User Guide

Released August 30, 2019

Developed by:

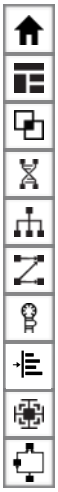
Table of Contents

Background and Introduction

Advaita develops bioinformatics tools that allow biomedical researchers to understand the underlying biological phenomena behind large and complex datasets. Our advanced computational methods emphasize a systems-biology approach that enhances researchers' understanding of disease mechanisms by placing gene expression, genetic variation, clinical data, and environmental effects into biological context.

Advaita's iPathwayGuide offers the most advanced pathway analysis available. Our patented Impact Analysis is the only method that exploits both pathway topology and enrichment in identifying significantly impacted pathways.

The eight main sections within iPathwayGuide are:



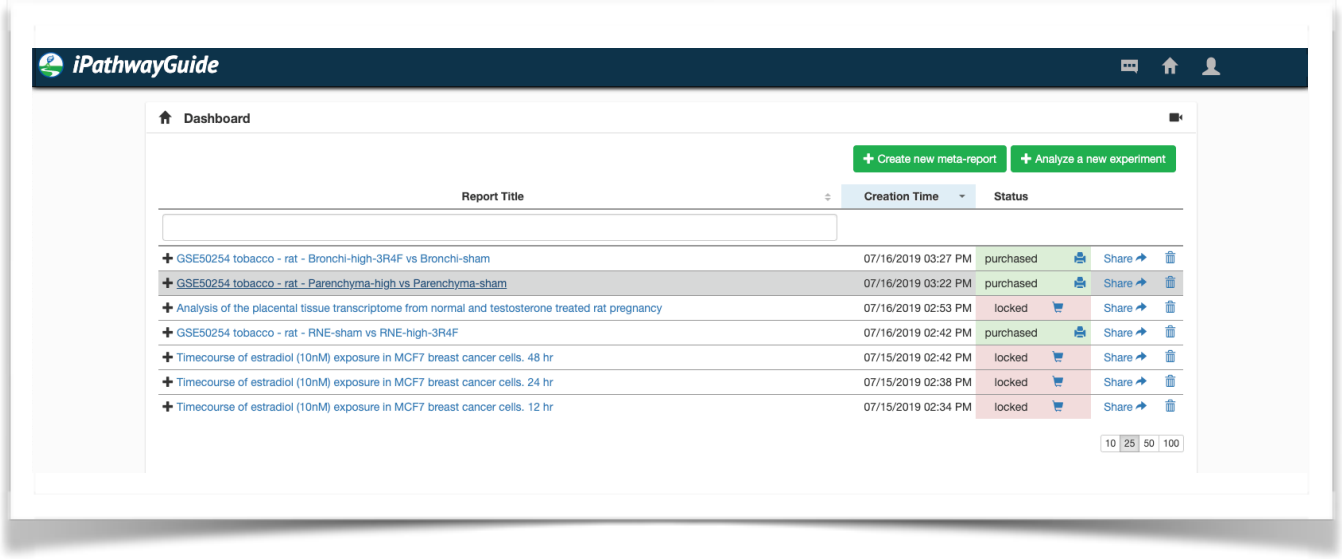
Dashboard
Summary
Meta-Analysis
Genes
GO terms
Pathways
miRNAs
Upstream Regulators
Diseases
Networks

Dashboard

Overview

The dashboard contains a list of all your reports and is the starting point for all new analyses. Each report is identified by the user-assigned name and creation time. There are 4 possible status indications for each report:

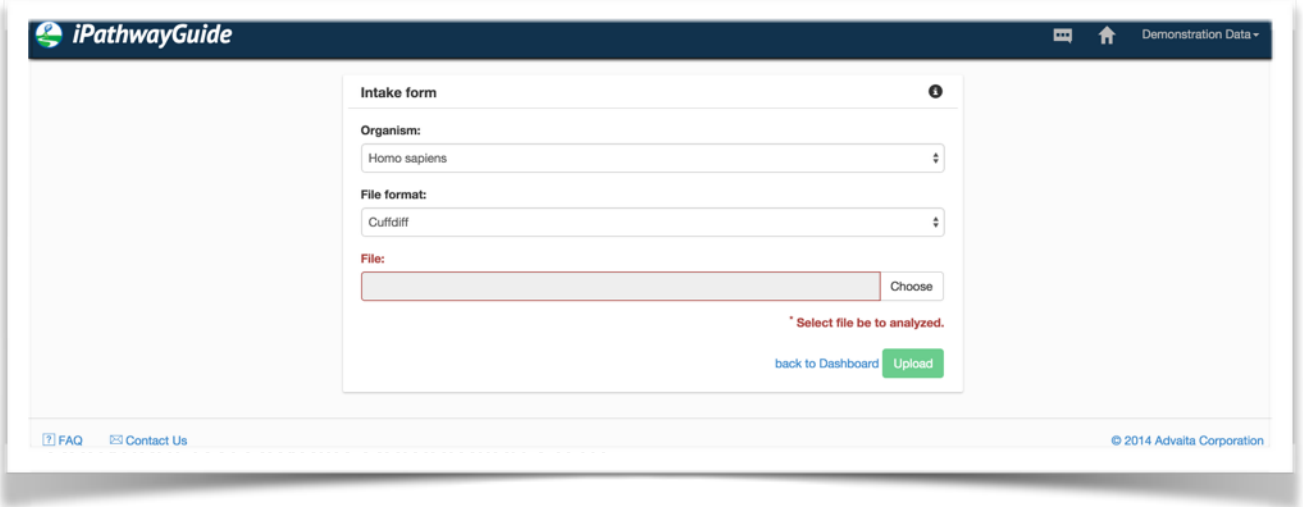
- **Pending** - an analysis that is currently being processed
- **Locked** - a completed analysis that has not been purchased
- **Purchased** - an analysis that has been purchased
- **Error** - an analysis that has failed processing



To generate a new analysis

[Note: These instructions pertain to generating a new analysis using the graphical user interface. To generate a new analysis using the API, see https://hub.docker.com/r/advaitabio/api-client/.]

Click on + Analyze a new experiment to open the data intake page:



Organism: Choose your organism.

iPathwayGuide supports analysis of **human, mouse, and rat** data because these are the organisms for which the best annotation is available.

For analysis of data from other organisms, we recommend that you convert genes from your organism of interest to orthologous human, mouse, or rat genes.

File Format: Choose your file format.

- **Text files:** In general, iPathwayGuide is set up to recognize tab delimited text files with columns for: Gene, Log2(Fold Change), and adjusted p-value (FDR or q-value). If your file has these columns, you can use the "Custom tab-delimited text" file format:

Gene Symbol	Log2 Fold Change	Adjusted p-value
Erdr1	-4.74343	0.01789
Galt	2.15787	0.23478
Myd88	0.44178	0.78945

For convenience, iPathwayGuide recognizes the following "standard" text file formats:

- **CuffDiff:** Output that ends with "_gene_exp.diff"
- **DESeq:** DESeq output - typically a text (.txt) file.
- **EdgeR:** edgeR topTags export: typically a tab-delimited text (.txt) file with columns for gene (label may be blank), logFC, logCPM, PValue, and FDR.
- **GEO2R/Limma:** GEO2R output text file
- **JMP Genomics:** iPathwayGuide output from JMP Genomics.
- **nSolver™ data:** From the export function in nSolver, choose which comparisons you wish to export in a single file with columns for: gene symbol, logRatio, and p-value. Export as a *.txt file.
- **Sciex SWATH Proteomics data:** "foldchange.csv" file that comes from the Sciex Protein Assembler or Sciex Protein Expression Workflow.

