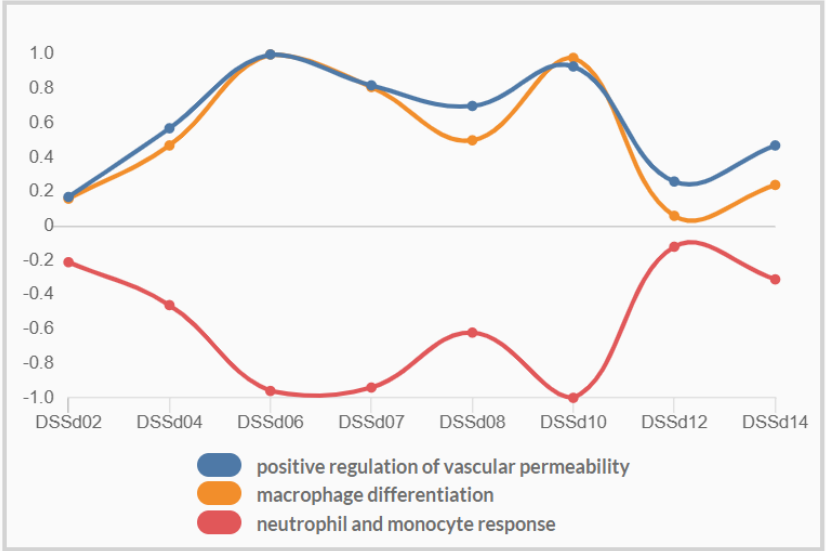

Figure 13

A line plot to provide a graphical overview of phenotype levels in all samples. Phenotypes to show in the graph can be selected in the results table (above).

Regulator Ranking

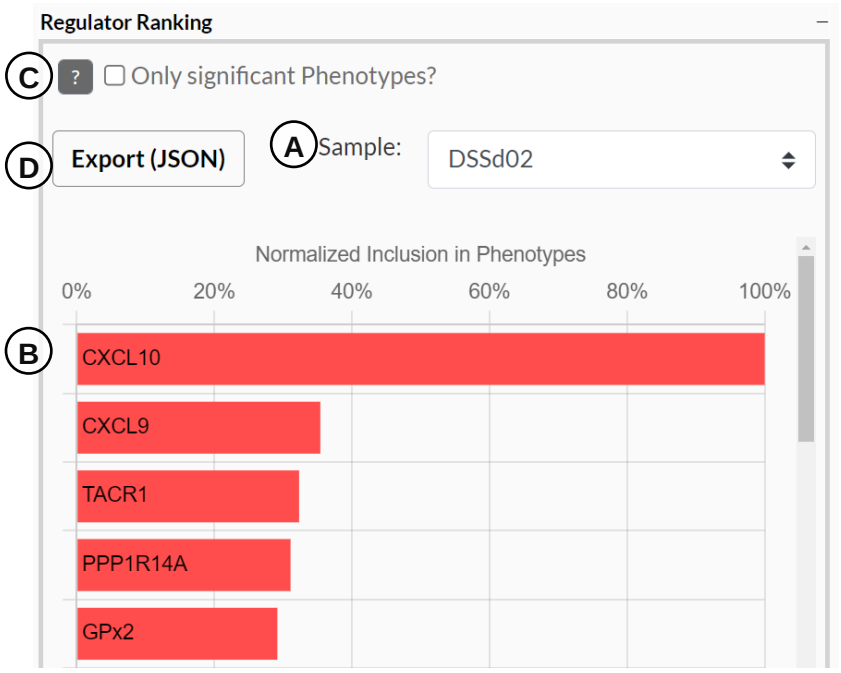

Figure 14

As the name implies, the regulator ranking panel gives an overview of how strong elements in the network contributed to the results, based on their influence scores and fold change values. The overview is for one sample only, which can be selected by the user (Figure 14A). The ranking is presented as a horizontal bar plot, with one bar for each element, showing in percent their relative importance (Figure 14B). Additionally, it can be specified whether only significant phenotypes (based on the p-value threshold defined in the table panel) will be considered. Finally, the result can be downloaded in JSON format (Figure 14D).

Highlight on Map

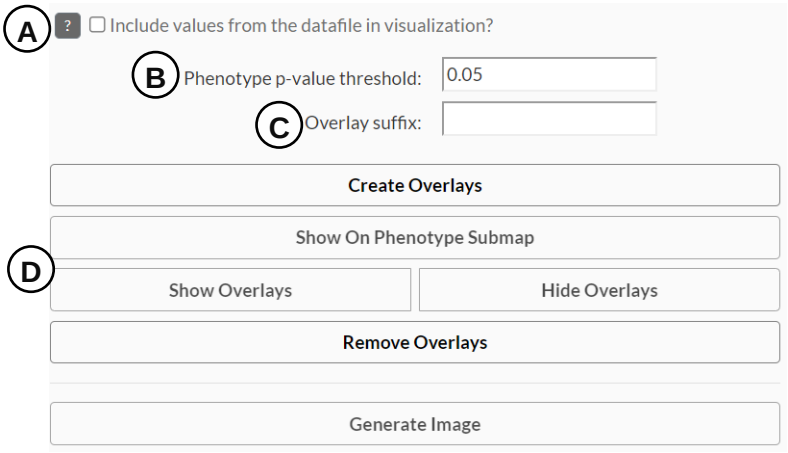

Figure 15

The third panel allows users to customize the visualization of the estimated phenotype levels as overlays in MINERVA, with the possibility to visualize the fold change values from the data (Figure 15A). By setting a phenotype threshold, the user can decide which phenotypes to include in the

overlays (Figure 15B). The name of the overlay is equal to the sample name in the data, with a user-specified suffix to be able to distinguish multiple analyses (e.g., with different p-value thresholds) from another (Figure 15C). There are buttons to hide, show or delete all overlays automatically (Figure 15D).

