

Application User Manual

goloco: A Web Application to Create Genome Scale Information from Surprisingly Small Experiments

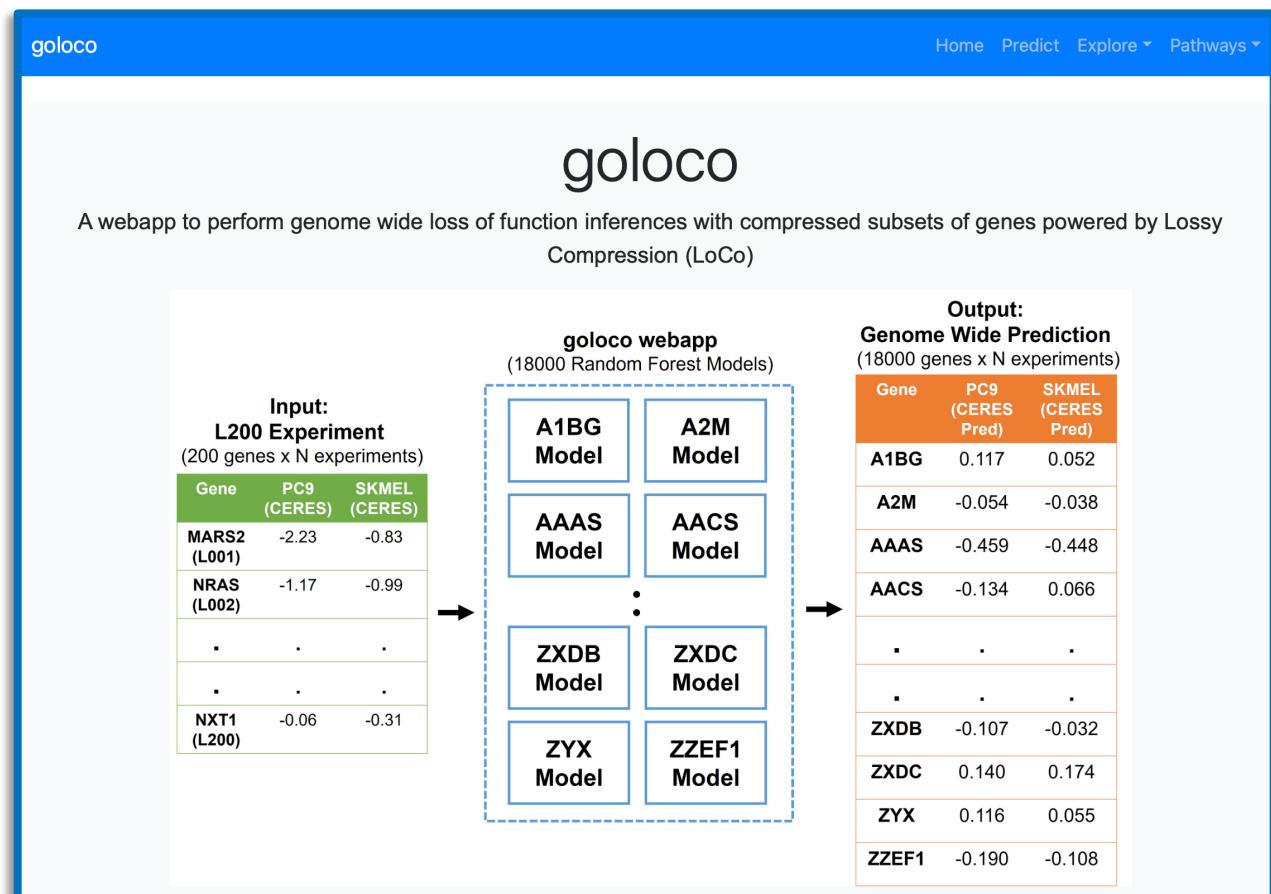


Table of Contents:

Section 1: Download Lossy Compression Sets.....	3
Section 2: Convert Experimental Data to Gene Effect Scores.....	4
2.1: Read counts Data frame.....	4
2.2: Sequence Map Data frame.....	5
2.3: Guide map Data frame.....	5
Section 3: Submit a Genome-wide Inference.....	7
3.1: Option 1: Submit a new Job.....	7
3.2: Option 2: Upload a previous prediction.....	9
3.3: Option 3: Load a sample prediction.....	10
Section 4: Explore the Visualizations.....	11
4.1: Overview.....	11
4.2: Volcano Plot.....	12
4.3: Z-Score Hits Plot.....	14
4.4: Multiple Pairwise Regression Plot.....	16
4.5: Single Pairwise Regression Plot.....	18
Section 5: Pathway Analysis.....	20
5.1: gProfiler2.....	20
5.2: Network Visualization.....	21
5.3: Community Cluster Analysis.....	23

Section 1. Download Lossy Subset Genes

- To navigate the goloco application, use the navigation banner at the top right.

- Navigate to the “Predict” page of goloco and scroll down to the bottom of the page.
- Download the “L200_landmark_genes.csv” file by selecting the link as below.

Download Lossy Subsets

L100, L200, and L300 are compressed subsets of genes consisting of 100, 200, and 300 genes, respectively, which tunably predict thousands of other CRISPR gene effects. Genome-wide loss-of-function predictions can be made by measuring the gene effect of only these compressed subsets. You will find downloadable links to the L100, L200, and L300 gene subsets below:

[L100_landmark_genes.csv](#)
[L200_landmark_genes.csv](#) ← Download this
[L300_landmark_genes.csv](#)

- The file contains a list of 200 genes in column A that represents the lossy 200 genes. The first 9 rows of this file are shown in the image below.

	A	B
1	feature	score
2	MARS2 (92935) [CERES]	
3	NRAS (4893) [CERES]	
4	SDHAF1 (644096) [CERES]	
5	IRF9 (10379) [CERES]	
6	ING5 (84289) [CERES]	
7	YPEL4 (219539) [CERES]	
8	NPIP85 (100132247) [CERES]	
9	SCYL1 (57410) [CERES]	

- A CRISPR sgRNA library for these 200 genes can be constructed as a standalone library or extracted from genome-wide libraries, including Brunello and Avana.
- CRISPR knockout experiments using this compressed library across multiple cell lines, and/or in different contexts, can be performed by the user prior to running genome-wide inferences on goloco.

Section 2. Convert Experimental Data to Gene Effect Scores

- Navigate to the “Gene Effect Calculator” tab in the “Predict” page of goloco.

goloco

Home Predict Explore Pathways

Submit Inference Gene Effect Calculator

Use this tool to convert your read-counts from CRISPR screens on L200 genes to CHRONOS scores. Converting raw reads to gene effect scores with the CHRONOS algorithm, developed by the Broad Institute, requires at least 3 dataframes including a matrix of raw readcounts, a sequence mapping, and an sgRNA guide mapping. Optionally, gene-level copy number calls can be submitted or selected from the DepMap data to correct gene effect scores. Submit your data below in the correct formats to convert to CHRONOS scores:

1) Readcounts Dataframe:
Upload a matrix as csv mapping count values to sgRNA sequences and sequence IDs with the following formats listed in the details below.
▶ Details Download Example Drag and Drop or Select Readcounts File

2) Sequence Map Dataframe:
Upload a dataframe, as csv, mapping sequence IDs to cell lines, pDNA batches, and timepoints with the following four columns and precise headers as listed in the details below.
▶ Details Download Example Drag and Drop or Select Sequencemap File

3) Guide Map Dataframe:
Upload a dataframe mapping sgRNA sequences to genes as a csv with at least the two columns with precise headers as listed in the details below.
▶ Details Download Example Drag and Drop or Select Guidemap File

Run CHRONOS Conversion

- Converting CRISPR knockout experimental data to gene effect scores requires three .csv files as inputs:
- 1. **Read Counts Dataframe:** Expand the “Details” arrow of the Read counts Dataframe and download the example. The user can read the details for creating the Readcounts .csv as below to generate their own file from their experiments.