In addition, iPathwayGuide recognizes some Affymetrix raw datasets (.cel files)

- **Afffymetrix .cel Files:** iPathwayGuide accepts the following platforms (only)

Human

Human Genome U133
Human Genome U133A 2.0
Human Genome U133 Plus 2.0
Human Genome U95
Human Genome U35K

Mouse

Mouse Expression Set 430
Mouse Expression Set 430 2.0
Mouse Genome 430A 2.0

Rat

Rat Expression Set 230
Rat Genome 230 2.0
Rat Genome U34

Data Upload

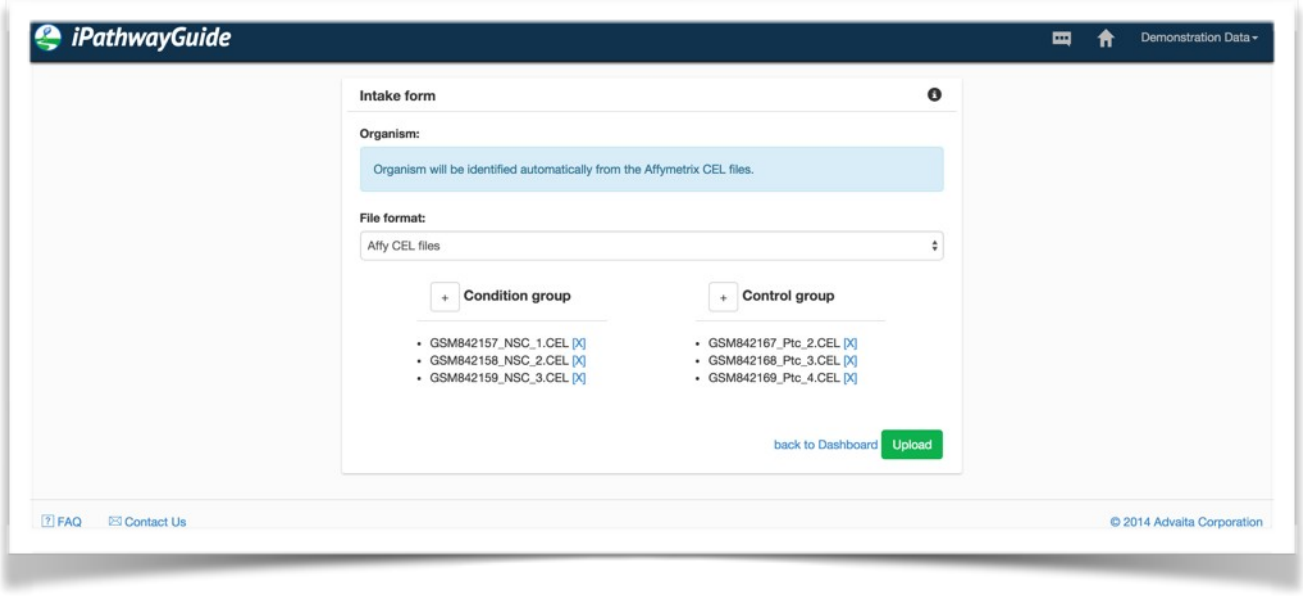
Input data File: Click on the Choose input data file to choose your input file and begin the uploading process.

Text File formats: The software will proceed with upload and quality checking of the data.

- **Note:** To assure the correct background, we strongly recommended that you upload differential expression data on all genes that were measured in your experiment, including genes that are differentially expressed and genes that are not differentially expressed. Cutoff thresholds are applied to select DE genes from the total set in the Contrasts page (below).

Affymetrix CEL file upload:

If you choose **Affymetrix CEL File** as your file type, a dialogue will prompt you to choose the CEL files that belong to the "condition" (e.g. treatment) and "control" (e.g. untreated, wild type) groups. You must select at least 3 files for each group.

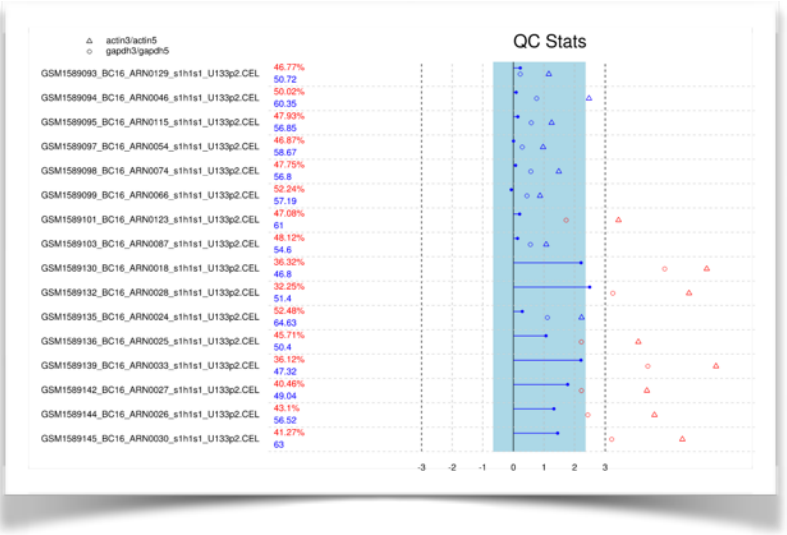


Once you have chosen the appropriate files, the data will be submitted for initial processing. Some of these files can be quite large and upload times can take several minutes depending on the quality of your data connection.

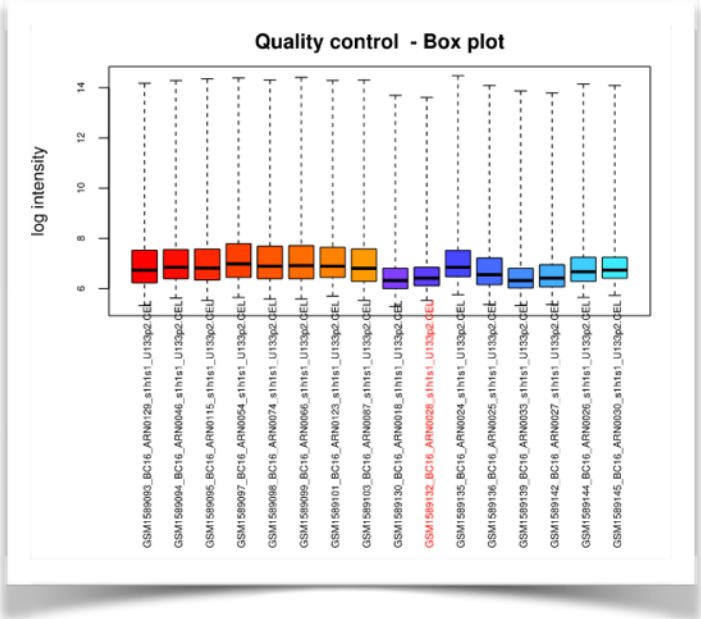
Affymetrix CEL file QC and Normalization

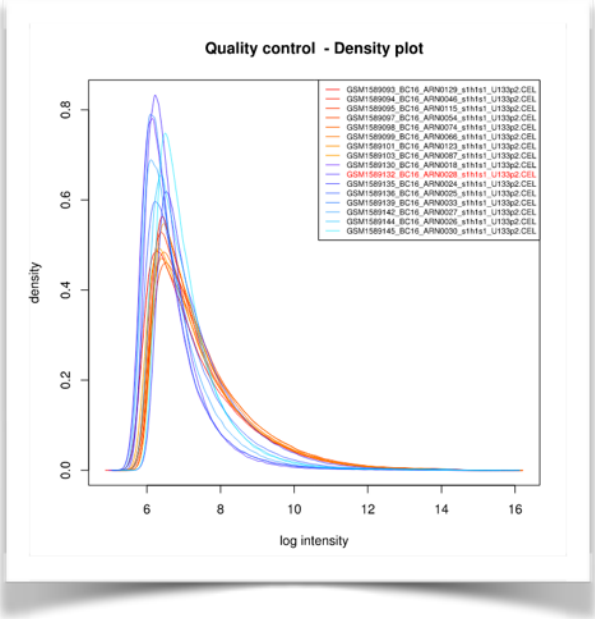
iPathwayGuide will perform Quality Control (QC) for Affymetrix CEL files and then normalize the data. A detailed printable report is available from the application once these steps are performed in real time.

Comprehensive information regarding which arrays (if any) were rejected, and the normalization scheme used along with pre- and post-analysis figures is provided.



