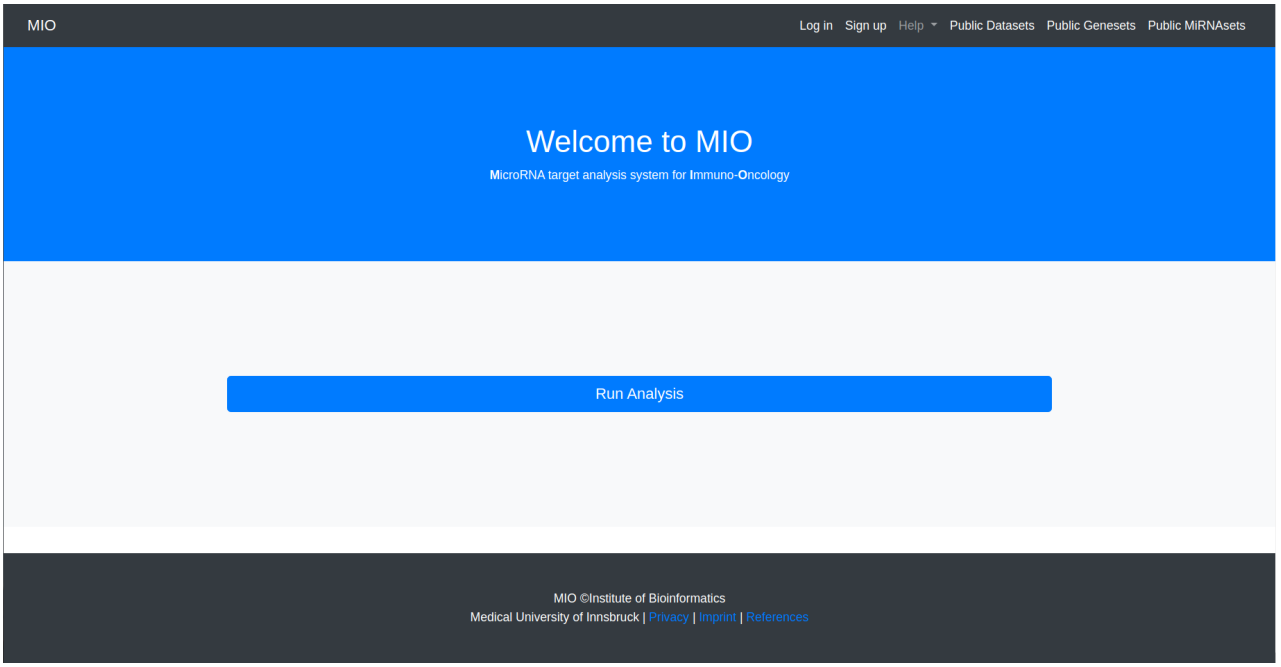
MIO

miRNA target analysis system for Immuno-Oncology



MIO Web portal: http://mio.icbi.at
Version: 1.0 (September, 2021)
INTRODUCTION

MicroRNAs (miRNAs, mir) have been shown to be able to modulate the tumor microenvironment and the immune response and hence could be interesting biomarkers or targets for therapeutic concepts in immune-oncology. A variety of systems and tools for miRNA target prediction even in the context of cancer have been developed, however, an integrated system taking the complex interactions between tumor and immune system into account is lacking.
In order to identify miRNA target genes of immune-related signatures and pathways in cancer samples, we have developed the user-friendly web platform MIO and the Python toolkit miopy integrating various analyses and visualization methods based on provided or custom bulk miRNA and gene expression data. We include regularized regression and survival analysis and provide information of 40 miRNA target prediction tools as well as a collection of curated immune-related gene signatures and analyses of >30 cancer types based on data from The Cancer Genome Atlas (TCGA). MIO allows the identification of genes affecting the vulnerability of cancer based on synthetic lethal interactions of miRNA target genes. We also integrated several machine learning methods to enable the selection of prognostic and predictive miRNAs and gene interaction network biomarkers. In MIO, users can generate testable hypotheses and identify miRNAs together with their target genes, which may play a potential role as biomarkers or represent direct or susceptible candidates for cancer immunotherapy.

Availability
MIO http://mio.icbi.at, miopy toolkit https://github/icbi-lab/miopy

1. Getting started

The Django web framework was used to develop the MIO website. MIO is the web portal to identify miRNA target genes of immune-related signatures and pathways in cancer samples.

1.1 The Home Page

As shown in Figure 1 the homepage presents the main menu bar that allows users to log in or log out. Through the user tab, in the main menu bar, users can modify their own data, and upload their own dataset, miRNA set, and gene set, to use in their own analysis, or download all the data on the server. In the main menu bar, we find the tab Public Dataset and Public Geneset, these two tabs allow any user to access the curated data present in the tool. The last tab present in the main menu bar is the Help button, this is a dropdown menu where users can access the tool manual, the Github repository, and the FAQ.

In the middle of the screen appears the Run Analysis button, which redirects the user to the Analysis index view.

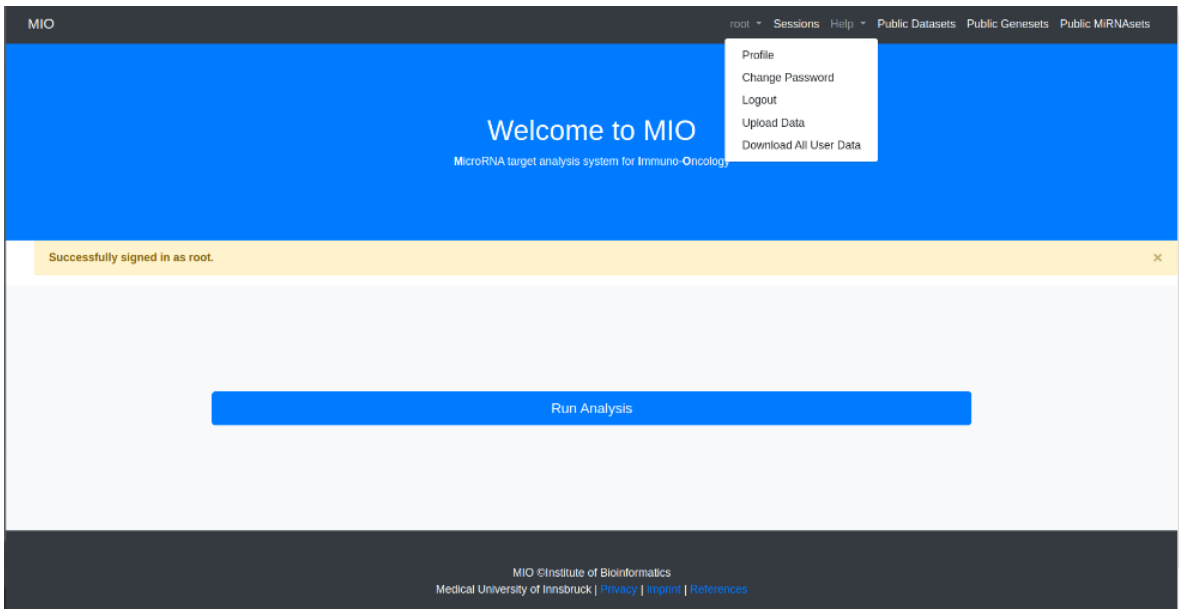


Figure 1. Index view from MIO.

1.2 The Public Genesets Page

Figure 2 shows the Public Geneset page. This page shows the list of every curated gene set available in the tool. The gene set is displayed as a table, where the user can find the name, a brief description, and the number of genes. The user can download one specific gene set, with Gene Symbol or Gene ID, with the download button. Any curated gene set can be

downloaded in GMT format by clicking the buttons below the table. Finally, with the Search box, the user can filter by specific words present in the gene set descriptions.

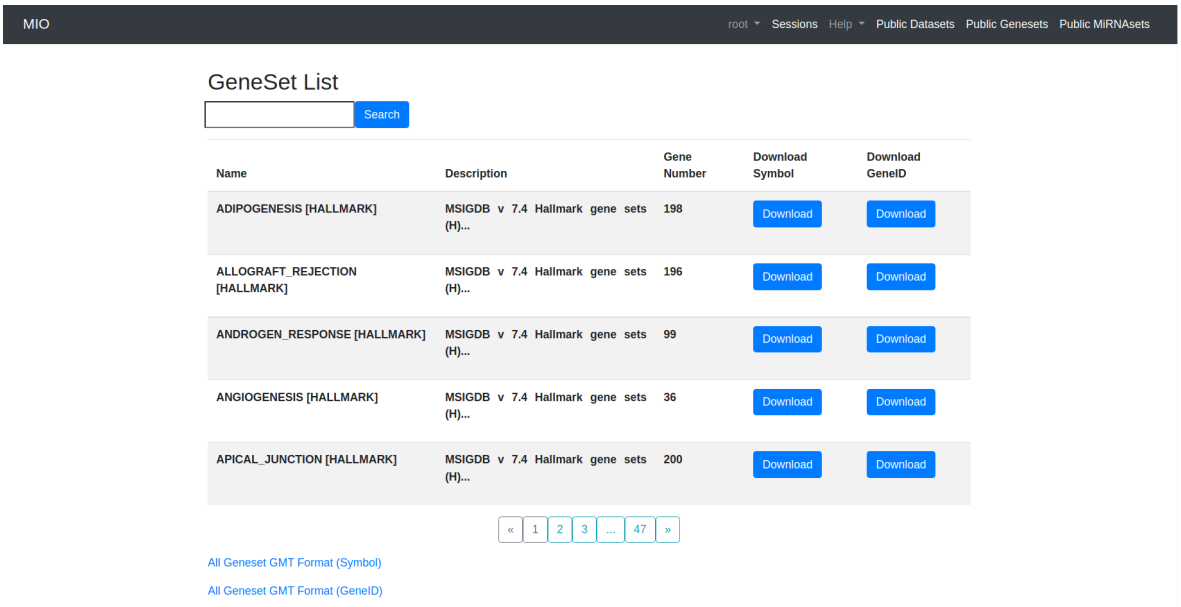


Figure 2. List view of all the public gene sets stored in MIO.