1) Readcounts Dataframe:
Upload a matrix as csv mapping count values to sgRNA sequences and sequence IDs with the following formats listed in the details below.
▼ Details
Indexes: unnamed index with records of sequenced entities including pDNA and biological replicates for your experiments
Columns: individual sequences of sgRNAs, first column in csv must be the unnamed index column
Values: number of reads counted for each sgRNA
Download Example Drag and Drop or Select Readcounts File

- An example of the first few columns of the Readcounts csv is included below:

	A	B	C	D	E
1		AAAAGAATAA	AAAGAGATAC	AAATAGCTAC	AACATGGCTC
2	pDNA		516	1704	1342
3	rep1		122	1106	339
4	rep2		142	1631	441
					426

2. Sequence Map Dataframe: Expand the “Details” arrow of the Sequence Map Dataframe and download the example. The user can read the details for creating the Sequence Map .csv as below to generate their own file from their experiments.

2) Sequence Map Dataframe

Upload a dataframe, as csv, mapping sequence IDs to cell lines, pDNA batches, and timepoints with the following four columns and precise headers as listed in the details below.

▼ Details

sequence_ID (str): sequenced entities, must match row indexes from readcounts matrix

cell_line_name (str): "pDNA" for pDNA or cell-line name, each pDNA batch must have at least one pDNA measurement, biological replicates of the same cell-line should share the same cell_line_name

pDNA_batch (int or str): pDNA measurements in same batch are grouped and averaged, then used as reference for replicates assigned to batch

days (int): days post infection, ignored for pDNA

Download Example

Drag and Drop or Select Sequencemap File

- An example of the Sequence Map .csv is included below:

	A	B	C	D
1	sequence_ID	cell_line_name	pDNA_batch	days
2	pDNA	pDNA		1 0
3	rep1	ACH-000030		1 21
4	rep2	ACH-000030		1 21

3. Guide Map Dataframe: Expand the “Details” arrow of the Guide Map Dataframe and download the example. The user can read the details for creating the Guide Map .csv as below to generate their own file from their experiments.

3) Guide Map Dataframe

Upload a dataframe mapping sgRNA sequences to genes as a csv with at least the two columns with precise headers as listed in the details below.

▼ Details

sgrna (str): sgRNA sequences, must match columns from readcounts matrix and can only appear once in this column

gene (str): the gene the sgRNA maps to in "geneSym (genelD)" format, i.e. "ACSL3 (2181)", should include the L200 genes

Download Example

Drag and Drop or Select Guidemap File

- An example of the first few rows of the Guide Map .csv is included below. This table maps sgRNAs to their corresponding genes:

	A	B
1	sgrna	gene
2	AAAAGATAAGAAGAAACG	C8orf33 (65265)
3	AAAGAGATAGCAGTGACCA	CDC25A (993)
4	AAATAGCTACGGTGAACCCG	DNM1L (10059)
5	AACATGGCTGAAGGAAGCCG	SNAPC3 (6619)

- For more details to create these 3 csv files from your data, please visit the chronos GitHub from the Broad Institute (<https://github.com/broadinstitute/chronos>).

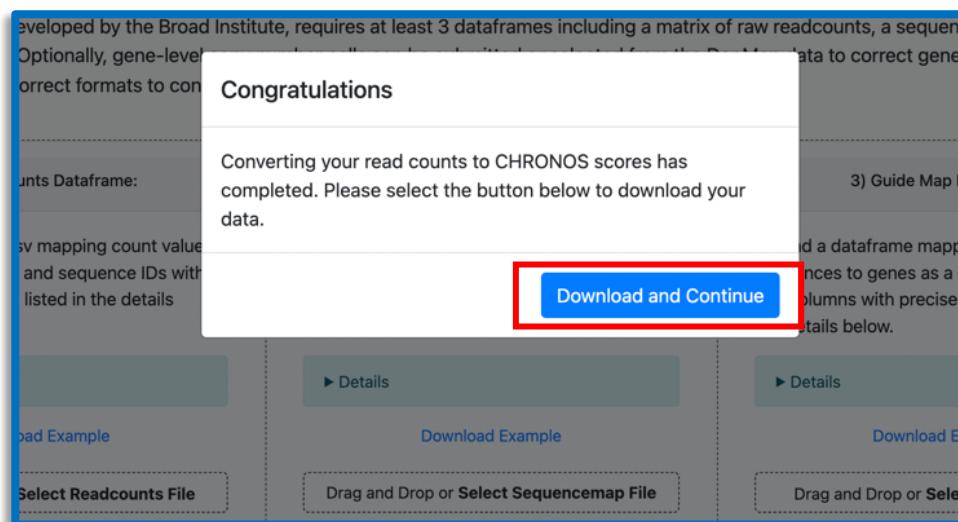
- The 3 csv files can be uploaded into their respective upload links in the “Gene Effect Calculator” tool and the “Run CHRONOS Conversion” button can be selected.

Submit Inference Gene Effect Calculator

Use this tool to convert your read-counts from CRISPR screens on L200 genes to CHRONOS scores. Converting raw reads to gene effect scores with the CHRONOS algorithm, developed by the Broad Institute, requires at least 3 dataframes including a matrix of raw readcounts, a sequence mapping, and an sgRNA guide mapping. Optionally, gene-level copy number calls can be submitted or selected from the DepMap data to correct gene effect scores. Submit your data below in the correct formats to convert to CHRONOS scores:

1) Readcounts Dataframe: Upload a matrix as csv mapping count values to sgRNA sequences and sequence IDs with the following formats listed in the details below. ▶ Details	2) Sequence Map Dataframe: Upload a dataframe, as csv, mapping sequence IDs to cell lines, pDNA batches, and timepoints with the following four columns and Download Example	3) Guide Map Dataframe: Upload a dataframe mapping sgRNA sequences to genes as a csv with at least the two columns with precise headers as listed in the details below. Download Example
1. Upload csv files to these 3 drag and drop links		
Drag and Drop or Select Readcounts File	Drag and Drop or Select Sequencemap File	Drag and Drop or Select Guidemap File
Everything looks good! File Uploaded Succesfully: sgRNA_raw_reads.csv	Everything looks good! File Uploaded Succesfully: sgRNA_mappings.csv	Everything looks good! File Uploaded Succesfully: sgRNA_gene_reference.csv
2. Select this button to start conversion		
Everything looks good! Ready to convert scores!		
Run CHRONOS Conversion		

- A pop-up will appear once the conversion is completed. Select the “Download and Continue” button.