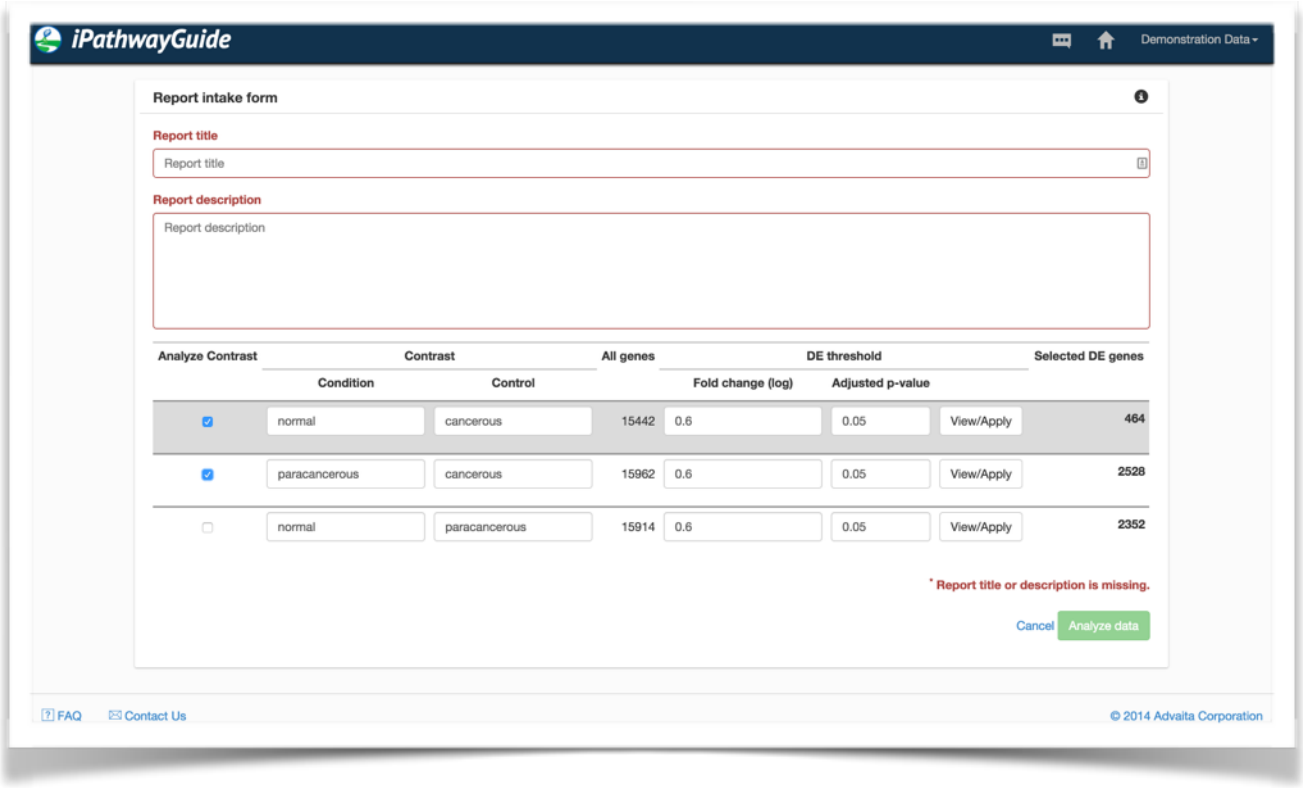
Once the data have finished the QC and Normalization process, they may be further processed to select the Differentially Expressed Genes (DEGs) by proceeding to the Contrasts Page.





Report Intake Form (Contrasts Page)

Once the data has been uploaded, the **Report Intake form** is displayed to gather additional information.



Report title: Provide a descriptive title. For example: "Tumor vs Normal Breast Cancer Experiment #2." This title will be displayed on your dashboard. In case you have a set of related contrasts, we recommend that you use a naming convention that makes it easy to see that they are related contrasts and to distinguish among them.

Report description: Provide description of your experiment. This section is important to provide context for the analysis. For reports that will be shared with colleagues, we recommend that you provide sufficient detail to allow them to understand all of the experimental conditions.

Contrasts: Provide labels for your contrast(s). For example, the condition might be "Tumor" and the control might be "Normal". For some file formats the contrast labels will be auto-detected from the input file. As with the report title, if you have a set of related contrasts, we recommend that you use a naming convention that makes it easy to see that they are related contrasts and to distinguish among them.

- Note that Report Title, Report Description, and Contrast labels can be edited after the report is generated.

DE Thresholds: Set the most appropriate thresholds to identify differentially expressed (DE) genes for your experiment.

Fold Change (log): Default value 0.6

- Fold change is used as a proxy for biological relevance (e.g. a change of at least this magnitude is assumed to have a biologically relevant effect). iPathwayGuide assumes that the input is in log2 format and the default threshold of log2(0.6) is equivalent to ~ 1.5 Fold Change in linear space.
- The software assumes that genes are interesting if they change in either direction (increase or decrease relative to control) so we use the absolute value of the selected threshold.
- Adjust the threshold as appropriate for your experimental conditions. You may opt to use a lower threshold (e.g. select a threshold of 0.4) to identify more genes as having biologically relevant changes of expression, or a higher threshold (e.g. 0.8) to identify less genes as having biologically relevant changes of expression. Having control genes (e.g. one or more genes that are known to be interesting) can be very helpful when adjusting the fold change threshold. If no controls are available, the default is a reasonable value for most genes.

Adjusted p-value: Default value 0.05

- The p-value (e.g. statistical significance) is used to control the false positive rate. iPathwayGuide assumes that the input values have been corrected for the number of hypothesis tests in the experiment (e.g. converted from p-value to False Discovery Rate or q-value).
- As with the Fold Change threshold, if appropriate adjust the p-value threshold to control stringency based on your experimental conditions.

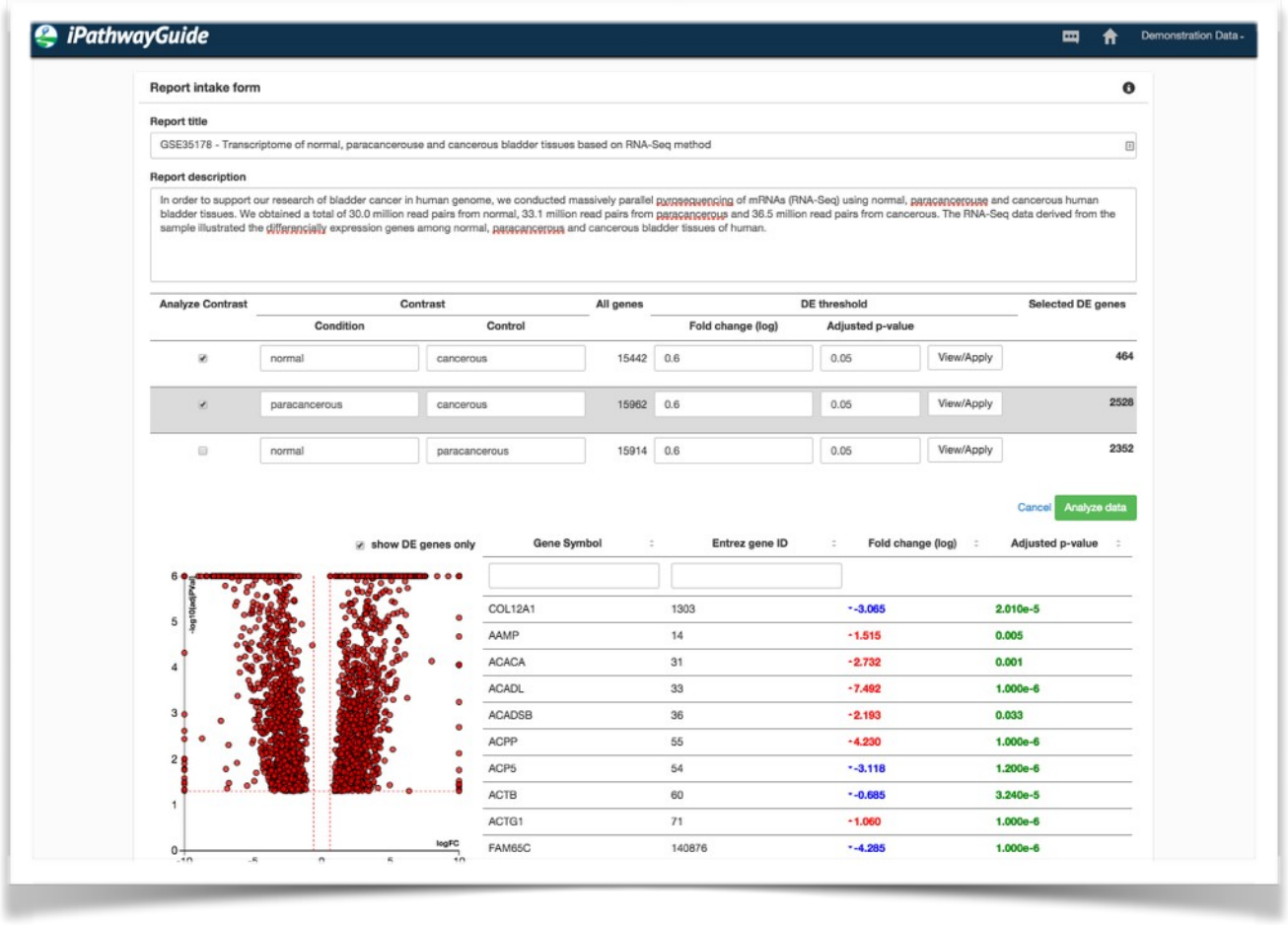
Clicking on "**View/Apply**" of a contrast will:

- Highlight the row in the table
- Display the number of genes that are DE (i.e. meet the combined thresholds for log(FC) AND Adjusted p-value)
- Display the volcano plot associated with that contrast
- Save your input to the contrasts labels and DE thresholds

❖ Note that DE thresholds CANNOT be edited after the analysis is complete and the report generated. Review the set of genes that meet your selected thresholds and adjust those thresholds if appropriate prior to running the analysis.

If your input file has multiple contrasts, you may select up to 5 contrasts in a single report. Not all file types support multiple contrasts in a single file. Please refer to the 'Multiple Contrasts' section below for more information.