1.3 The Public MiRNAsets Page

Figure 3 shows the Public MiRNAsets page. This page shows the list of every curated miRNA set available in the tool. The miRNA set is displayed as a table, where the user can find the name, a brief description, and the number of miRNAs. The user can download one specific miRNA set, with mirbase accession or mirbase id, with the download button. Any curated miRNA set can be downloaded in GMT format by clicking the buttons below the table. Finally, with the Search box, the user can filter by specific words present in the miRNA set descriptions.

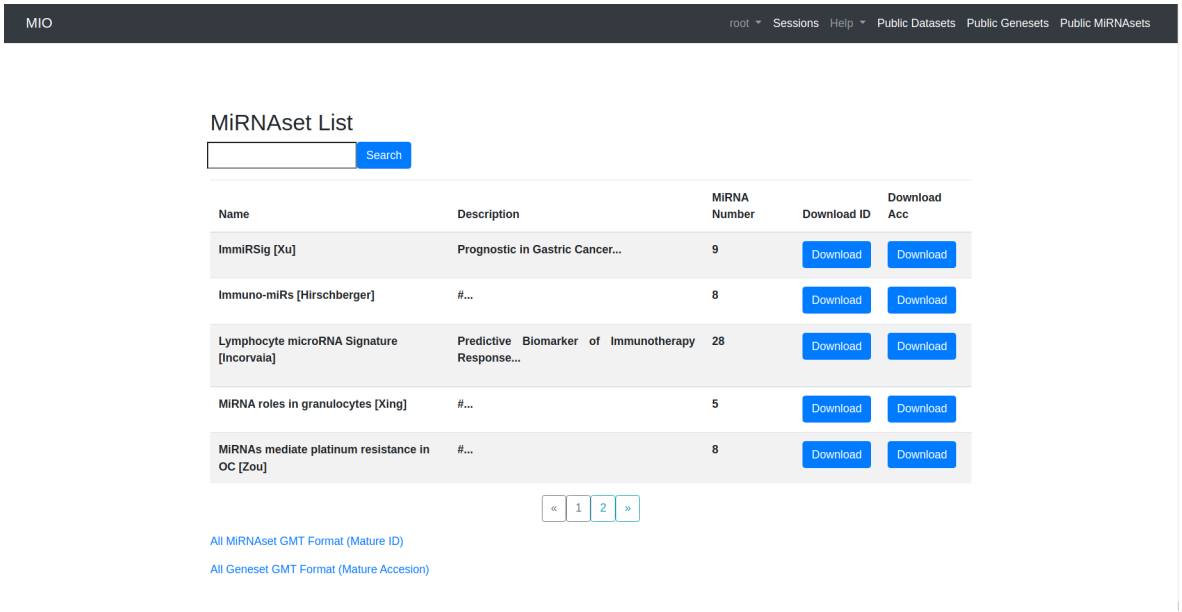


Figure 3. List view of all the public miRNA sets stored in MIO.

1.4 The Public Dataset Page

Figure 4 shows the Public Dataset page which lists all of the available datasets. The datasets are displayed as a table, where the user can find the name, the sequencing technology (sequencing or microarray), the number of genes, miRNAs (Mir) and samples, and the columns present in the metadata. Then, with the Search box, the user can filter by specific words present in the dataset description.

Finally, the user can view one specific dataset, with the view button. Figure 5 shows the Dataset View page. The user can find which data is stored in the metadata, and the name of the sample used in the analysis. The user can download all the datasets.

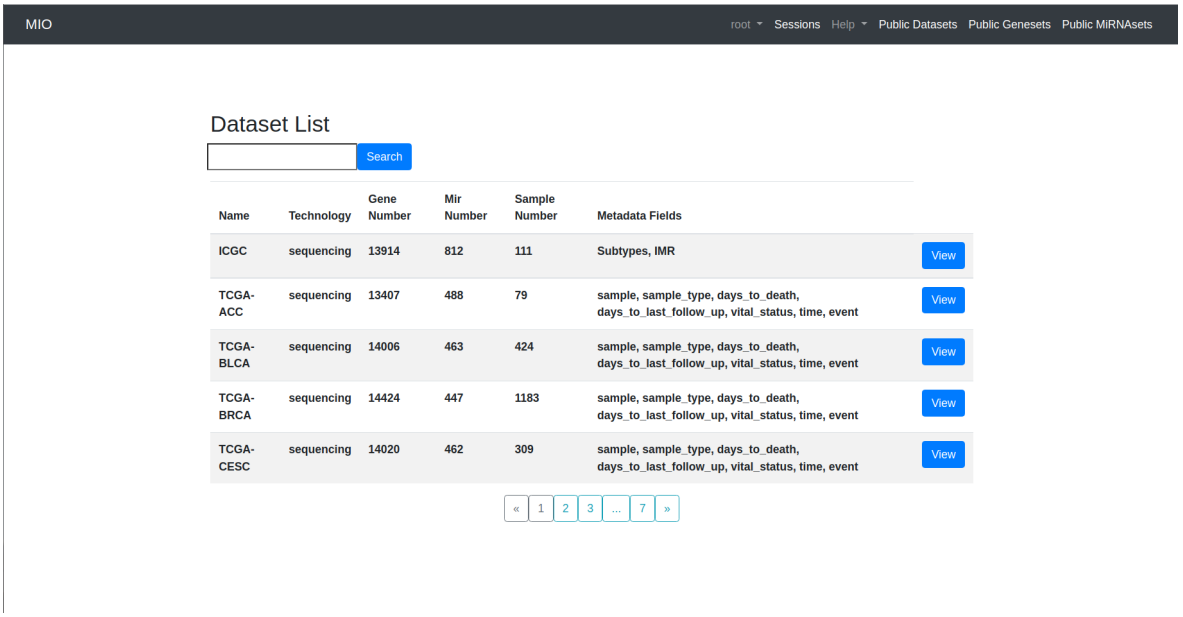
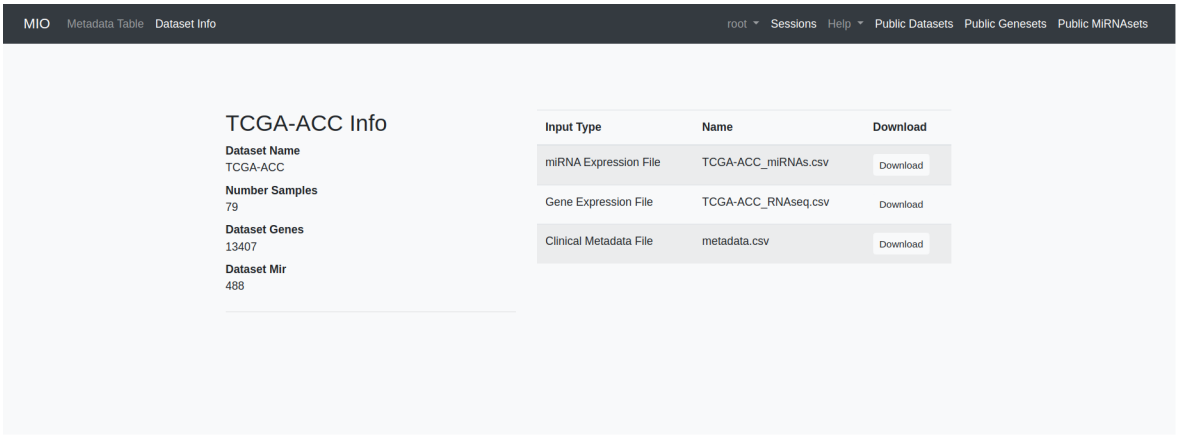


Figure 4. List view of all the public datasets stored in MIO.



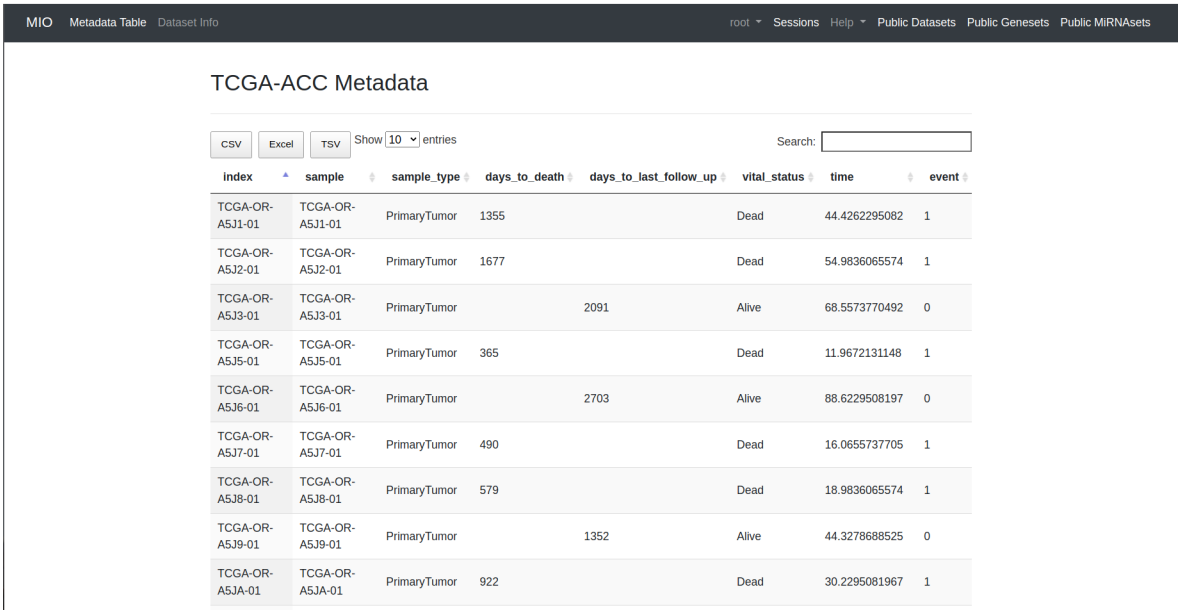


Figure 5. Dataset view of MIO. In the first image, the view shows a brief summary of the dataset, the links from the different files: miRNA expression set, gene expression set, and the metadata. The second image shows the metadata of the dataset.