Target Inference

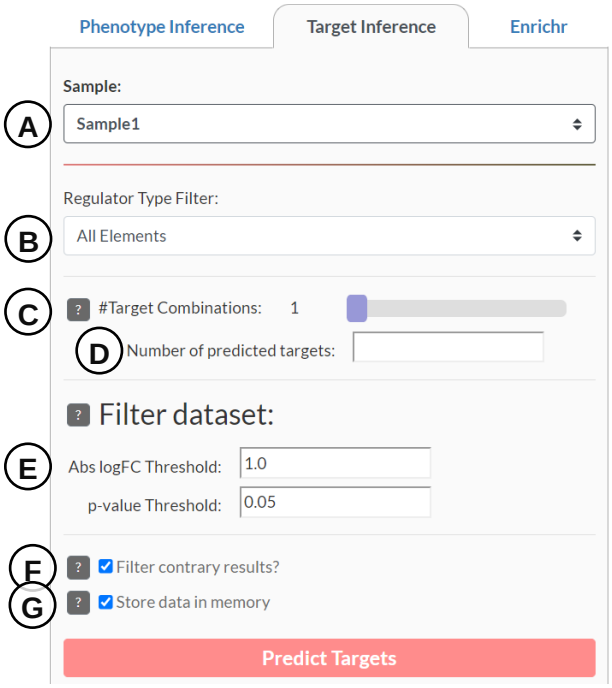

Figure 16

This tool predicts elements in our dataset that may be the most probable regulators for the observed changes (i.e., FC values) of the data sample using network topology features,. Highly potential targets interact with elements in the dataset the same way as their fold change (positive target) or the opposite way (negative target). To perform the analysis, the user selects the sample (Figure 16A) and, if desired, filters the target by their molecule type (e.g., transcription factor, miRNA, ...) (Figure 16B).

The target inference not only predicts single targets but is also able to analyze combinations of up to four different targets (Figure 16C). However, because this strongly impacts the calculation time, the number of targets used to iterate through the combinations needs to be capped (Figure 16D). The number supplied here is the number of significant targets with the highest sensitivity value from the single target inference parsed to the combinatory analysis. The supplied DCEs can be filtered by their fold change and p-value (Figure 16E). Additionally, targets with different signs of their FC in the data compared to their sensitivity can be filtered out (Figure 16F). Because the target inference tool requires fetching many data for the analysis, we provide the option to not store this data in memory for the time of plugin use (Figure 16G).

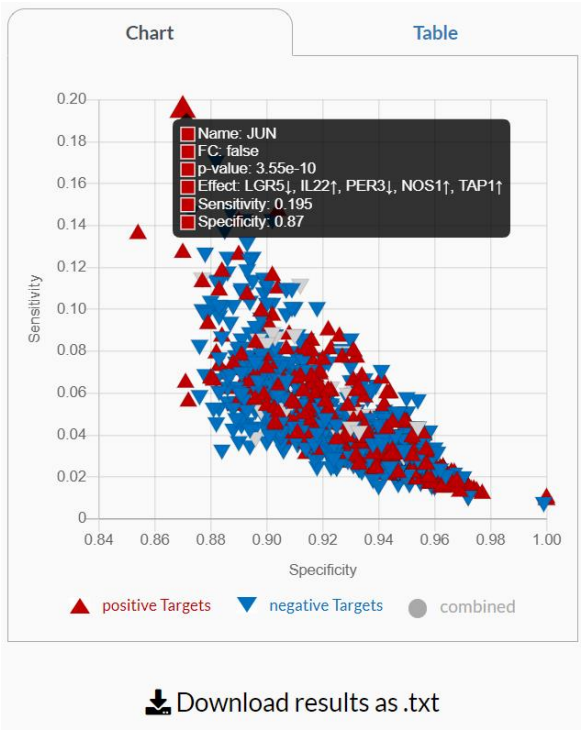

Figure 17

The predicted regulators are then displayed in a scatter plot (Figure 17) by their sensitivity (ability to regulate the elements in the data file as their FC values describe them) and specificity (inability to regulate elements with no or a zero FC value). The colors of the regulators are based on their type of regulation: *positive*, if the element induces the FCs, and *negative* if it induces the exact opposite changes. Elements marked as '*external*' are not included on the maps and will link to their entry on public databases.

References

[1] C. N. Serhan *et al.*, “The Atlas of Inflammation Resolution (AIR),” *Mol. Aspects Med.*, vol. 74, p. 100894, Sep. 2020, doi: 10.1016/j.mam.2020.100894.

[2] S. Köhler *et al.*, “The Human Phenotype Ontology in 2021,” *Nucleic Acids Res.*, vol. 49, no. D1, p. D1207, Jan. 2021, doi: 10.1093/NAR/GKAA1043.

[3] T. U. Consortium *et al.*, “UniProt: the universal protein knowledgebase in 2021,” *Nucleic Acids Res.*, vol. 49, no. D1, pp. D480–D489, Jan. 2021, doi: 10.1093/NAR/GKAA1100.

[4] X. Watkins, L. J. Garcia, S. Pundir, M. J. Martin, and U. Consortium, “ProtVista: visualization of protein sequence annotations,” *Bioinformatics*, vol. 33, no. 13, pp. 2040–2041, Jul. 2017, doi: 10.1093/BIOINFORMATICS/BTX120.