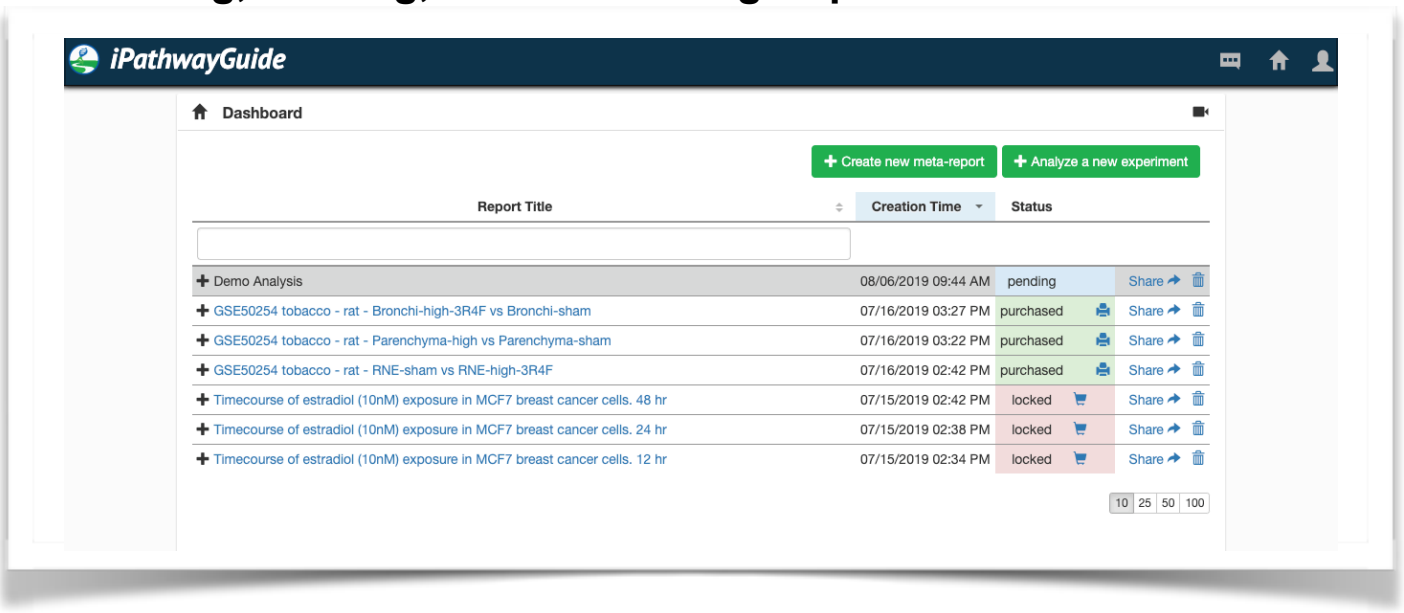
Once the report title and report description are entered and an input data file has been chosen, the Analyze data the button in the bottom right corner will turn green. Set the contrast names and adjust DE thresholds as appropriate, then click Analyze data.



Pressing the 

button will submit the file for analysis. Each contrast generally takes 10-15 minutes to analyze. As soon as data is submitted for analysis, an estimated time of completion will be displayed. ***You will receive an automated email*** with a link to the analysis ***once the analysis is complete*** and ready to view.

Viewing, Sharing, and Purchasing Reports



Viewing a Report:
To view an analyzed report from the Dashboard, click on the title of the report. Locked reports must be purchased before viewing, and pending reports must finish processing before they can be viewed.

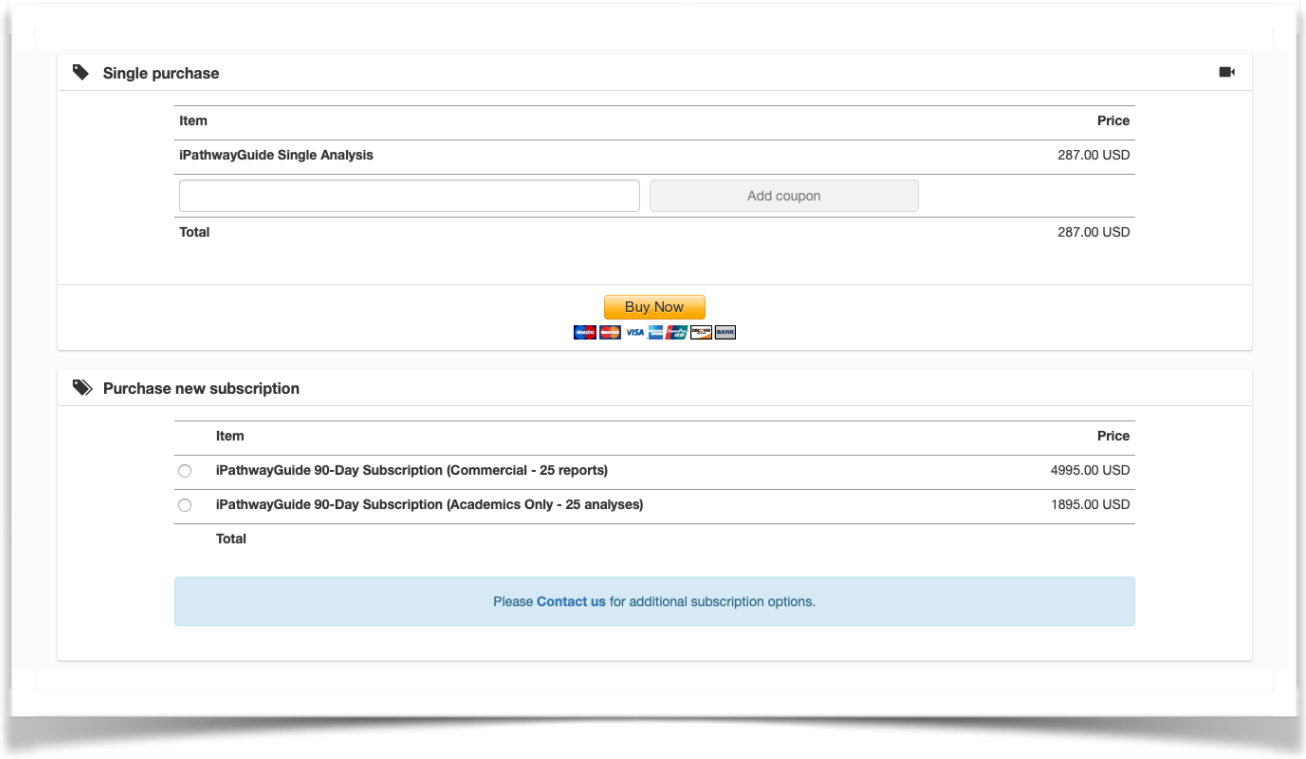
Sharing a Report:
To share a report, click on the "Share" link next to the report title in your Dashboard or in the report. A window will pop up where you may add multiple emails of recipients for the report. You may also set whether a shared report may continue to be shared to others by checking the "Allow to re-share" box.

If you see "~~Share~~", the report either has not finished or the person who shared it with you did not give you permission to share it onward.

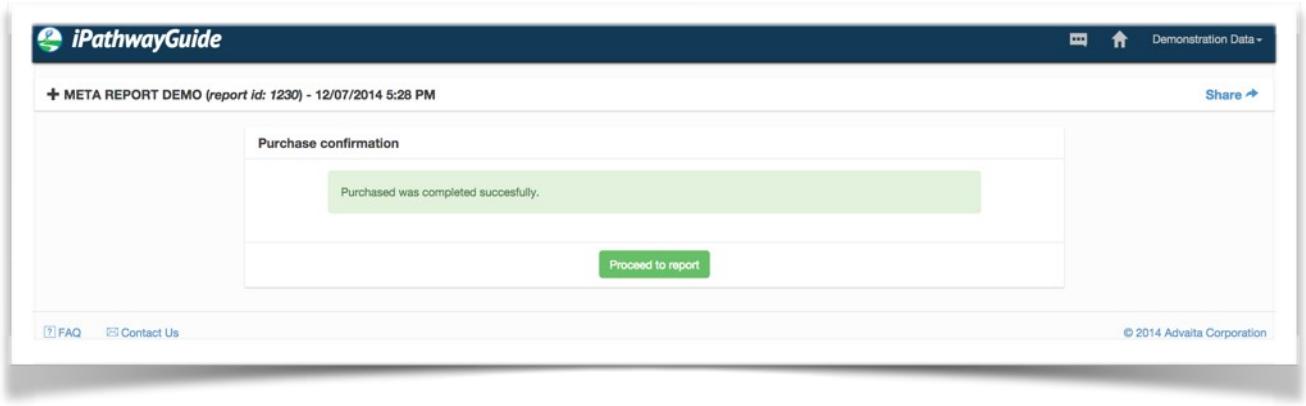
Purchasing Reports:

Credit Card
- You may purchase a limited number of reports with a credit card.