1.5 The Analysis Index Page

Figure 6 illustrates the Analysis Index view. This view provides the user the choice to start a new analysis or select a previously submitted analysis. In the available analysis box, the user can choose between the pre-calculated analyses provided by the tool or the user-submitted analyses. Once the user has selected the analysis, they will be redirected to

the Analysis Details page by clicking the “View Analysis“ button. In order to create a new analysis, the user can click the “New Analysis“ button.

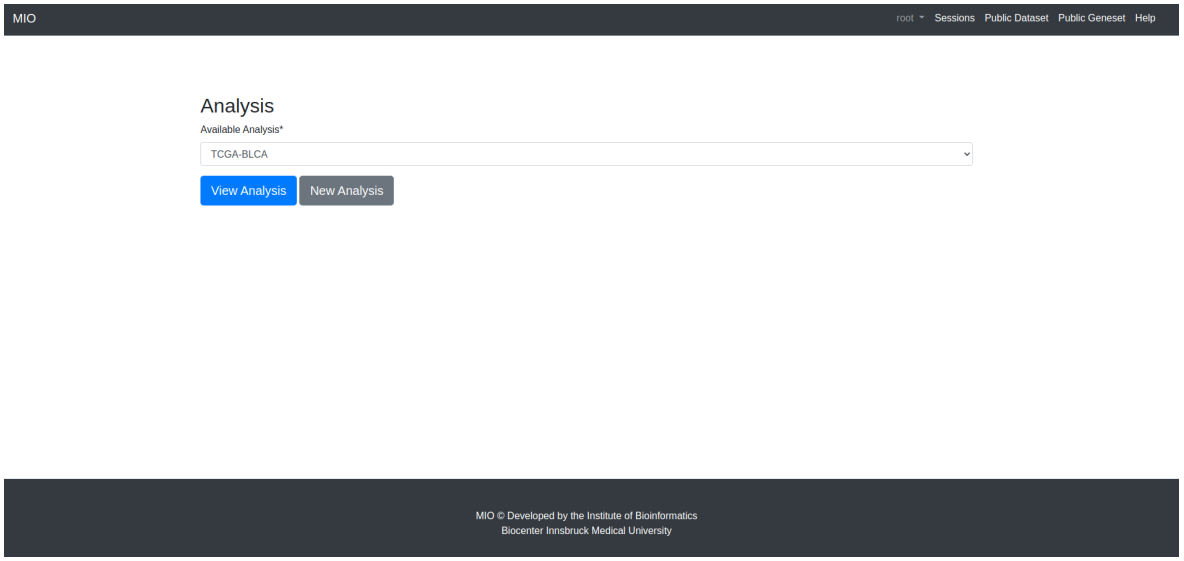


Figure 6. Analysis list view of MIO. In the select menu all the analyses available on the web are shown.

1.6 The Analysis Detail Page

On the analysis detail page, the user can choose between running a Correlation Analysis, a Target Prediction Analysis, a Survival Analysis, and a Classification Analysis. When a user runs one analysis, this appears in the Analysis List as can be seen in Figure 7. There is no limit to the number of analysis pages, or the number of submitted analyses for any user.

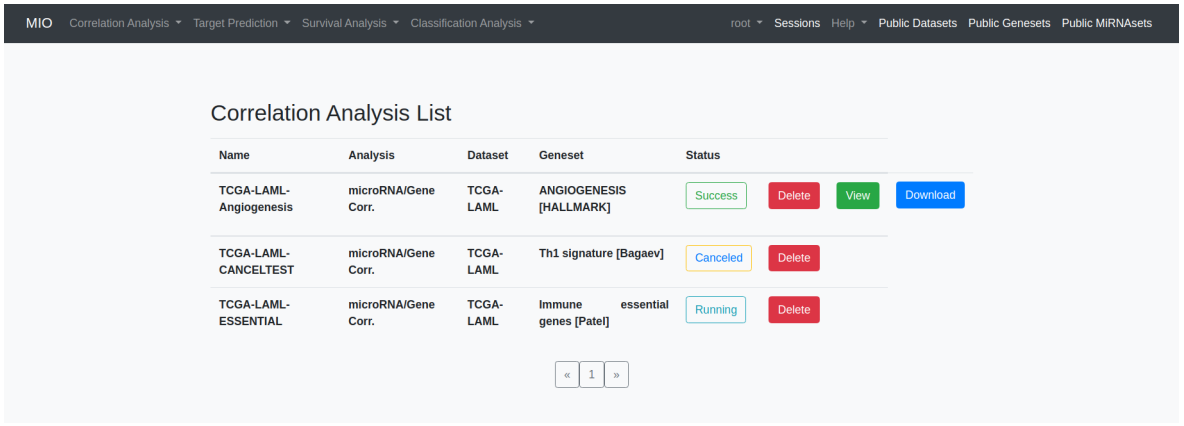


Figure 7. Analysis list view of the running, successful and canceled analyses of the user.

In Figure 7, the user can also find the main top bar with the different available analyses in the tool. Each tab is a dropdown menu with the different workflows for each analysis. For the correlation, the user can choose between miRNA/Gene Correlation or miRNA/Geneset Correlation. The first option obtains different correlation coefficients for each miRNA-gene pair. The second one obtains the module score for the selected gene set for each sample. This module score is used to calculate the Pearson coefficient for each gene set/miRNA.

In the second tab, Target Prediction, the user can choose between miRNA/Gene Target Prediction or miRNA Synthetic Lethal Prediction. The Target Prediction allows the user to determine the best match for one gene or miRNA. The synthetic lethality prediction allows the user to obtain the miRNA that targets different synthetic lethal genes for a given gene.

In the Survival Analysis tab, the user can choose between Kaplan Meier plot, Feature Selection, or miRNA/Gene Ratio feature selection. The Kaplan Meier plot generates the survival plot to a given miRNA, gene, or miRNA/Gene ratio, splitting the sample into higher and lower expressions. The Feature Selection uses a machine learning ensemble method to select the most important gene, miRNA, or both in the patient overall. The miRNA/Gene Ratio Feature Selection uses the miRNA/gene ratio, which is an approximation of the pair interaction, as a feature in the survival model.

Finally, the last tab is the Classification Analysis. The Classification Analysis uses a machine learning model to test one gene set or miRNA set as a predictor in a binary classification. The Feature Selection uses a machine learning ensemble method to select the most important gene, miRNA, or both in the binary classification. The miRNA/Gene Ratio Feature Selection uses the miRNA/Gene ratio, which is an approximation in the pair interaction, as a feature in the classification model.

In the middle of the page, the user can see the different submitted analyses. An analysis may have one of several states:

- Queue, if the analysis is waiting to be processed
- Running, if the server has started to analyze the job
- Success, if the analysis ended without problems
- Canceled if problems occurred during the analysis or the user interrupts it.

The “View” and “Download” buttons will appear next to the analysis name. With the download button, the user will be able to download the analysis result, and with the view button, the data will be visualized.

2. Data Upload

2.1 Upload a Dataset

The user can upload their own data to the server in two different ways in the Upload Data menu or during job submission. The Upload Data is part of the user menu in the main top bar as depicted in Figure 8.

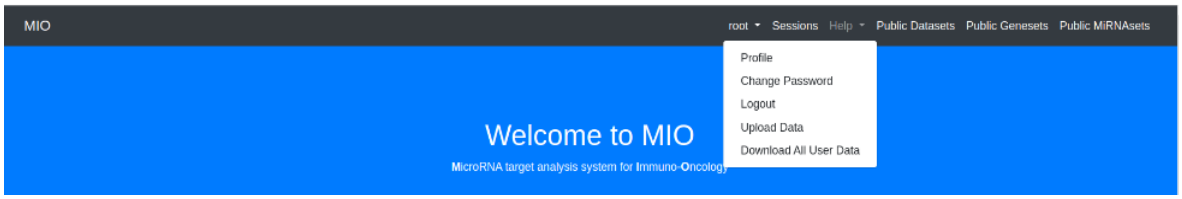


Figure 8. User top bar menu.

The user will be redirected to the User Data Page (Figure 9). In the User Data Page, the user can manage their own gene sets, miRNA sets and datasets. To upload a new dataset, the user has to click the Upload Dataset button. When a user uploads a Dataset, it is necessary to assign a name to the dataset, the technology (sequencing or microarray), the metadata, and the miRNA/gene expression table, as can be seen in Figure 10.