Section 3: Submit a Genome-wide Inference

- Navigate to the “Submit Inference” tab in the “Predict” page.
 - Use **Option 1** to submit a new inference job
1. **Option 1:** Expand the “Details” arrow of Option 1 and download the example as reference. The user can read the details for creating the csv input for Option 1 as below. **Note:** This is also the output from the previous CHRONOS conversion step in Section 2

Option 1: Submit a new job

Upload a csv or excel file of L200 CERES scores for your experiment to the box below. Select the run inference button to run your inference:

▼ Details

feature (str): first column must list all the L200 genes in the following format example: "NRAS (4893) [CERES]"

name exp 1 (int): all other columns can be named by cell-type or experiment name, row values under column should be chronos scores corresponding to feature column

name exp 2 (int): all other columns can be named by cell-type or experiment name, row values under column should be chronos scores corresponding to feature column

Download Example

- An example of the first 5 rows of the input required for option 1 is shown below:

	A	B	C	D
1	feature	ACH-000030	ACH-000615	ACH-001042
2	ACSL3 (2181)	-0.293079	-0.3738258	-0.2563117
3	ACTR1A (101)	-0.9919657	-0.6193841	-0.4850361
4	ADNP (23394)	0.35332279	-0.0909388	0.04529878
5	AHCYL1 (107)	-0.9265004	-0.9211116	-0.5124891

- Upload the properly formatted csv input into the Option 1 upload link and select the “Run Inference” button. A pop-up will appear. Select “Continue”.

Option 1: Submit a new job

Upload a csv or excel file of L200 CERES scores for your experiment to the box below. Select the run inference button to run your inference:

▼ Details

Download Example

Drag and Drop or Select L200 File

File Uploaded Successfully:
PC9SKMELCOLO_L200_CERES_Features.csv

Run Inference

1. Upload csv file here

2. Select this button to start the inference

3. Select this button

Information

This inference may take over an hour. Please keep your browser and this page open as the inference runs. Once completed, you will be redirected to the overview page and a pop-up will prompt you to download your prediction. Select the button below to continue.

Continue

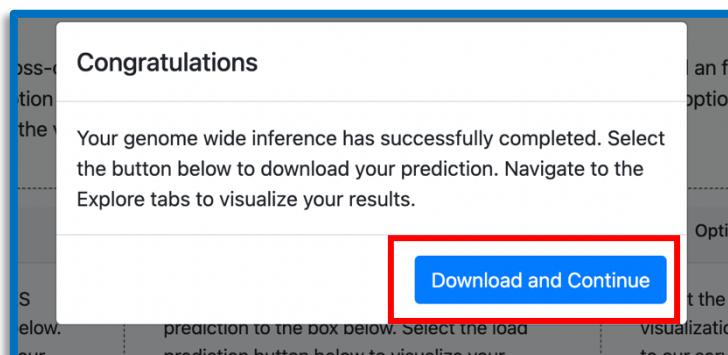
- The user will see the following screen while the inference is running. A progress bar will appear indicating the estimated time remaining.

The screenshot shows the Gene Effect Calculator interface. At the top, there are two tabs: "Submit Inference" (selected) and "Gene Effect Calculator". Below the tabs, a descriptive text explains the tool's purpose: "Use this tool to generate genome-wide loss-of-function predictions and data visualizations. Use option 1 to upload an file containing L200 CERES scores across your cell-line experiments, use option 2 to upload a csv of a prediction that was performed previously with option 1, or use option 3 to load a sample prediction if you are interested in testing the visualizations and predictions that this tool will generate." The interface is divided into three main sections:

- Option 1: Submit a new job**: Instructions: "Upload a csv or excel file of L200 CERES scores for your experiment to the box below. Select the run inference button to run your inference." Buttons: "Details", "Download Example", "Drag and Drop or Select L200 File". A green message box says "File Uploaded Succesfully: PC9SKMELCOLO_L200_CERES_Features.csv". A blue "Run Inference" button is at the bottom.
- Option 2: Upload a previous prediction**: Instructions: "Upload a csv or excel file of a previous prediction to the box below. Select the load prediction button below to visualize your previous prediction data." Buttons: "Details", "Download Example", "Drag and Drop or Select Prediction File", "Load Prediction".
- Option 3: Run a sample visualization**: Instructions: "Select the load sample button below to load visualizations of a sample prediction preloaded to our servers." Button: "Load Sample".

A progress bar at the bottom indicates "Estimated Time Remaining: 00:04:37". A "Cancel" button is also present.

- Once the prediction is complete, a pop-up will appear prompting the user to download the prediction. Select the “Download and Continue” button.



- Use **Option 2** to upload a previous prediction:

2. Option 2: Expand the “Details” arrow of Option 2 and download the example as reference. The user can read the details for the csv input for Option 2 as below. **Note:** This is also the output from the previous step in Option 1 from above.

Option 2: Upload a previous prediction

Upload a csv or excel file of a previous prediction to the box below. Select the load prediction button below to visualize your previous prediction data:

▼ Details

gene (str): column with ~18000 gene names in format "NRAS"

gene_category (str): either conditional, common essential, or common nonessential

avg (int): average CERES score for gene

std (int): standard deviation of CERES scores for gene

name_exp_1 (CERES Pred) (int): each experiment should have a column with ceres predictions

name_exp_1 (Z-Score) (int): each experiment should have a column with z-score predictions

Download Example

- An example of the input required for option 2 is shown below:

	A	B	C	D	E	F
1	gene	gene_catego		ACH-000030	ACH-000030 (Z-	
2	A1BG	conditional e:	0.102813528	0.121791592	0.116841597557	0.11518093178080
3	A1CF	conditional e:	0.0676545	0.121071289	0.076605552	0.073932093
4	A2M	conditional e:	-0.0607705	0.108204266	-0.054029421	0.062299341
5	A2ML1	conditional e:	0.174291337	0.113386643	0.189192190607	0.13141629747645

- Upload the properly formatted csv input into the Option 2 upload link and select the “Load Prediction” button. A pop-up will appear. Select “Continue”.

Option 2: Upload a previous prediction

Upload a csv or excel file of a previous prediction to the box below. Select the load prediction button below to visualize your previous prediction data:

► Details

Download Example

Drag and Drop or Select Prediction File

File Uploaded Successfully:
PC9SKMELCOLO_L200_CERES_Features_prediction (2).csv

Load Prediction

1. Upload csv file here

2. Select this button to load prediction

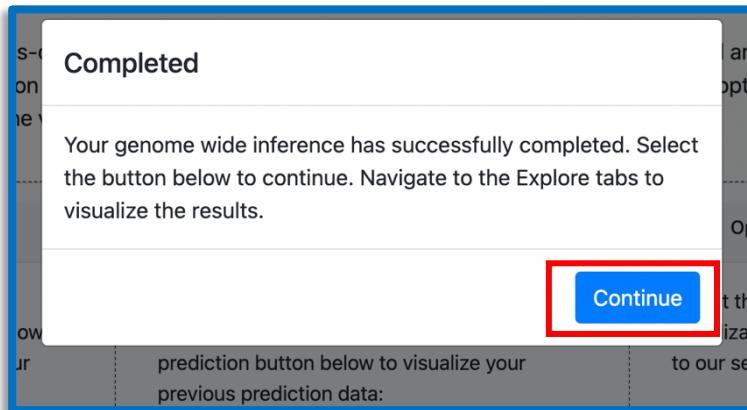
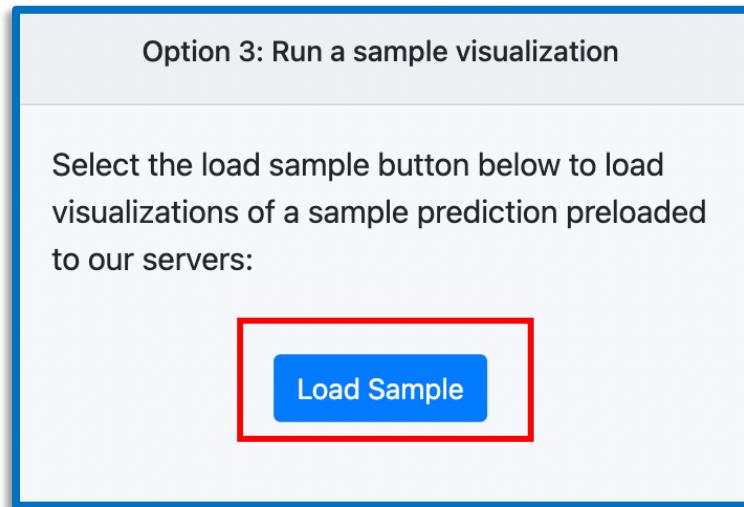
Completed

Your previous prediction has been uploaded. Select the button below to continue. Navigate to the Explore tabs to visualize your results.