Subscriptions
- There are significant advantages to having a subscription to iPathwayGuide. We have subscriptions designed to meet the needs of all users. Contact us and we'll be happy to help you find the right fit for your organization.



When a report has been purchased successfully, you will get the following message:

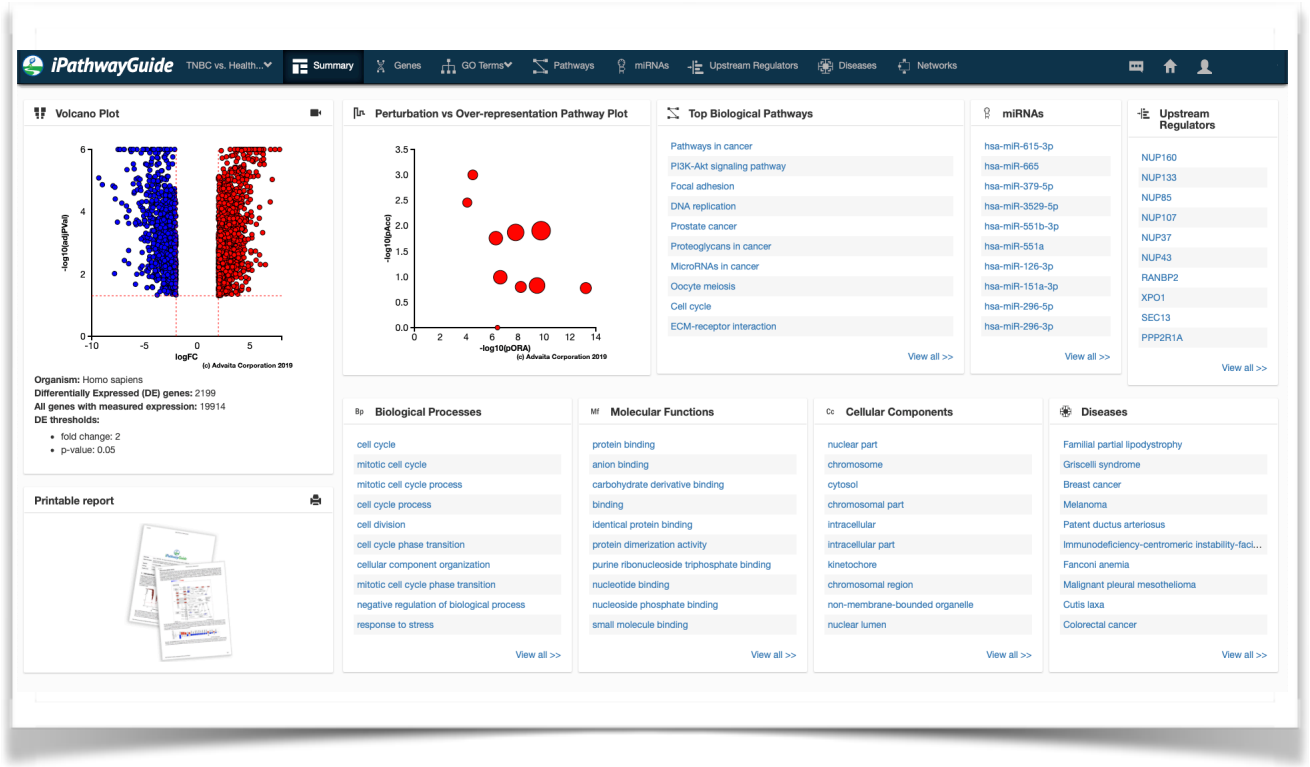


Summary

The Summary tab is designed to provide a brief overview of your data. The summary page consists of:

- The report title and description
- Volcano plot of DE genes
- Two-way evidence plot of pathways: Perturbation vs. Over-representation
- Top 10 significant pathways
- Top 10 Predicted micro RNAs (miRNAs)
- Top 10 upstream regulators
- Top 10 enriched biological processes (GO)
- Top 10 enriched molecular functions (GO)
- Top 10 enriched cellular components (GO)
- Top 10 enriched diseases
- Printer friendly summary

Clicking on the individual terms in these sections will take you to the respective detailed analysis. You can also use the navigation bar on the top of the screen to switch to and explore these sections.



Editing and Updating Projects

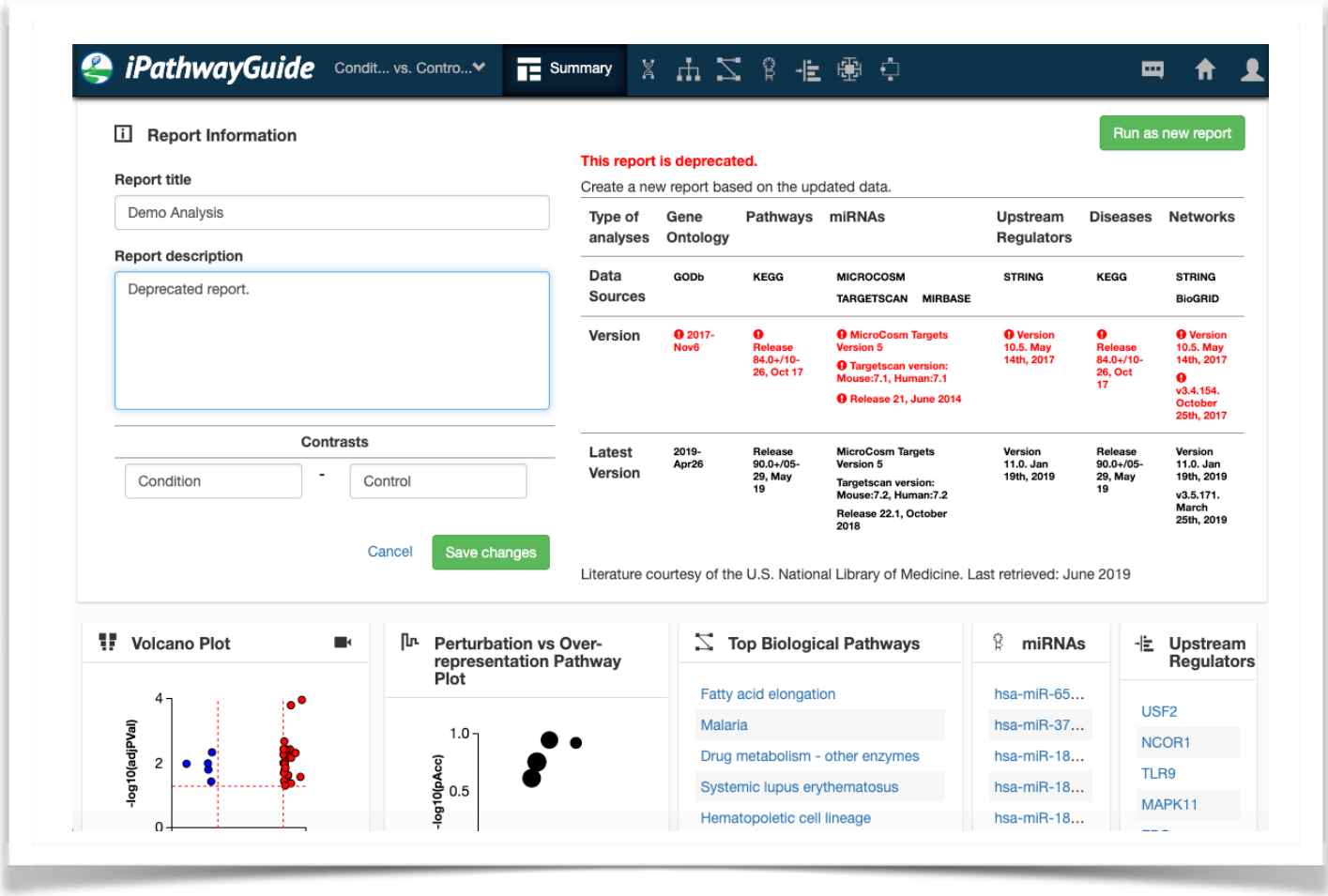
Editing Title, Description, Contrast labels

You may edit project information by clicking the "**+**" next to the Project title. Click "**Save Changes**" after you edit the text. If you are not the project owner, you cannot edit the project, but you may send an automated email to the owner to request an edit or update.

Updating a Project

We continuously update iPathwayGuide input datasets and periodically add new functionality to the software. These updates provide significant value to you as the user and we recommend that you update the analysis. When a new version of the knowledgebase or software is available, a message that "This report is deprecated" will appear and the report owner can "Run as new report" to get a new analysis based on updated underlying databases and/or with new capabilities.

- Note: If you do not have a subscription, updates will be available for purchase.
- If you have a subscription, updates are included in your quota.
- Whether you choose to update your analysis or not, the old analysis will be preserved and you may continue to use it.
- A history of each analysis is available from your dashboard. Click the "+" sign to the left of the analysis title to see the history.