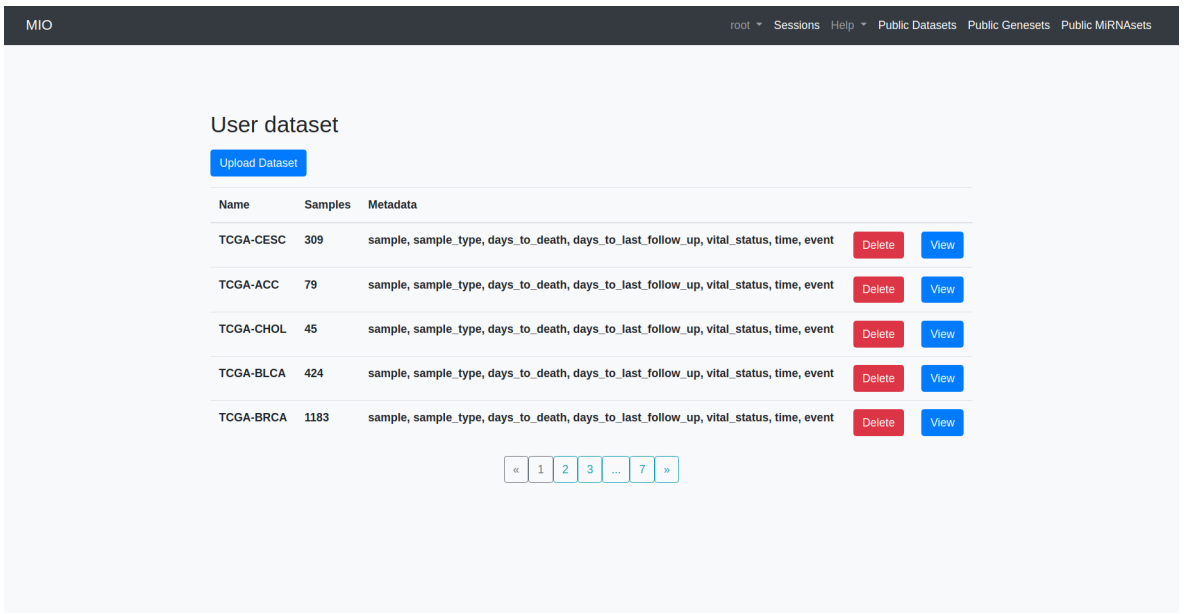


Figure 9. User data view.

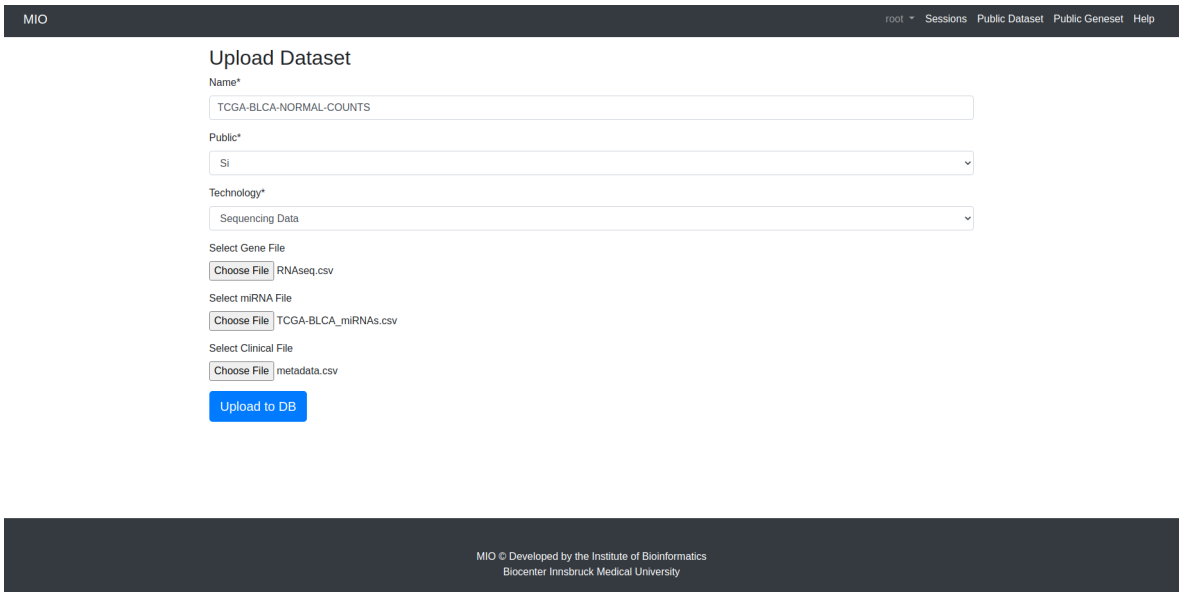


Figure 10. Upload dataset view.

Data Format

The expression files may be of a csv, tsv, or txt extension. It should be a tab-separated file, where the first row pertains to the samples in the data, and the first column pertains to the Gene Symbol in the gene expression file, or the mature mirbase id in the miRNA expression file as depicted in Figure 11.

	TCGA-AB-2805-03	TCGA-AB-2806-03	TCGA-AB-2808-03	TCGA-AB-2810-03	TCGA-AB-2811-03	TCGA-AB-2812-03	TCGA-AB-2813-03	TCGA-AB-2814-03	TCGA-AB-2815-03	TCGA-AB-2817-03
DPM1	4.725781	4.583641	4.420862	4.994760	4.766534	4.837009	4.707247	4.674473	4.553948	4.045273
SCYL3	4.757294	4.793835	4.626136	3.452037	4.071263	4.709653	4.634654	4.450155	4.878366	4.628749
C1orf112	4.103280	5.533229	4.529604	3.405279	2.601448	4.264259	4.365979	3.879948	5.174730	4.229162
FGR	8.673282	3.793228	7.217589	6.392921	9.990586	4.981187	9.477212	5.982639	8.010053	7.547292
CFH	0.469518	-0.537233	2.933139	-1.419150	0.960060	5.335964	4.007778	-3.028277	1.692433	5.299988
FUCA2	4.327956	5.204462	5.282520	5.327700	5.826520	4.994576	5.862225	4.169212	5.855972	4.175871
GCLC	5.013650	6.051037	6.166010	5.265482	5.350027	5.985552	5.425442	5.627949	6.112084	5.663526
NFYA	5.648606	6.988851	6.416664	6.072169	4.051145	6.010152	6.026928	5.761437	5.918593	6.434545
STPG1	2.197189	-1.210423	2.225567	2.586849	2.213817	2.800121	1.889494	2.681381	2.365213	2.577272
NIPAL3	5.459974	4.961126	6.157033	5.401029	4.415001	5.702736	5.476990	6.748887	5.727836	5.860460

Figure 11. Gene expression matrix from the TCGA-LAML dataset.

The metadata file may be a csv, tsv, or txt extension. As shown in Figure 12, it should be a tab-separated file, where the first row pertains to the variables in the data, and the first column pertains to the sample name. The names in the metadata have to be identical to the ones in the expression file.

	event	time	Fab_Category
TCGA-AB-2802-03	1	365.0	M4
TCGA-AB-2803-03	1	792.0	M3
TCGA-AB-2805-03	1	577.0	M0 Undifferentiated
TCGA-AB-2806-03	1	945.0	M1
TCGA-AB-2807-03	1	181.0	M1
TCGA-AB-2808-03	0	2861.0	M2
TCGA-AB-2810-03	1	31.0	M2
TCGA-AB-2811-03	1	243.0	M4
TCGA-AB-2812-03	1	366.0	M2
TCGA-AB-2813-03	1	31.0	M4

Figure 12. Metadata table from the TCGA-LAML dataset.

2.2 Upload a Gene set

After clicking the “Upload Data” option, present in the user menu, in the main top bar, the user can upload their gene set on the User geneset page as shown in Figure 13. The user can upload only one gene set at a time using this method, however several gene sets may be uploaded simultaneously, using one GMT file with the Upload GMT button.

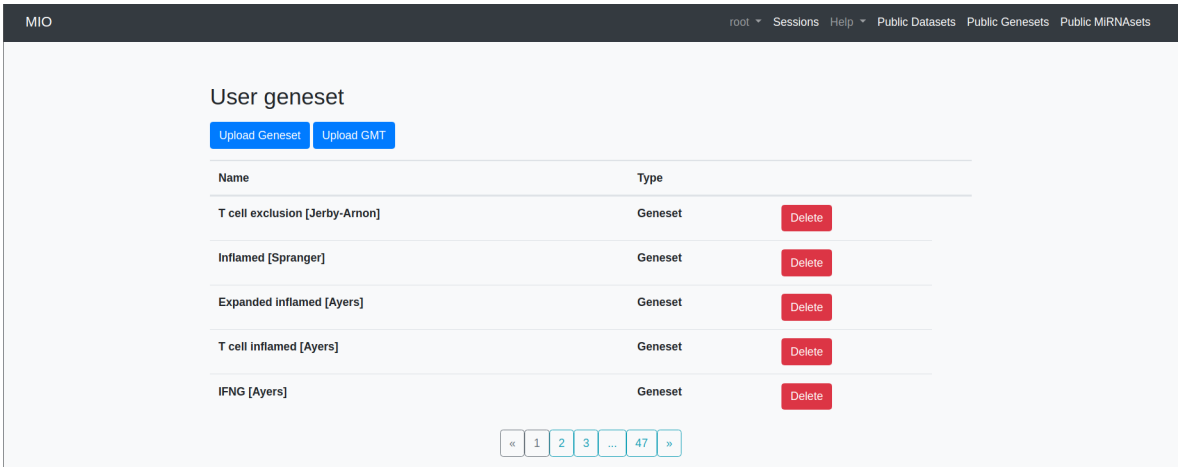


Figure 13. Upload geneset view.

Upload a Gene set

To upload a gene set, the user has to fill in a name, a description, a reference of the gene set, indicate the identifier of the genes (Symbol, or Gene Id), and choose a txt file. The file with the genes should contain one gene per row.

Upload a GMT File

To upload a GMT file, the user only needs to indicate the identifier of the genes (Symbol, or Gene Id) and choose a GMT file. The GMT file format is a tab-delimited file format that describes gene sets. In the GMT format, each row represents a gene set. Each gene set is described by a name, a description, and the genes in the gene set.

2.2 Upload a miRNA set

After clicking the "Upload Data" option, present in the user menu, in the main top bar, the user can upload their miRNA set, as depicted in Figure 14. The user can upload only one miRNA set at a time using this method.