Continue

3. Select this button

- Use **Option 3** to run a sample visualization:
3. **Option 3:** Select the “Load Sample” button to load a pre-loaded prediction from goloco server. A pop-up will appear. Then select the “Continue” button.

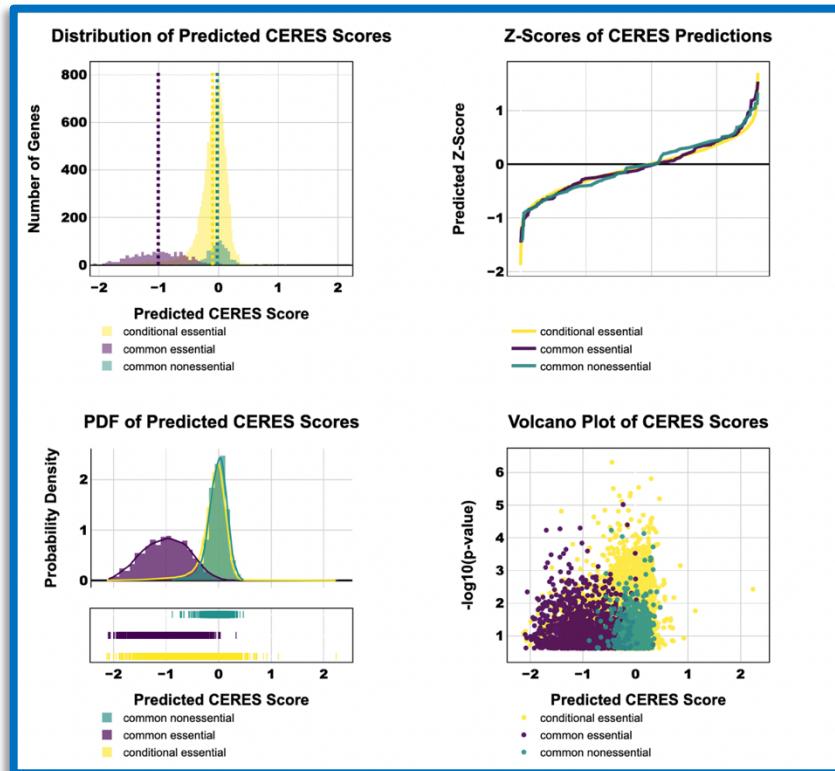


Section 4: Explore the Visualizations

- **Overview:** Navigate to the “Overview” tab by selecting the “Explore” dropdown menu in the navigation bar and selecting the “Hits” page. Generating the overview plots requires 3 inputs as below. Recommended selections for example are shown.

Overview	Volcano Plot	Z-Score Hits	Table
Select an experiment below to visualize the distribution of CERES scores and z-scores of predictions			
<input type="text" value="PC9"/> <input type="button" value="x"/>	1. The experiments from your prediction will auto populate in this drop-down menu. Select one experiment.		
<input type="text" value="viridis"/> <input type="button" value="x"/>	2. Selecting a color palette from the drop-down menu will apply discrete colors from the palette to the gene categories in step 3		
<input checked="" type="checkbox"/> Conditional Essential <input checked="" type="checkbox"/> Nonessential <input checked="" type="checkbox"/> Essential	3. Select one or more of the gene categories using the check boxes		

- Once the inputs are selected, the visualization will update automatically. The overview plot provides histograms and probability distribution functions of all predicted scores separated by gene categories (top left and bottom left). This may be useful as quality control.



- It is expected that common nonessential genes will have an average prediction around 0 while common essential genes will have an average prediction around -1.

2. Volcano Plot: Navigate to the volcano plot by selecting the “Explore” dropdown menu in the navigation bar and selecting the “Hits” page. Then select the “Volcano Plot” tab. The volcano plot requires 8 inputs described below. The recommended selections are shown.

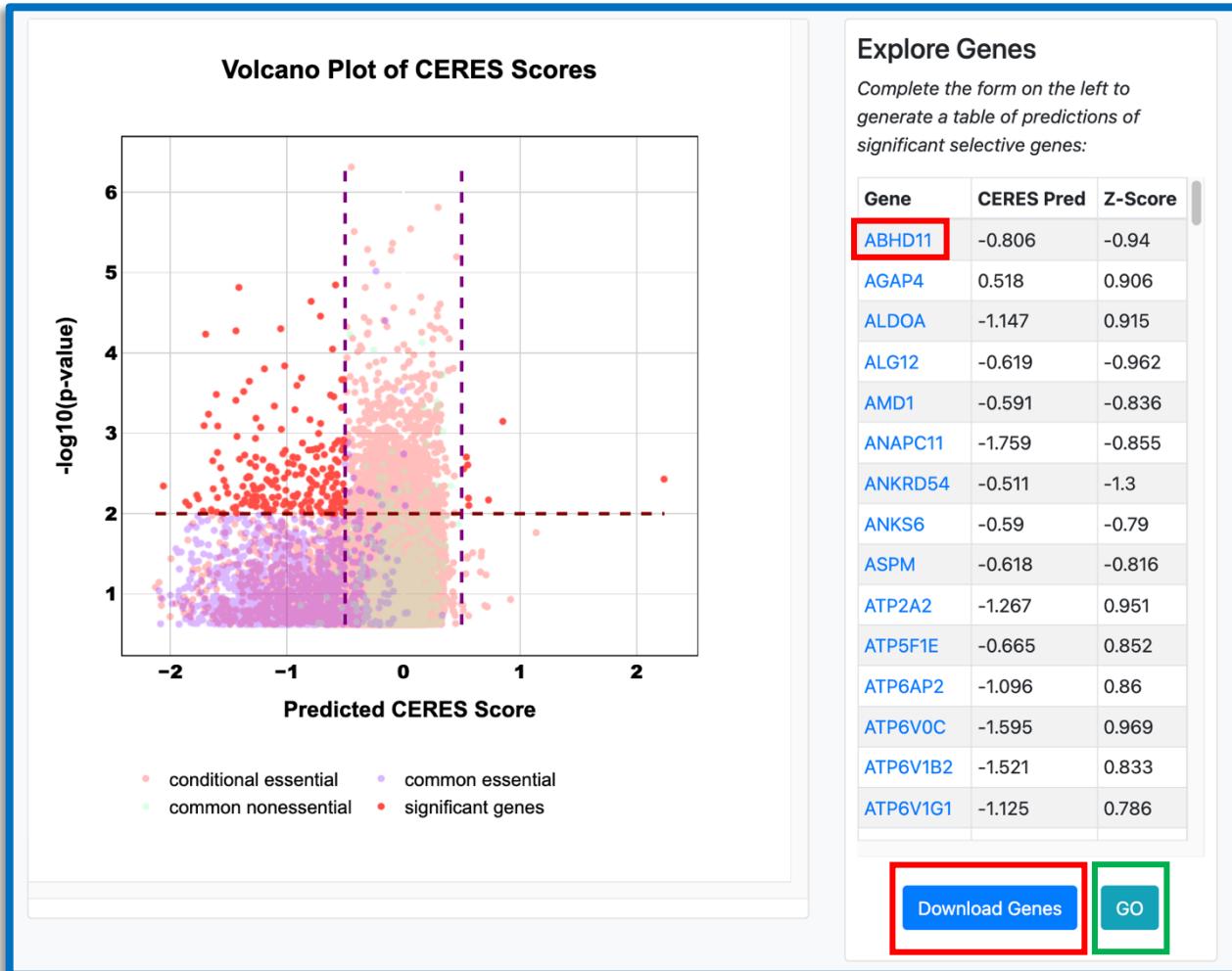
The screenshot shows the Volcano Plot interface with the following steps highlighted:

- Data:** PC9 (highlighted with a red border)
- Colors:** rainbow (highlighted with a green border), red (highlighted with a blue border), purple, maroon
- Thresholds:** select CERES thresholds (highlighted with a dark blue border), select p-val threshold (highlighted with a red border)

On the right, numbered steps explain the process:

1. The experiments from your prediction will auto populate in this drop-down menu. Select one experiment.
2. Select one or more of the gene categories using the drop-down menu. Multiple selections can be made.
3. Selecting a color palette from the drop-down menu will apply discrete colors from the palette to the gene categories in step 2.
4. Select a color to highlight significant genes.
5. Select a color to create vertical dashed lines on the plot corresponding to the CERES thresholds selected in step 7.
6. Select a color to create a horizontal dashed line on the plot corresponding to the p-value threshold selected in step 8.
7. Significant genes will be defined as those less than the lower border of this CERES threshold and greater than the upper border of this CERES threshold. (i.e. significant genes will be less than -0.5 or greater than 0.5.)
8. Significant genes will be defined as those with a $-\log_{10}(p\text{-value})$ greater than the selected threshold.

- Once the inputs are selected, the plot will automatically update (see next page) with CERES predictions of genes against their p-values. Based on the selected threshold values, significant genes will be highlighted in the top left and top right sextants based on p-value and CERES prediction thresholds.
- The “Explore Genes” frame on the right (see next page) will auto populate with the significant genes based on the filter criteria selected. These genes can be downloaded by selecting the “Download Genes” button.
- The “Gene” column in the “Explore Genes” frame will contain gene specific links to the DepMap website. Select a gene link to view.



- gProfiler2 can be run on this set of filtered genes by selecting the “Go” button in the “Explore Genes” frame. For the next steps on running gProfiler, visit section 5.1.

- 3. Z-Score Hits:** Navigate to the “Z-Score Hits” plot by selecting the “Explore” dropdown menu in the navigation bar and selecting the “Hits” page. Then select the “Z-Score Hits” tab. The Z-Score plot requires 9 inputs described below. Recommended selections for example are shown.

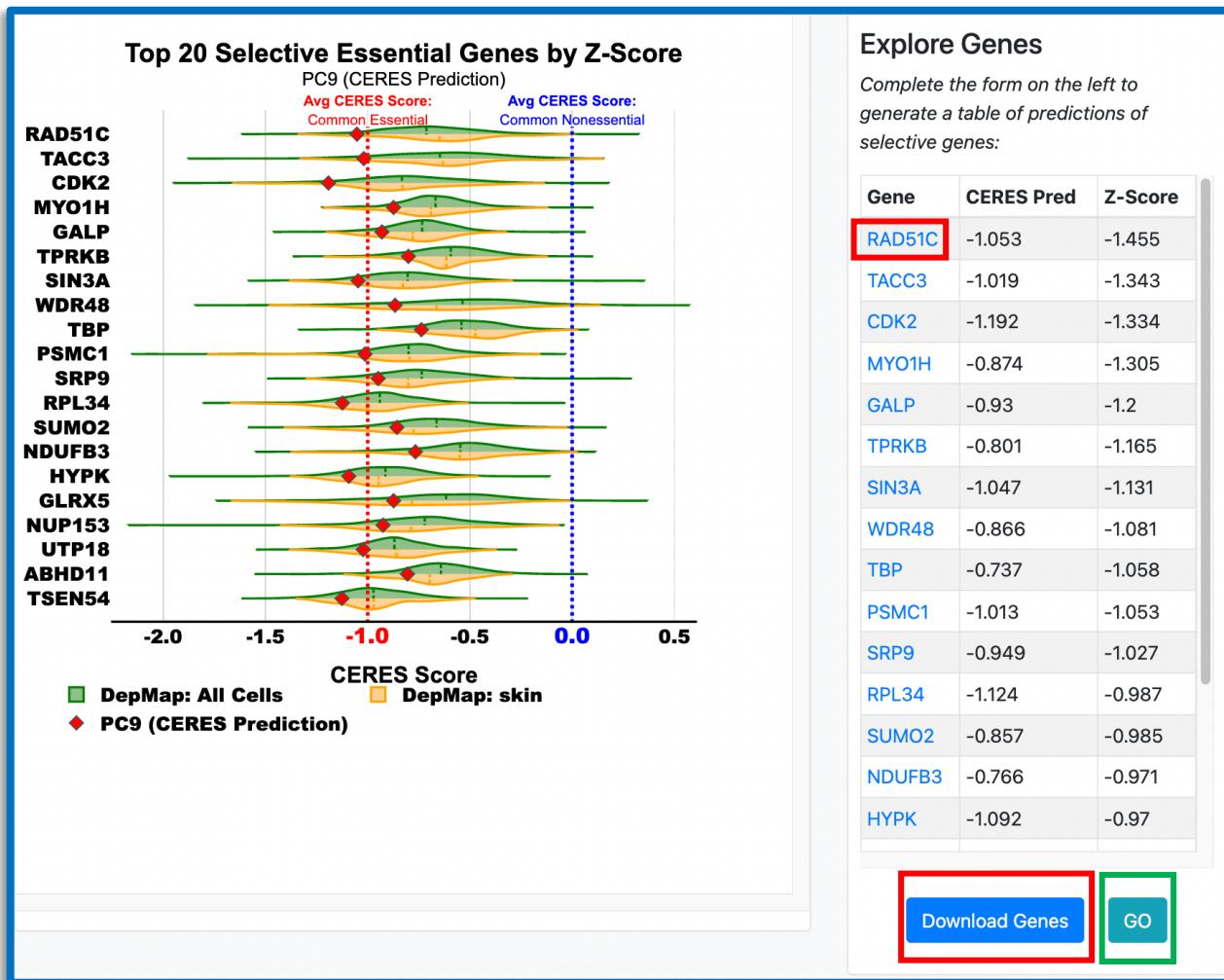
The screenshot shows the 'Z-Score Hits' tab selected in the top navigation bar. On the left, there are several input fields with recommended selections: 'Select Experiment' (PC9), 'Select Experiment Color' (red), 'Select Primary Histogram Color' (green), 'Select Max Genes' (100), 'Select CERES Threshold' (a slider from -1.5 to 0.5 with a blue dot at 0.0), 'Select Average Range Across Cell-lines' (a slider from -1.5 to 0.5 with a blue dot at 0.0), 'Select Category' (Lineage Subtype), 'Select Subcategory' (exocrine), and 'Select Secondary Histogram Color' (peachpuff). To the right of these fields are nine numbered steps with corresponding descriptions:

1. The experiments from your prediction will auto populate in this drop-down menu. Select one experiment.
2. The CERES prediction for the gene within each row of the figure, for the experiment selected in step 1, will be highlighted as a diamond with the color selected here.
3. For each gene row in the figure, the distribution of all CERES scores for that gene across all experiments will be in this color selected here along the positive y-axis.
4. Select the maximum number of genes to display in the figure.
5. Select genes with CERES predictions less than the threshold selected here. Genes with CERES predictions greater than the threshold will be filtered out.
6. Only genes with an average CERES score across all cell-lines within this range will be included.
7. Select from the drop-down menu to narrow down the subcategories in step 8.
8. Select from the drop-down menu to select a cell type or lineage or disease site depending on the selection in step 7.
9. For each gene row in the figure, the distribution of CERES scores across experiments of the subtype selected in step 8 will be in this color in the negative y-axis.

- Once the inputs are selected, the plot will automatically update (see next page) with rows corresponding to filtered genes, ranked by their Z-Score (decreasing extremity in Z-Score going down the plot).
- In each row, the diamond will represent the CERES prediction of the gene for the experiment selected. The positive y-axis for each row will be a histogram of the CERES scores for all experiments for that gene (green) and the negative y-axis will be a histogram of the CERES scores for a subset of experiments based on the

subcategory selected in step 8 (yellow). The predicted CERES score (red diamond) can be compared against these histograms.

- The “Explore Genes” frame will auto populate with the genes selected based on the user inputs and ranked by extremity in predicted Z-scores. These genes can be downloaded by selecting the “Download Genes” button.
- The “Gene” column in the “Explore Genes” frame will contain gene specific links to the DepMap website. Select a gene link to view.



- gProfiler2 can be run on this set of filtered genes by selecting the “Go” button in the Explore Genes frame. For the next steps on running gProfiler, visit section 5.1.

4. Multiple Pairwise Regression Plot: Navigate to the plot by selecting the “Explore” dropdown menu in the navigation bar and selecting the “Regressions” page. Then select the “Multi Linear Regression” tab. This tool requires 7 inputs as described below and some recommended selections as example.

1. Select “ALL” to maintain all genes possible in step 2. Or selecting a Louvain community to narrow the selectable genes in step 2 to those in the community selected.

2. Select two or more genes for a multiple pairwise regression analysis.

3. On the bottom left, beneath the diagonal of the figure, each pairwise regression will include a scatter of all experiments in this color.

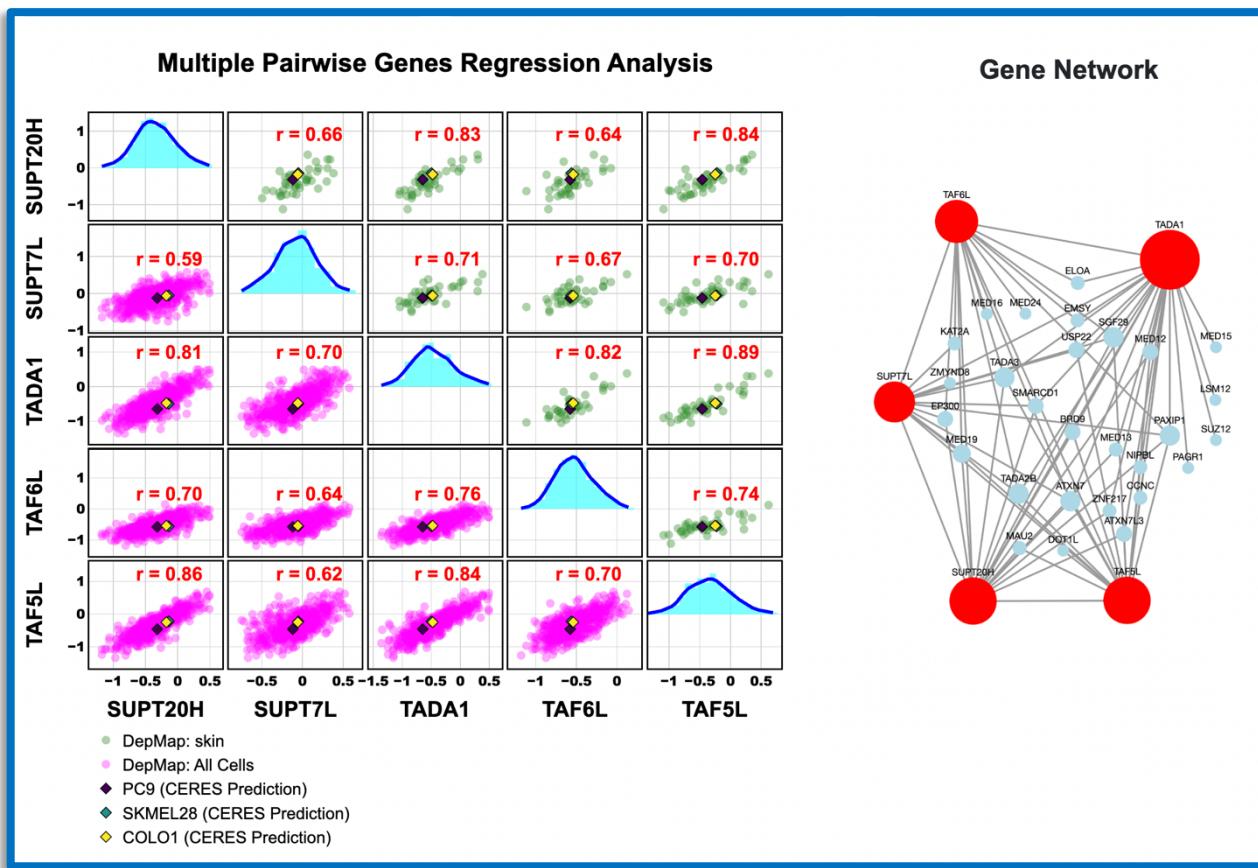
4. The CERES predictions from the user’s experiments will be scattered in each pairwise regression. Each of the experiments will be assigned a discrete color based on the color palette selected.

5. On the top right, above the diagonal of the figure, secondary scatterplots will be created for a subset of experiments. To narrow down a subset in step 6, select a category here.