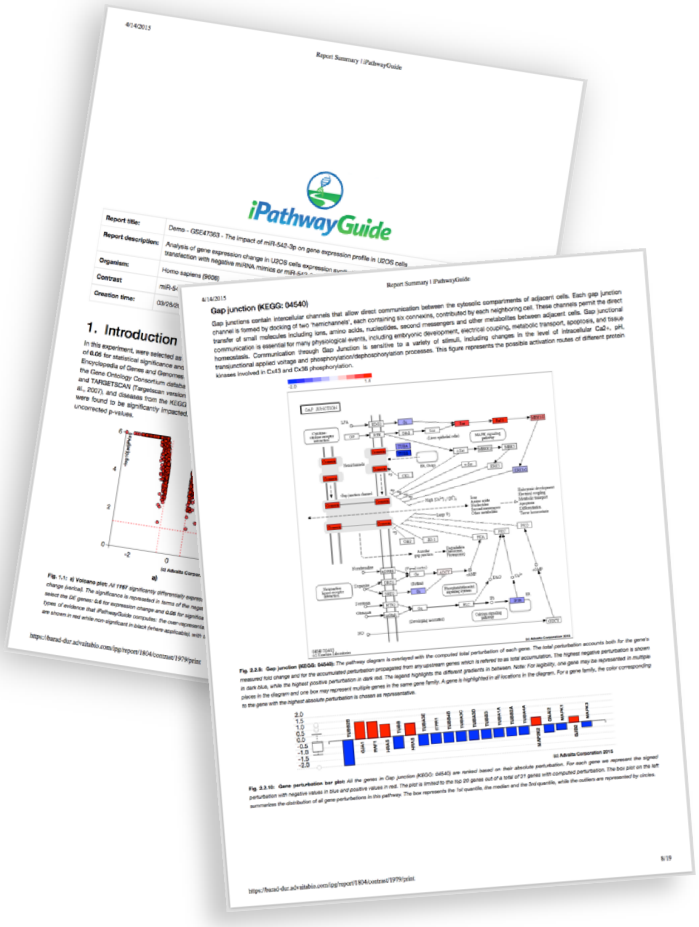


Printer friendly summary

Each purchased analysis report comes with an auto-generated, printable report. From the Summary page (just below the Volcano plot) click the printer icon to generate the printable report. Each report includes the following sections for the selected contrast:

- Introduction with input parameters and descriptive statistics of the experiment
- Detailed methods for each analysis performed
- Top 5 results for each analysis section with graphs, figures, and descriptions
- Analyses for:
  - Pathways
  - Predicted miRNAs
  - Gene Ontologies (3)
  - Diseases
  - Upstream Regulators
- Literature references

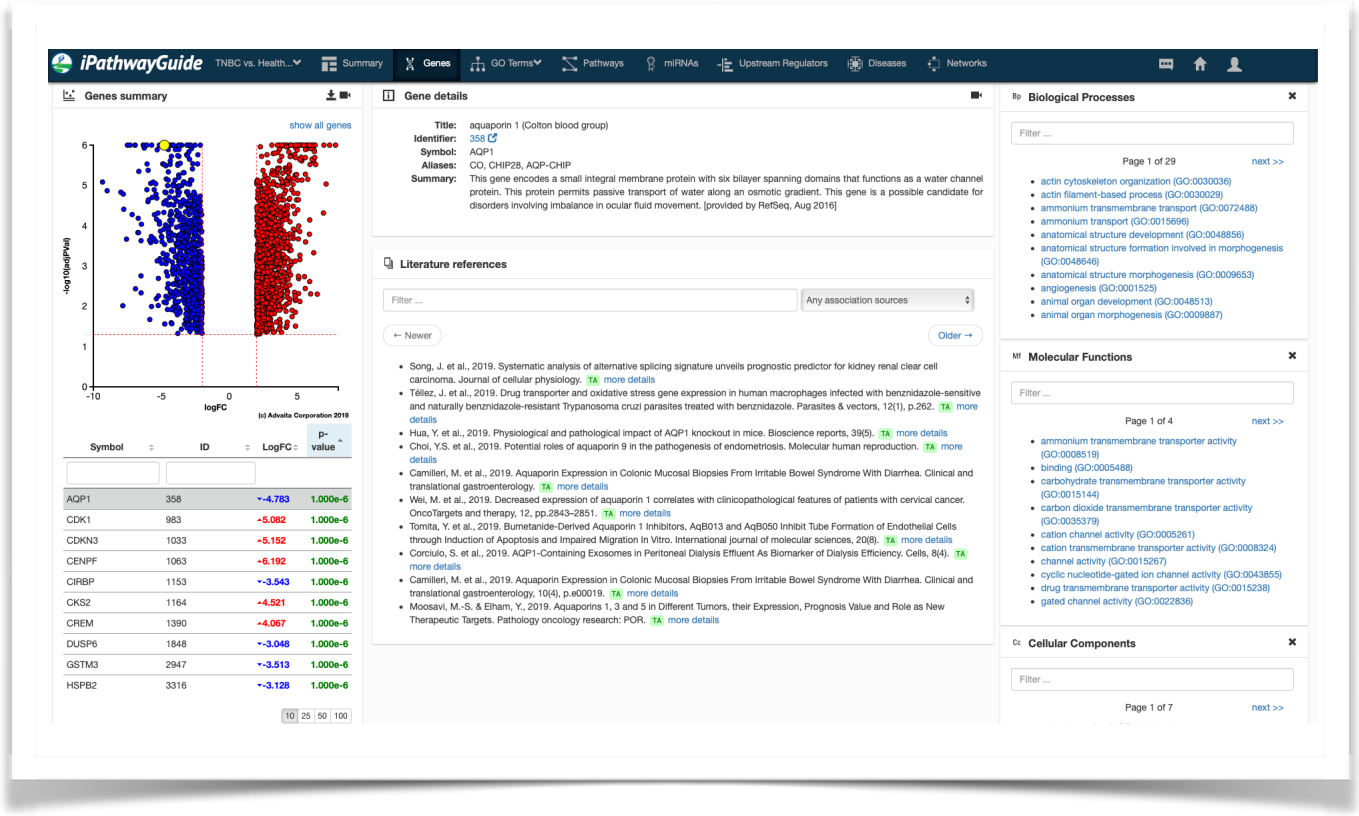
The printable report is an excellent resource to include as an executive summary for a client or collaborator. Also, the .pdf version could be used as supplemental information for a grant application or publication.



Genes

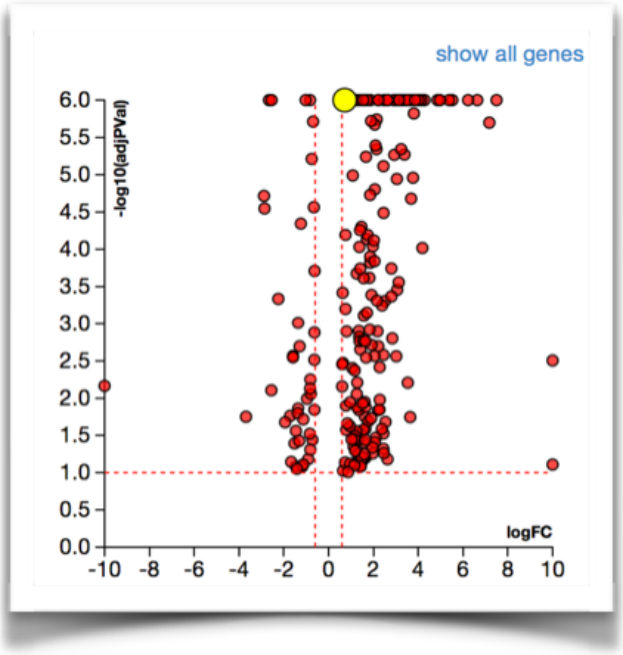
The Genes section provides detailed information about each gene that was input to the analysis. Genes may be selected either by clicking on the gene in the volcano plot or by selecting from the gene table below. You may also search for a gene by its official gene symbol or Entrez gene ID.

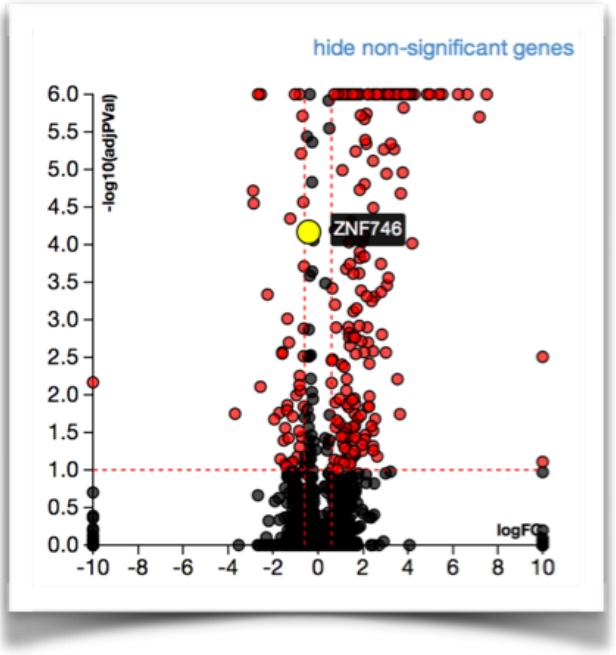
When a gene is selected, information about that gene is displayed, along with literature references and relationships to: biological processes, molecular functions, cellular components, pathways, drugs, miRNAs, SNPs, and diseases. Clicking on any of the links pertaining to another module will take you to that respective analysis within the application.