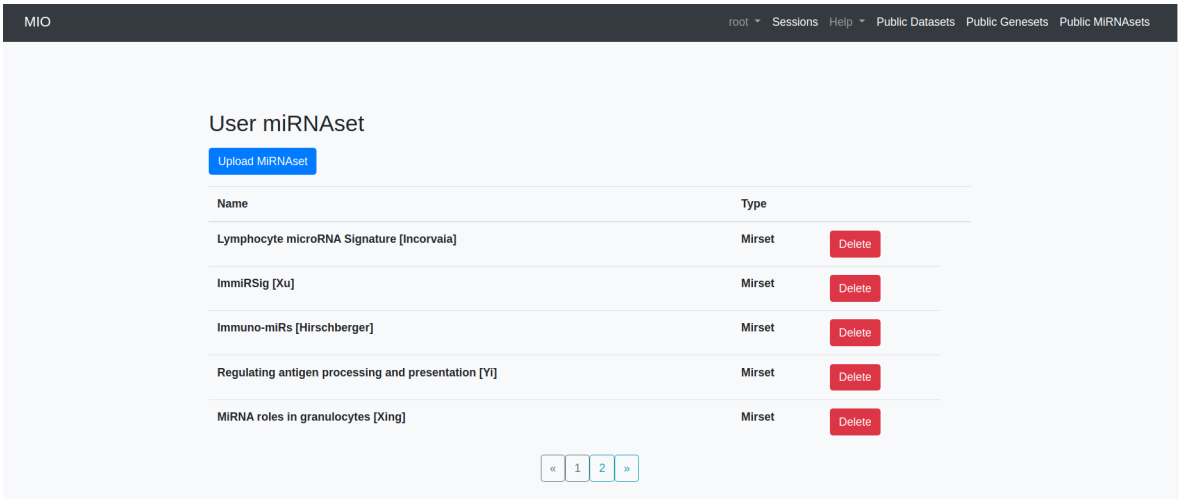


Figure 14. Upload miRNAset view.

Upload a miRNA set

To upload a miRNA set, the user has to fill in a name, a description, a reference of the miRNA set, indicate the identifier of the genes (miRBAse Accesion, or miRBAse Id), and choose a txt file. The file with the genes should be one gene per row.

3. Submit an analysis

3.1 Submit a mir/gene correlation analysis

When an analysis is created, the user can access the Analysis View. To run a mir/gene Correlation analysis, the user has to access the miRNA/Gene Correlation Analysis present in the Correlation tab.

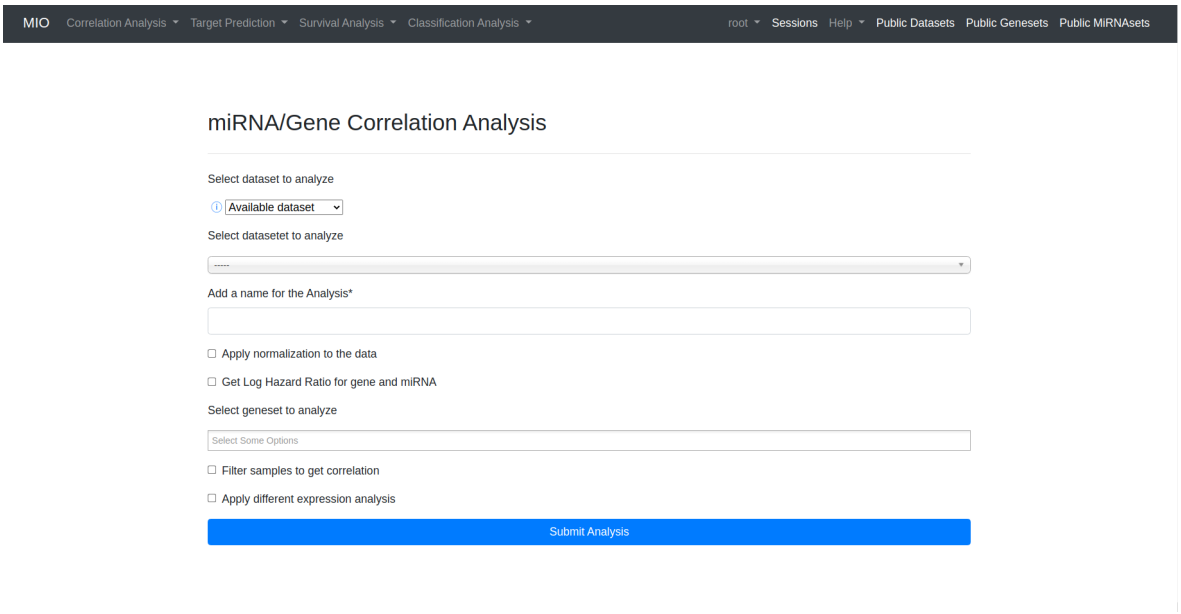


Figure 15. miRNA/Gene correlation analysis form.

In the miRNA/Gene Correlation Analysis Page, which is depicted in Figure 15, the user can choose between using an available dataset (public datasets in the tool, and datasets uploaded by the user) and uploading a new dataset. In the select available dataset field, the user can select the available dataset in the dataset list. Then the user needs to assign a name for the analysis. If the data provided is not normalized the user can click the "Apply Normalization to the Data", and the tool will use the Voom normal normalization to the data. If the survival data (time and survival status) is provided in the metadata, the user can check the "Get Log Hazard Ratio for Gene and miRNA". In this case, the tool will obtain the $\log_2$(HR) splitting the gene/miRNA ratio into higher and lower. To restrict the number of genes used to search for miRNAs, the user can choose a different gene set from the list.

In general, different conditions may be present in the same expression set, for this reason the user can check the "Filter Samples to get correlation". If it is checked, two new fields will appear and the user needs to indicate which variable and which group from the metadata will be used in the analysis.

In order to filter the genes and miRNAs, and keep only the differentially expressed genes and miRNAs, the user can apply the DEA check box “Apply differential expression analysis”. In these cases, the user can select if the analysis should be paired or unpaired, which column in the metadata is used in the analysis, and which adjusted p-value and $\log_2$ fold change are used to filter the genes and miRNAs. Only genes and miRNAs fulfilling the defined criteria will be used for further analyses.

When the analysis has been submitted the name of the analysis will appear on the Analysis View Page.

Visualize the results

When the analysis ends, the status will change from “Running” to “Success”, in the Analysis View page. The user can download the result by clicking the “Download” button. To filter and visualize the result with the tool, the user should click the “View” button, which will redirect them to the filter form, as can be seen in Figure 16. In the filter table view, the user can choose the adjusted p-value and the cut-off value for the coefficients (higher and lower). Finally, the user can choose a combination of 40 different Target Prediction Tools to filter the miRNA/Gene pairs.

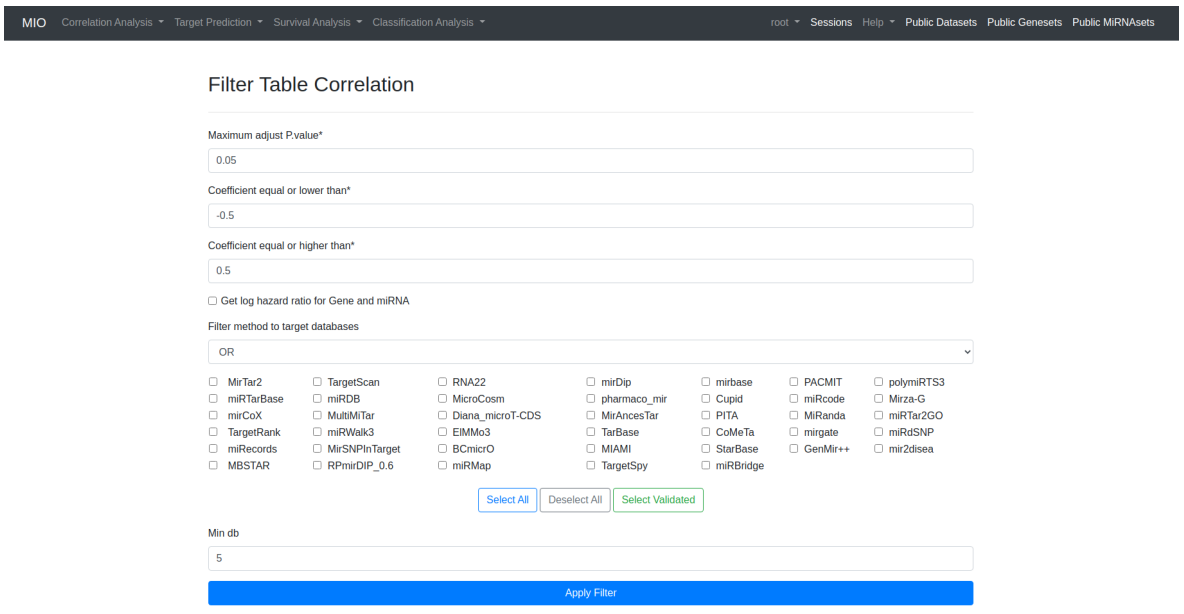


Figure 16. Filter miRNA/Gene correlation result from view.