6. For the secondary scatterplots, select a cell type, lineage, or disease based on the selection in step 5.

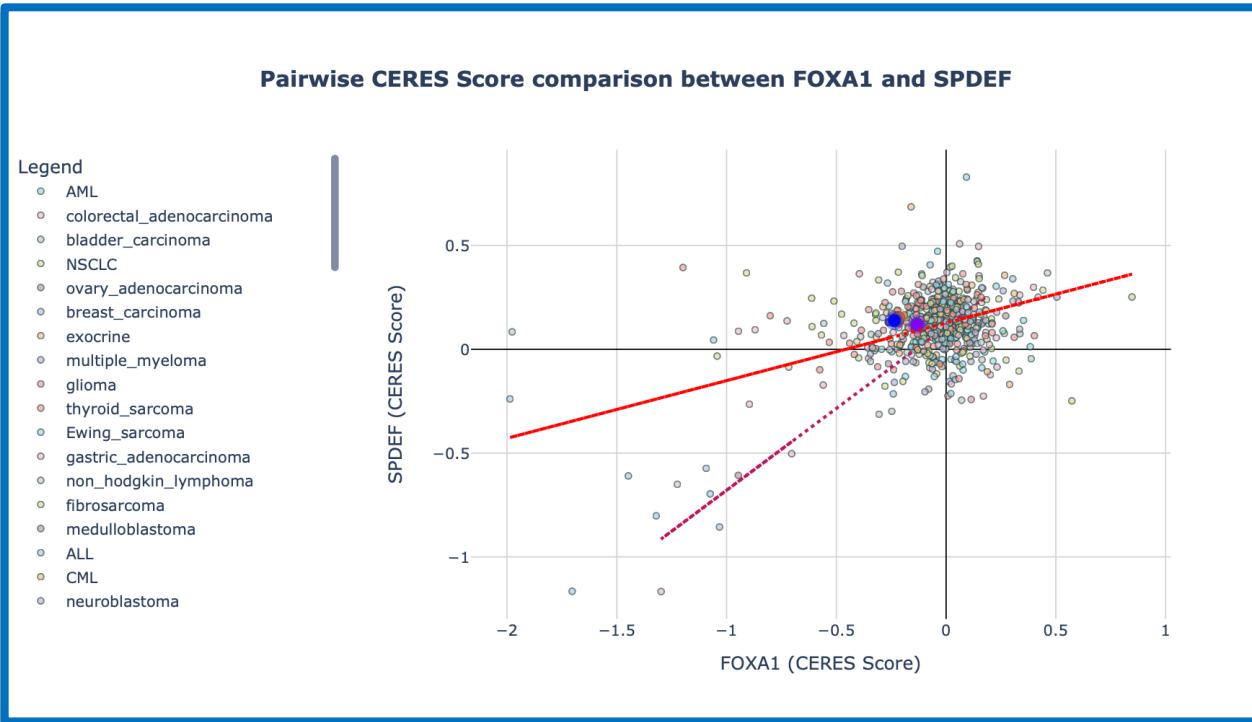
7. On the top right, above the diagonal of the figure, secondary pairwise regressions will be created for a subset of experiments based on the selection in step 6. The scatter plot will be in the color selected here

- Once the inputs are selected, the plots will automatically update. The plot on the left will show a matrix of pairwise regressions based on the selected genes.
- Beneath the diagonal histograms will be a scatter of the pairwise regressions for all experiments (pink), whereas above the diagonal will be a scatter of pairwise regressions of experiments of a subset of experiments selected in Step 6 above (green). Over this background scatter will be the CERES predictions from the users experiments in diamond shapes.
- On the right, and network will appear demonstrating the connections between the different genes, where a connection represents that the loss of function phenotype of one gene predicts the other.



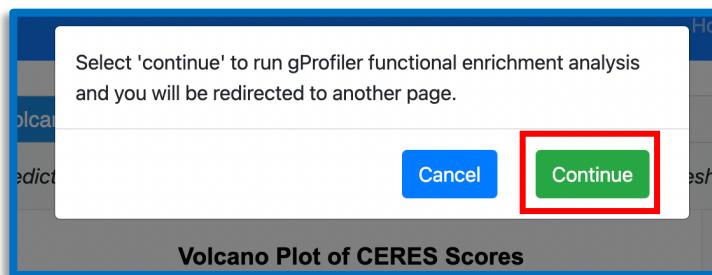
5. Single Pairwise Regression Plot: Navigate to the plot by selecting the “Explore” dropdown menu in the navigation bar and selecting the “Regressions” page. Then select the “Pairwise Linear Regression” tab. This tool requires 5 inputs as described below and some recommended selections for example.

- Once the inputs are selected, the plot will automatically update (see next page). A scatterplot of all experiments across the two genes selected will be generated.
- A linear regression of the two genes across all experiments will be displayed in a red dashed line while a regression of the two genes across the experiments of a subcategory selected in step 5 will be in a different color dashed line.
- In the example on the next page, the regression of gene effect scores of SPDEF and FOXA1 is performed across all experiments. The red linear regression line indicates positive correlation observed between the two genes across all experiments, whereas the darker purple linear regression line indicates a greater positive correlation for the two genes across just experiments of breast adenocarcinoma cell lines.
- Additionally, the users’ predictions across their experiments will be highlighted as circular larger points.



Section 5: Pathway Analysis

1. **gProfiler2:** Navigate to the enrichment analysis page by selecting the “Pathways” dropdown menu in the navigation bar and selecting the “gProfiler” page.
- As mentioned in section 4.2 and 4.3, gProfiler can be run on a set of filtered genes or significant genes in the user experiments by selecting the “GO” button in the “Explore Genes” frame from both the “Volcano Plot” and “Z-Score Hits” tabs.
- After selecting “GO”, the user will be prompted to be redirected to the gProfiler page. Select the “Continue” button.



- After being redirected, the gene set enrichment analysis will be performed, producing charts as seen below. On the top a Manhattan plot showing the functionally enriched pathways and a table beneath it will list the important biological GO terms enriched in the set of genes. On the right, a list of the genes queried will appear with gene specific links to the HGNC and Ensemble websites.



2. Community Visualizations: Navigate to the “Community Analysis” plot by selecting the “Pathways” dropdown menu in the navigation bar and selecting the “Communities” page. Then select the “Community Analysis”. This tool requires 10 inputs as shown below.

Community Analysis Cluster Analysis

Use this tool to visualize functional connections between different genes within a Louvain community for a DepMap cell

Network:

x ▾

1. Select one or more Louvain communities to visualize in the network. Multiple selections of communities can be named. The communities are numbered.

x ▾

2. Select a network layout for the visualization. “Spring” is the layout used in the next figure.

Data:

x ▾

3. For data, either select a cell-line from the DepMap dataset or a user experiment in step 4. This data will be applied to the network node colors and selected thresholds.

x ▾

4. Selecting a user experiment instead of a cell-line in step 3, will use the predictions for the experiment to apply to the network for node colors and selected thresholds.

Nodes:

x ▾

5. Select a value to color nodes in the network by the selected value. i.e. if “Louvain Community” is selected, nodes of the same community will be in the same color.

x ▾

6. Select a color palette. Discrete values from the palette will be applied to the selection in Step 5.

Select Color Resolution:

x ▾

7. If a quantitative variable is selected for node color, such as CERES scores, increasing the resolution will increase the difference in colors between similar values

8. Select a value to indicate how nodes should be sized. i.e. if “Betweenness Centrality” is selected, nodes with greater centrality will be bigger in the network

Gene Effect Threshold:

Select a CERES Threshold:

x ▾

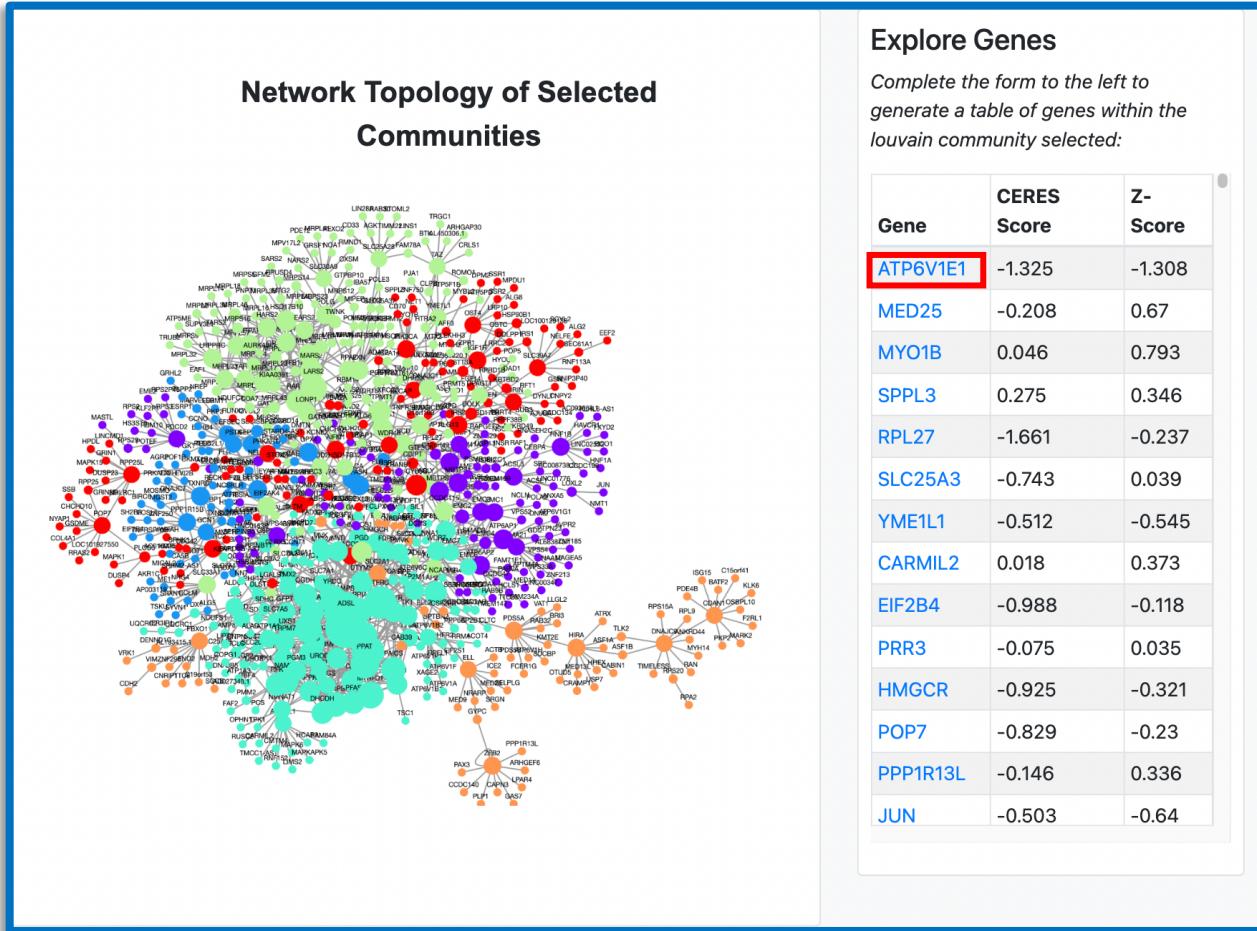
9. Only include nodes in the network if the CERES score for that node (gene) in the experiment selected in step 3 or 4 is less than the threshold selected here.

Select two-tail Z-Score Threshold:

x ▾

10. Only include nodes in the network if the $|Z\text{-Score}|$ for that node (gene) in the experiment selected in step 3 or 4 is greater than the threshold selected here.

- Once the inputs are selected, the visualization will automatically update (next page).



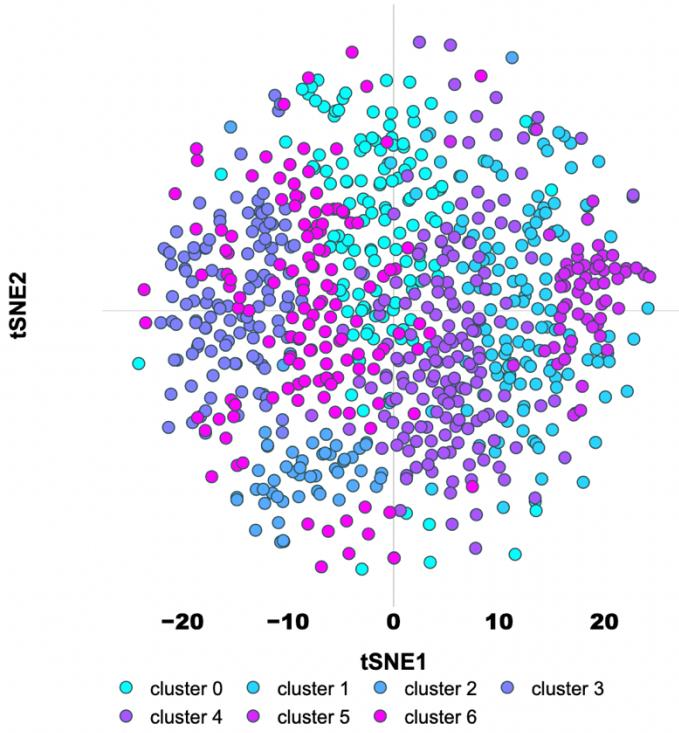
- In the above visualization, different colors represent different Louvain communities. Each node represents a gene and a connection between two nodes represents that one gene's loss of function phenotype predicts the other. Nodes are sized by their betweenness centrality so larger nodes indicate greater betweenness centrality, and thus increased predictive power for other nodes.
- Depending on the user selection in step 3 and 4 on previous page, size and color of the nodes can be modified to include data from user predictions or experiments.
- On the “Explore Genes” table to the right of the network visualization, gene specific links to the DepMap portal can be accessed. The table will include all genes in the network selected across all Louvain communities selected.

3. Cluster Analysis: Navigate to the cluster analysis visualization tab by selecting the “Pathways” dropdown menu in the navigation bar and selecting the “Communities” page. Then select the “Cluster Analysis” tab. The tool requires 6 inputs as shown.

- Once the inputs are selected, the plots will automatically update as shown in the next page. The cluster plot will show the clusters formed by different experiments from the DepMap data (each point indicates a different experiment), with genes only from the Louvain community selected. Each cluster may represent a different phenotypic utilization of the underlying Louvain network.
- The feature importance plot on the bottom of the next page will show which features (genetic nodes in the Louvain network) distinguish the indicated cluster from the other clusters.
- Depending on the user selection in step 4, 5 and 6 above, the table in the “Cluster Features” frame to the right of the cluster plot will update with genes in the Louvain network selected, and the data from either step 5 or 6 selected. Gene specific links to the DepMap portal can be accessed in the table.

PHATE Clusters for Louvain Community: 0

tSNE Projection



Cluster Features

Complete the form on the left to generate a table of top features for specified cluster and cell line:

Gene	CERES Pred	Z-Score
ABHD17C	0.1	0.05
ACACA	-0.242	0.527
ACLY	-0.429	0.296
ACSL1	-0.074	-0.203
ACSL4	-0.16	-0.21
ACSS1	0.081	0.094
ALG6	-0.444	-0.544
AMFR	-0.244	0.057
ANXA5	0.144	0.122
AP2M1	-0.75	0.293
AP2S1	-0.977	-0.165
ARHGAP5	-0.183	0.057
ATF7IP2	0.012	0.278
ATP6AP1	-1.262	0.52
ATP6AP2	1.000	0.00

Feature Importance Scores for Clusters of Louvain Community: 0