If you do not see your gene(s) of interest listed, you may choose to “show all genes” at the top right corner of the volcano plot. This will display all of the genes (significant and non significant) from the input file used to generate the analysis. It may take a few seconds to load all the genes from your input file. When this option is selected, both black and coloured dots are displayed. Red dots (up) and blue dots (down) identify significantly DE genes, while black dots identify non-significant genes. Note that scale of the volcano plot is limited to 10e-6.

When a black dot is selected, relationships to other elements like GO Terms and pathways will lead to external sources because that gene was not considered in the analysis.



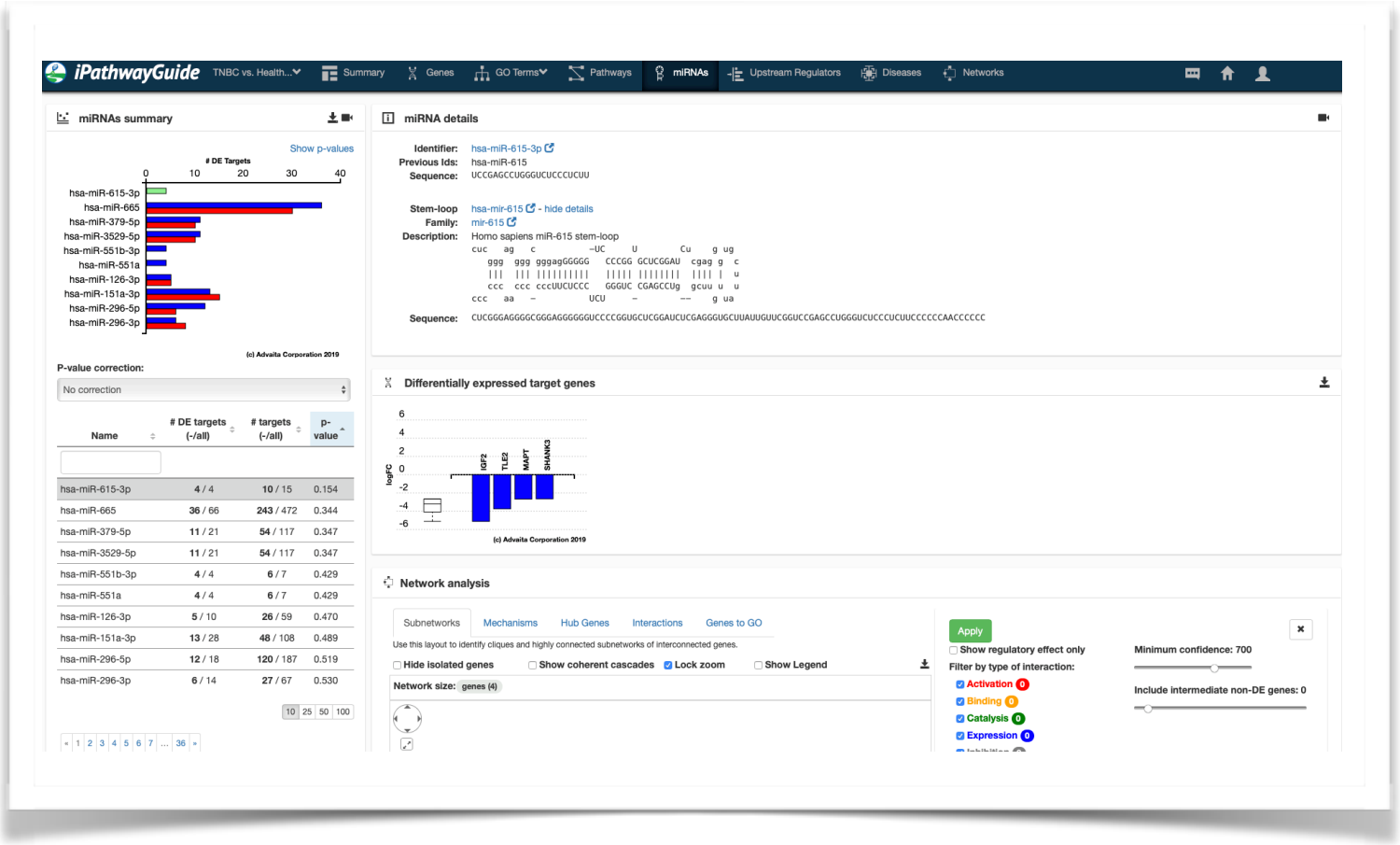


Predicted Micro RNAs (miRNAs)

Based solely on differential gene expression data, iPathwayGuide predicts miRNAs that might be active in your experiment (e.g. hypothesized to cause the observed down regulation of target genes). In many cases, these represent novel hypotheses on mechanisms active in your experiment.

The miRNA analysis has 5 key sections:

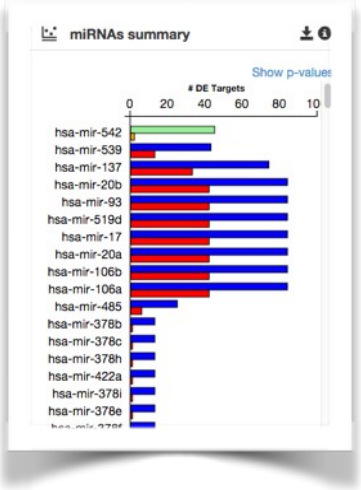
- miRNAs summary
- miRNA details
- Differentially expressed target genes
- Network Analysis
- Literature references

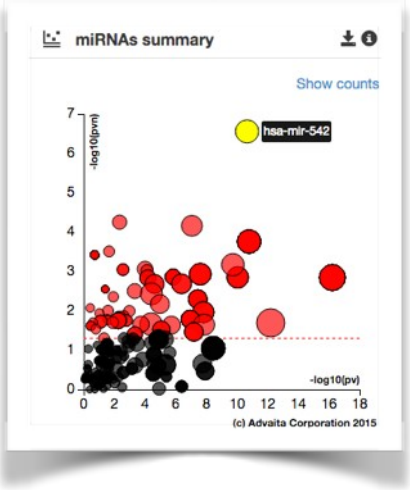


miRNA summary

In the miRNAs summary section, you may choose between two different views by clicking on the label in the top-right corner of the graphic.

Counts View **p-value View**





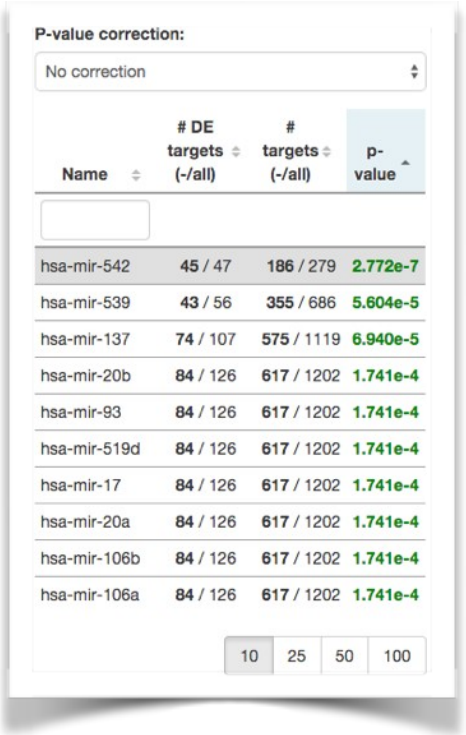
Counts View: The "Counts View" is a bar graph that shows the count of DE target genes that are down-regulated (blue bars) and the number of DE target genes that are up-regulated (red bars) for each miRNA. The selected miRNA bars change to green and orange to indicate its selection. By default the miRNAs are ordered by significance (most significant to least) without correction.

P-value View:
The vertical axis (-log10(pvn)) is the p-value based on the number of downwardly expressed DE targets versus the total number of DE targets. This p-value is used for identifying significant miRs.

The size of each dot is directly proportional to the number of target genes for that miRNA relative to other ones.

The horizontal axis (-log10(pv)) is the p-value based on the total number of DE target genes versus the total number of target genes (ignoring whether they are up or down regulated) This p-value is not used for identifying significant miRs.