The results are then shown in the Result Page view. All the miRNA/Gene pairs that passed filtering are displayed in a table. The user can download this table in Excel, csv, and tsv format. The fields in the table are:

- Borda: This is a ranked voting system. Borda integrates all the correlation methods used in the analysis, to sort the miRNA/Gene pairs.
- Gene: Gene Symbol ID

- miR: Mature mirbase ID.
- Lasso, Ridge, and Elastic net: These are regression analysis methods. Lasso performs the L1 regularization technique which allows a feature selection. Ridge regression performs the L2 regularization technique. In this case, the coefficients might be really small but never 0. Elastic next is a mix of both regularization techniques. In all the models, each gene is presented as a linear expression of all miRNAs, and the coefficient of a miRNA in the regression model is used as the association strength between the miRNA and the mRNA.
- Pearson (R), Spearman (Rho), Kendall (Tau), Hoeffding, and RDC. These correlation methods are used to analyze the relationship between one miRNA and one gene. Pearson's method measures linear dependency. If there is no linear relationship, Spearman and Kendall methods are recommended. These methods return a value between (-1,1). The RDC is a measure of nonlinear dependence between two variables. Finally, Hoeffding's D statistic provides a test for independence.

The second plot created is a heatmap. The heatmap is created with Plotly and is an interactive plot. The rows are miRNAs, and the columns are genes. The color of the heatmap depends on the Pearson Correlation coefficient. Genes and miRNAs are clustered to facilitate interpretability by the user.

Finally, a network plot is created using the Pearson Correlation coefficient. The network is created with Cytoscape and is interactive. The color of the node depends on the $\log_2$(HR)(orange, violet), the color of the edge depends on the correlation coefficient (red, blue), and the width on the absolute value of the correlation coefficient (1px,10px).

3.2 Submit a mir/gene set correlation analysis

When an analysis is created, the user can access the Analysis View. To run a miRNA/Geneset correlation analysis, the user has to access the Correlation tab.

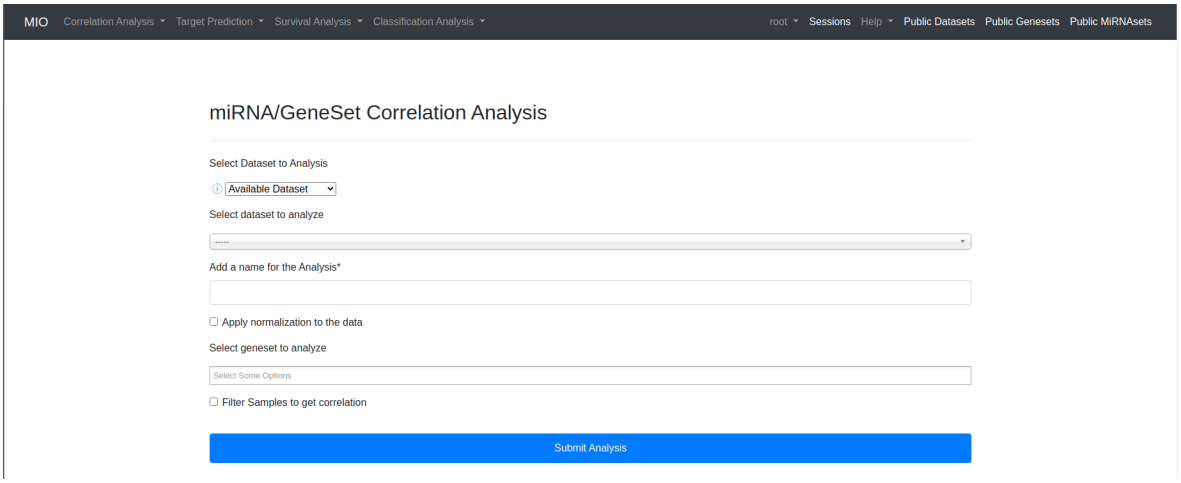


Figure 17. Run miRNA/Geneset correlation view.

In the miRNA/Geneset Correlation Analysis page, which is depicted in Figure 17, the user can choose between using an available dataset (public datasets in the tool, and the datasets uploaded by the user) and uploading a new dataset. When clicking the “Public Dataset” field, the user can select an available dataset from the dataset list. Then the user needs to assign a name to the analysis. If the data provided is not normalized the user can click the "Apply Normalization to the data”, and the tool will use the Voom normal normalization to the data. The user can choose different gene sets from the list. For the miRNA/gene set correlation, the module score will be calculated (Buffa et al. Cancer Res 2011). A module score is a number that summarizes the relation between each miRNA and all the target genes present in the geneset. The module score depends on the number of hits in the prediction databases.

In general, different conditions may be present in the same expression set, for this reason the user can check the “Filter Samples to get correlation”. When it is checked, two new fields will appear and the user needs to indicate which group from the metadata should be used in the analysis.

Visualize the results

When the analysis ends, the status will change from “Running” to “Success”, in the Analysis View Page. The user can download the result from the analysis by clicking the “Download” button. To filter and visualize the result with the tool, the user should click the “View” button. After clicking, the user will be redirected to the filter form which is shown in Figure 18. In the filter table view, the user can choose the adjusted p-value and the cut-off value for the coefficients (higher and lower).

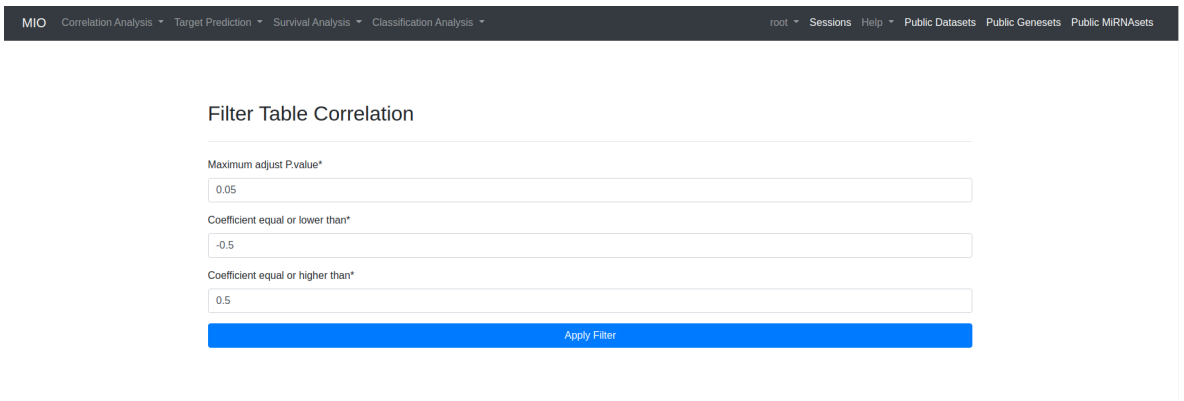


Figure 18. Filter miRNA/Geneset correlation result form.

Finally, the results will be shown in the Result Page View. The miRNA/gene set pairs which passed filtering are displayed in a table. The user can download the filter table in Excel, csv, and tsv format. The fields in the table are:

- Rank: The miRNA/gene set pairs, sorted by the absolute value of the Pearson Correlation
- Gene: Gene Set Name
- miR: Mature mirbase ID.
- Pearson (R): Pearson's method measures linear dependency

The second plot created by the view is a heatmap. The heatmap is created with Plotly and is an interactive plot. The rows are miRNAs, and the columns are gene sets. The color of the heatmap depends on the Pearson Correlation coefficient. Gene set and miRNAs are clustered to facilitate interpretability by the user.

Finally, a scatter plot is created using the Pearson correlation coefficient. The network is created with Plotly and is interactive. The user can choose between the gene set and the miRNA to visualize the correlation. In the scatter plot, each dot represents a patient. For each patient, the value of the module score and the expression of the miRNA is presented.

3.3 Submit a Feature Selection Analysis

Once the analysis view has been created, it can be accessed. The Feature Selection Analysis or Gene/Ratio Feature Selection Analysis is available in the Survival Analysis, and in the Classification Analysis. In both the analysis menu remains the same.

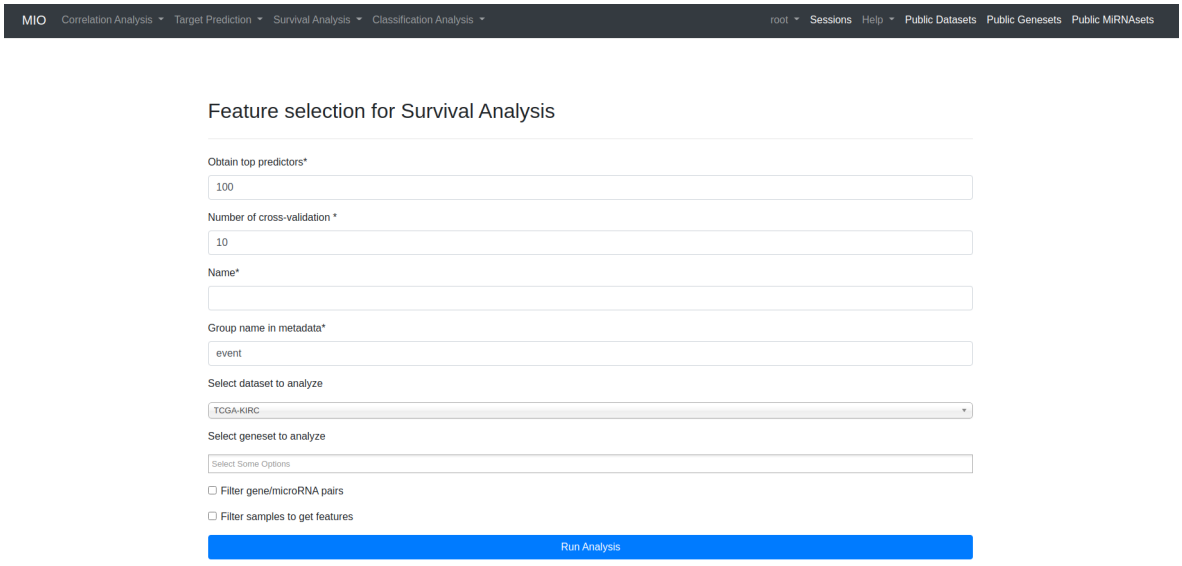


Figure 19. Gene/miRNA ratio feature selection form.

On the Feature selection page, which is depicted in Figure 19, the user can choose the dataset and a name for the analysis. The user needs to indicate which column of the metadata should be used for the classification. If a survival analysis is chosen, the user needs to specify the time column too. Then the number of top predictors has to be set, by

default this value is 100. In this case, the analysis will return the top 100 predictors for the classification, or for the survival prediction. The user needs to select the number of splits to use in the cross-validation. In order to obtain the most robust estimator, each model is run several times with different parts of the dataset to avoid overfitting. In each run, the model extracts the top features. If the analysis is a miRNA/Gene ratio, the gene sets need to be selected to reduce the computational requirements. In this case, the miRNA/Gene ratio is used as a feature. Finally, the analysis takes the most important common features in all the splits.

In general, different conditions may be present in the same expression set, for this reason, the user can check the "Filter Samples to get correlation". When it is checked, two new fields will appear and the user needs to indicate which group from the metadata should be used in the analysis.

Finally, for the mRNA/Gene Ratio Feature Selection to maintain the more relevant miRNA-Gene pairs the user can check the Filter gene/microRNA pairs. When it is checked, two new fields will appear and the user needs to indicate which is the maximum coefficient, and which is the lower number of prediction tools to allow for the miRNA/gene correlation filtering.

Visualize the results

When the analysis ends, the status will change from Running to Success, on the Analysis View page. The user can download the result from the analysis by clicking the "Download" button. To visualize the result the user should click the "View" button. After clicking, the user will be redirected to the result view.

All the kept features will be displayed as a table. The user is able to download the table in Excel, csv, and tsv format. In the table, the fields depend on the two types of analysis.

For the feature selection in the Classification Analysis:

- Percentage: This is the number of times this feature has appeared in all models with all partitions.
- Feature: This is the Gene Symbol, the miRNA mature mirbase ID, or the miRNA/gene.
- Random Forest: This is the percentage of times that this feature has appeared in the Random Forest (with 300 estimators) in all the partitions.
- Logistic Regression: This is the percentage of times that this feature has appeared in Logistic Regression (with L2 penalty and 10000 interactions) in all the partitions.
- Ridge Classifier: This is the percentage of times that this feature has appeared in Ridge Classifier (with 10000 interactions) in all the partitions.
- Support Vector Machine Classifier: This is the percentage of times that this feature has appeared in Ridge Classifier (with linear Kernel) in all the partitions.
- Ada Classifier: This is the percentage of times that this feature has appeared in Ada Classifier (with 300 estimators) in all the partitions.

- Bagging Classifier: This is the percentage of times that this feature has appeared in a Bagging Classifier (with 300 estimators) in all the partitions.

For the feature selection in the Survival Analysis:

- Borda: This is a ranked voting system. Borda integrates all the survival methods used in the analysis to sort the features.
- Percentage: This is the number of times this feature has appeared in all models with all partitions.
- Feature: This is the Gene Symbol, the miRNA mature mirbase id, or the miRNA/Gene.
- Gradient Boosted Models: This is the average of the importance of the feature in all partitions, for a Gradient Boosting Survival Analysis with 1000 estimators, and a learning rate of 0.5.
- Support Vector Machine Classifier: This is the average of the importance of the feature in all partitions, for a Fast Survival SVM with 1000 interactions.
- Penalized Cox: This is the average of the importance of the feature in all partitions, for a Cox net survival analysis with 1000 interactions.

The user can save these features like a gene set/miRNA set by clicking the "Save Feature as Gene/miRNAset" button. After clicking these features will be saved automatically. The name of the name set will be the name of the analysis. The user can access this new gene set/miRNA set on all the other analyses.

The second plot created by the view is a heatmap. The heatmap is created with Plotly and is an interactive plot. The rows are the models used, and the columns are the features. The color of the heat map depends on the percentage of times that this feature has appeared in the models for all the partitions. Finally, features are clustered to facilitate interpretability by the user.

The last plot in this view is the principal component analysis (PCA). It is created with Plotly and is an interactive plot. The PCA is created with the selected features. The PCA reduces all these features to only 3 dimensions. Each sample is colored using the label from the metadata.

3.4 Submit a classification analysis

Once the analysis view has been created, it can be accessed. The Classification Analysis is available in the Classification Analysis menu.

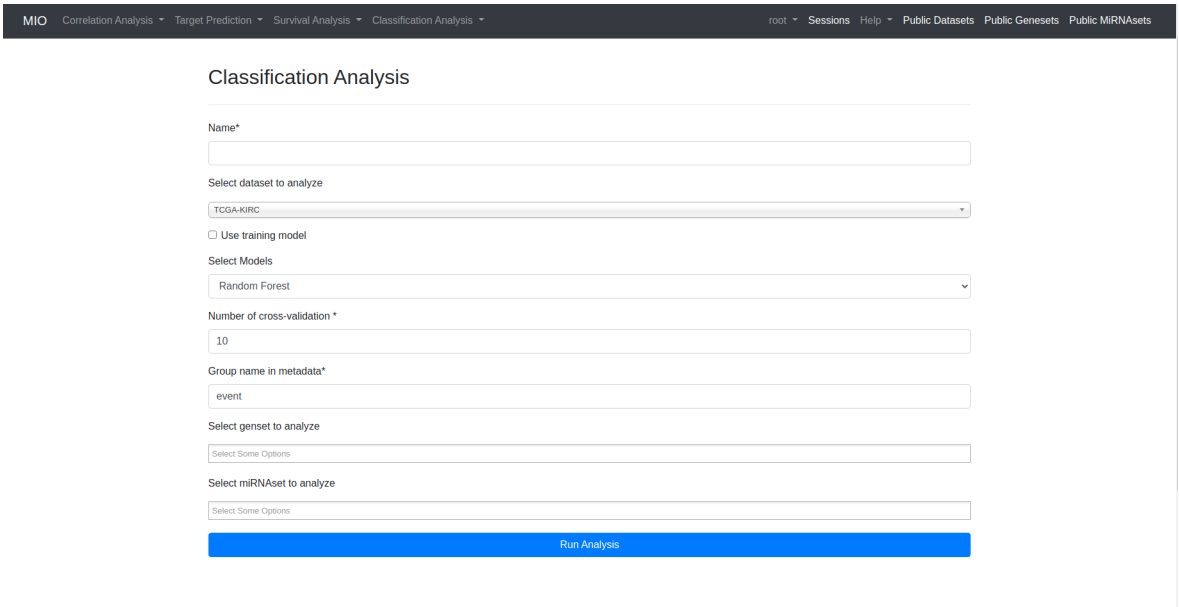


Figure 20. Classification analysis form view.

The Classification Analysis allows knowing the performance of the specific model with a specific gene set/miRNA set. In the Classification Analysis menu, which is depicted in Figure 20, the user needs to specify which column of the metadata is used for the classification. Then the user needs to select the number of splits to use in the cross-validation. In order to obtain the most robust model, the model is run several times with different parts of the dataset to avoid overfitting. The user can choose which gene set, miRNA set, or both are used to train the model. The models available in the tool are:

- Random Forest: Random Forest Classifier from scikit-learn, with 1000 estimators.
- Logistic Regression: A Logistic Regression from scikit-learn with L2 penalty and 100000 iterations.
- Support Vector Machine: A Support Vector Machine model from scikit-learn with a linear kernel.

MIO allows the use of a pre-training model in the Classification Analysis. The user can apply these models by clicking the "Use training model" button. When it is checked, one new field will appear and the user needs to indicate which model to use.

Visualize the result

When the analysis ends, the status will change from "Running" to "Success", on the Analysis View page. The user can download the result of the analysis by clicking the "Download" button. To visualize the result the user should click the "View" button. After clicking, the user will be redirected to the result view.

In the first part of the result view, the user gets a summary of the classification: The dataset used in the analysis, the gene set/miRNA set, and the performance in the train and test split for the cross-validation.

The second plot is a ROC Curve with the mean performance of the model for all the splits. The ROC Curve is created with Plotly and is an interactive plot. The AUC is shown in the bottom corner with the legend.

The next plot is the PCA It is created with Plotly and is an interactive plot. The PCA is created with the selected features. The PCA reduces all these features to only 3 dimensions. Each sample is colored using the label from the metadata.

The last plot is the Feature Weight. This is a horizontal bar chart with Plotly. This plot shows the feature importance for the selected model.

3.5 Running a target prediction analysis

In the Analysis View, the miRNA/Gene Target Prediction is in the Target Prediction tab. The target prediction can run in two different ways. Either by using the 40 predictions tools from MIO, or by improving the results using the previous correlation analysis from MIO.

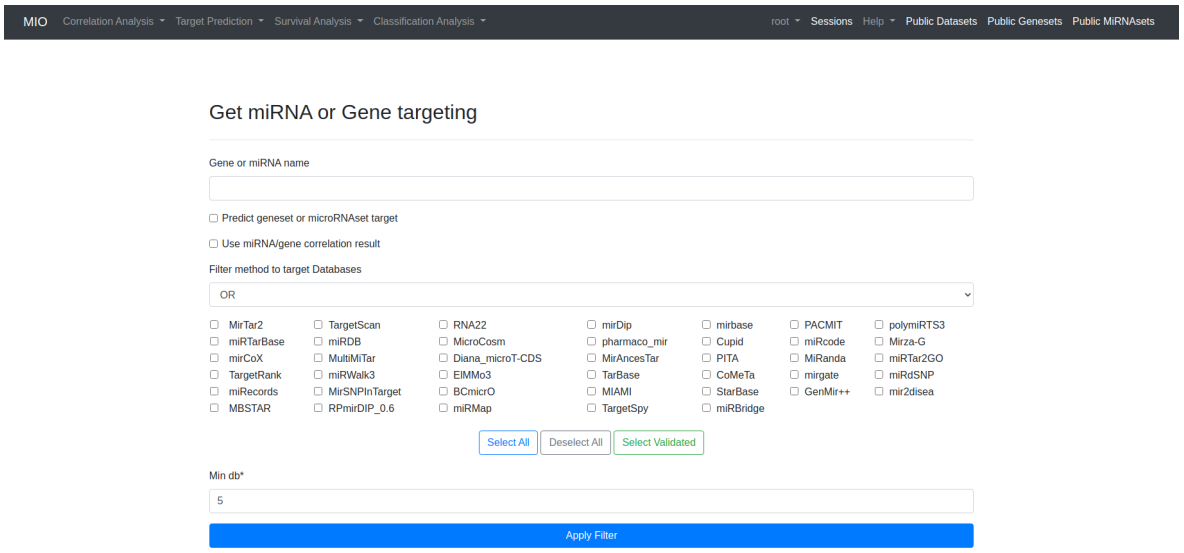


Figure 21. Target prediction form view.

In the miRNA/Gene Target prediction, the user needs to introduce the Gene Symbol or the miRNA mature ID, as shown in Figure 21. The user can use a specific gene set, or miRNAset by clicking the “Predict geneset or microRNAset target” button. The user can use a correlation analysis by clicking the “Use miRNA/gene correlation result” button. In the last step, the user can choose a combination of 40 different Target Prediction Tools to filter the miRNA/Gene pairs.

Finally, the results will be shown on the Result Page View. All the miRNA/gene pairs which passed will be displayed in a table. The user can download the filter table in Excel, csv, and tsv format. The fields in the table are explained in the previous section (3.1).

3.6 Running a miRNA Synthetic Lethal Prediction

In the Analysis View, the miRNA Synthetic Lethal Prediction is in the Target Prediction Tab. For miRNA Synthetic Lethal Prediction, every possible synthetic lethal gene pair was previously predicted with ISLE. In order to get the miRNA Synthetic Lethal for a given gene, the user needs to enter the Gene Symbol or the gene set into the menu depicted in Figure 22. Then, the tool gets the possible synthetic lethal genes for the chosen genes, and it will obtain all the targeting miRNAs for the gene list. In order to get the most specific miRNA, MIO performs a Fisher's exact test, by an "Overrepresented miRNA analysis", to obtain the miRNAs most enriched in these genes.

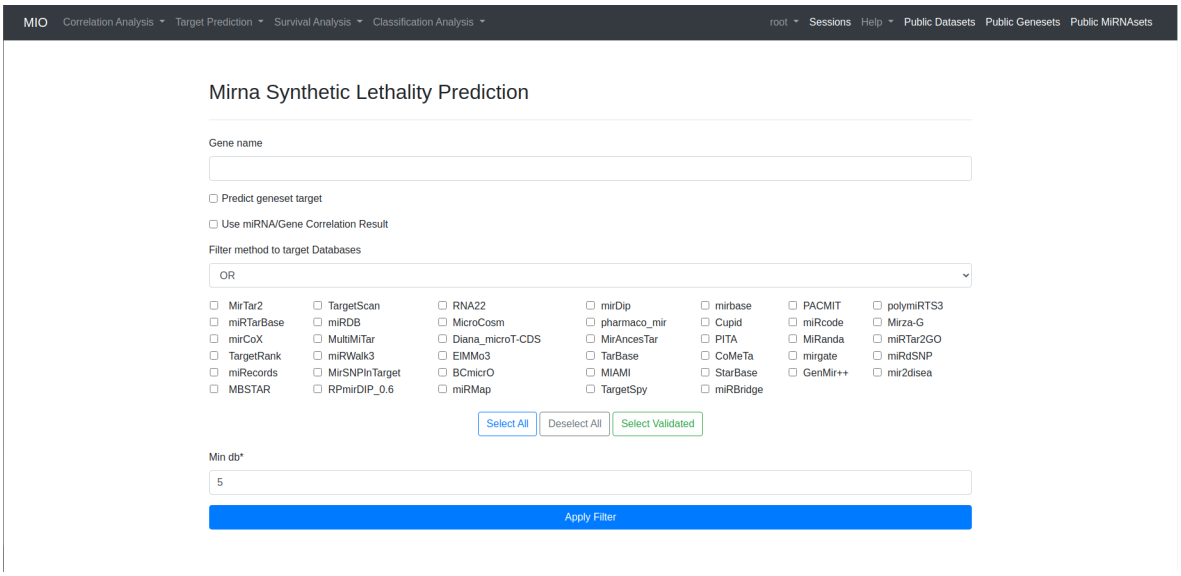


Figure 22. miRNA Synthetic Lethality Target Prediction form view.

Finally, the results will be shown in the Result Page View. All the miRNA/gene pairs which passed will be displayed in a table. The user can download the filter table in Excel, csv, and tsv format. The fields in the table are explained in the previous section (3.1). The results for the Overrepresented Analysis will be displayed in a second table, and all the interactions between the synthetic lethal genes and the miRNA target will be displayed in a network.

3.7 Running a Kaplan Meier Plot

In the Analysis View, the Kaplan Meier Plot is in the Survival Analysis tab. The Kaplan Meier Plot allows getting the survival plot to a given miRNA, gene or miRNA/Gene ratio, splitting the sample into higher and lower expressions. As can be seen in Figure 23, the user needs to indicate which gene or miRNA will be used for the analysis, and the user has to indicate the quantile which will be used as a cutoff to split the sample into higher and lower expressions.

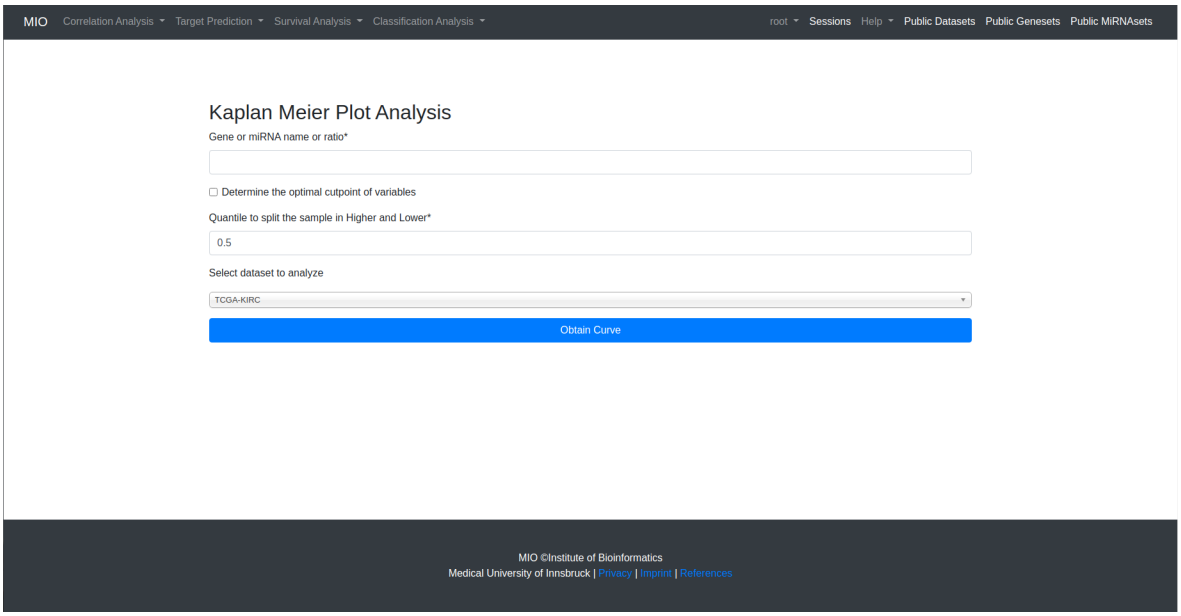


Figure 23. Kaplan Meier Plot form view.

The results will be shown in the Result Page View. The result is an interactive plot created with Plotly, an example of which is depicted in Figure 24. On the Y-axis, the Survival Probability is represented. On the X-axis, the survival time is depicted. The survival curve for high-expression samples is plotted in red, and low-expression samples in blue. Finally, inside the plot, the log2(HR) and the p-value are plotted.

Figure 24. Kaplan Meier Plot with the optimal cutpoint for CD276 in the TCGA-LUAD

dataset.

4. User Management

4.1 Logging in

In order to use MIO the user needs to be logged in. If the user does not have an account, creating a new user or logging in as a temporary user, is available to the user. All analyses, datasets, genesets, and results are connected to the user.

To create a new user, click “Sign up” and fill in the form.
To log in as a temporary user, click “Log in”, and then “Continue with the temporary account”.
To log in, click “Log in” and give your username and password.

What difference is there between a temporary user and a signed-up user?

The anonymous (temporary) user can access all the functionalities and tools of MIO. However, the data will be lost 7 days after the creation of the account, so the user will need to download the processed data and the results to their local computer. The anonymous user can become a normal (signed-up) user at any moment, by setting a username, and email.

4.2 Log in as temporary User

To log in as a temporary user, the user needs to click the “Log in“ button or try to run any analysis in the tool without logging in. Then, the user will be redirected to the Log in page. On the Log in page, the user needs to click the “Continue with temporary account” button.

When the user clicks the “Continue with temporary account” button, it will be redirected to the temporary user page. The user will be logged in as a temporary user, and the password and username will be shown on the page. On this page, the tool will show a warning message.

Transform a temporary user into a normal user

A temporary user can be transformed into a normal user clicking the “Make Permanent” option in the user main menu in the top bar. The user will be redirected to an Update Profile Menu. In this menu, the user needs to enter a username, first name, last name, and email address. When the user updates the user data, it will be transformed into a normal user.

4.3 Sign up

To sign up the user needs to click the “Sign up“ button in the main menu in the top bar. On the sign up page, the user needs to enter an email and a password.