Below the plots, is a table with the relative counts and global p-values for each miRNA. By default, no correction is applied, but you may choose to apply FDR or Bonferroni correction factors.



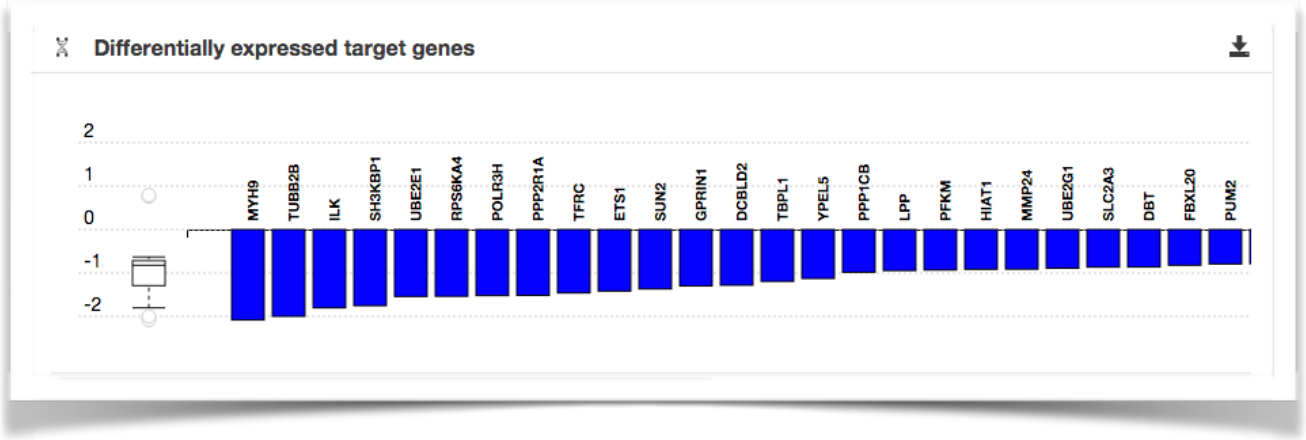
Name	# DE targets (-/all)	# targets (-/all)	p-value
hsa-mir-542	45 / 47	186 / 279	2.772e-7
hsa-mir-539	43 / 56	355 / 686	5.604e-5
hsa-mir-137	74 / 107	575 / 1119	6.940e-5
hsa-mir-20b	84 / 126	617 / 1202	1.741e-4
hsa-mir-93	84 / 126	617 / 1202	1.741e-4
hsa-mir-519d	84 / 126	617 / 1202	1.741e-4
hsa-mir-17	84 / 126	617 / 1202	1.741e-4
hsa-mir-20a	84 / 126	617 / 1202	1.741e-4
hsa-mir-106b	84 / 126	617 / 1202	1.741e-4
hsa-mir-106a	84 / 126	617 / 1202	1.741e-4

10 25 50 100

miRNA Details provide the description of the miRNA and a link to miRBase for additional information about the selected miRNA including the sequence and stem loop.

Differentially expressed target genes

The differentially expressed target genes plot displays target DE Genes for the selected miRNA with the relative measured fold change from the experiment. The DE targets are ordered from greatest negative FC to greatest positive FC. Depending on the number of DE targets associated to the selected miRNA, you may need to scroll the table to the right or left to view all the targets. Clicking on any one of the bar plots for the genes will take you to the genes tab for additional information about that gene.



miRNA Network Analysis

The network analysis included in the miRNA module is characteristic of the network analysis included in the other modules (e.g. GO, Pathways, Upstream Regulators, and Diseases). The network analysis in the miRNAs module will automatically display differentially expressed target genes for the selected miRNA. More information about how to use this section of the analysis can be found in the Networks section of this guide.

Literature references

Related literature references are provided for the selected miRNA. These references can be filtered by entering a term into the filter bar.

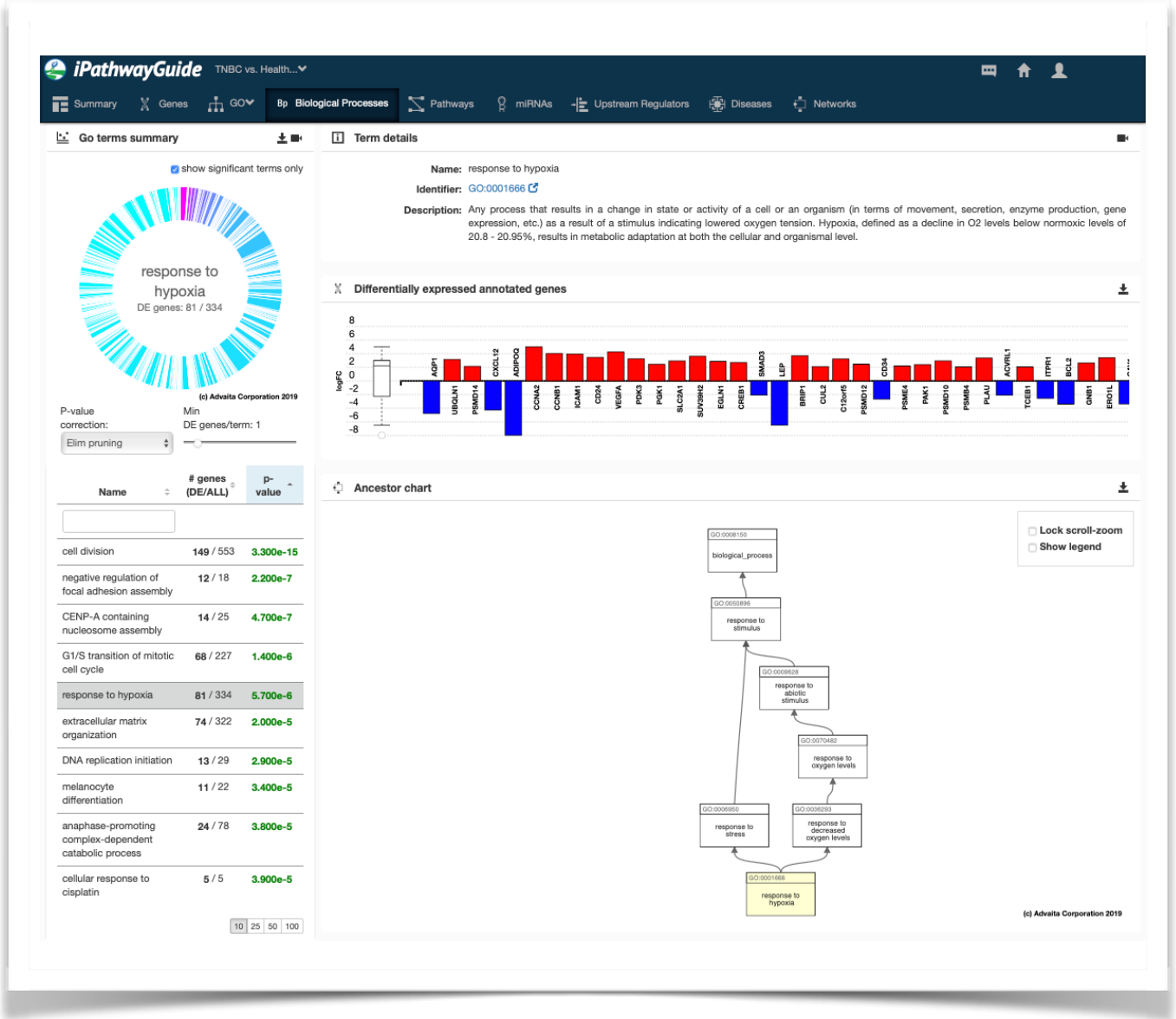
Gene Ontology

From the Gene Ontology (GO) analysis menu, you can select one of following domains:

- Biological Processes
- Molecular Functions
- Cellular Components

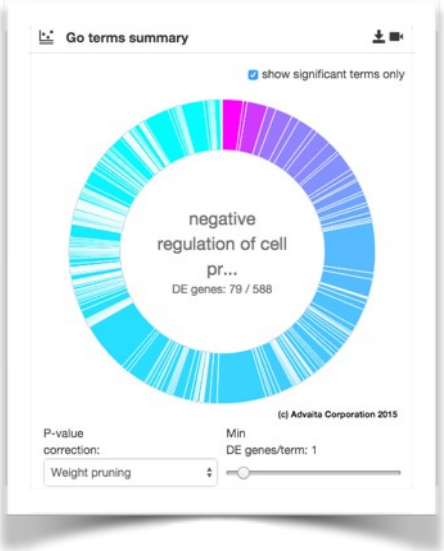
The layout and navigation is identical for all three GO analyses.

On the left is the GO plot with the table with GO terms ranked by significance and color coded. On the right is the detailed information about the selected GO terms and below that is the ancestor chart. More information about GO Analysis can be found at http://www.ebi.ac.uk/QuickGO/.



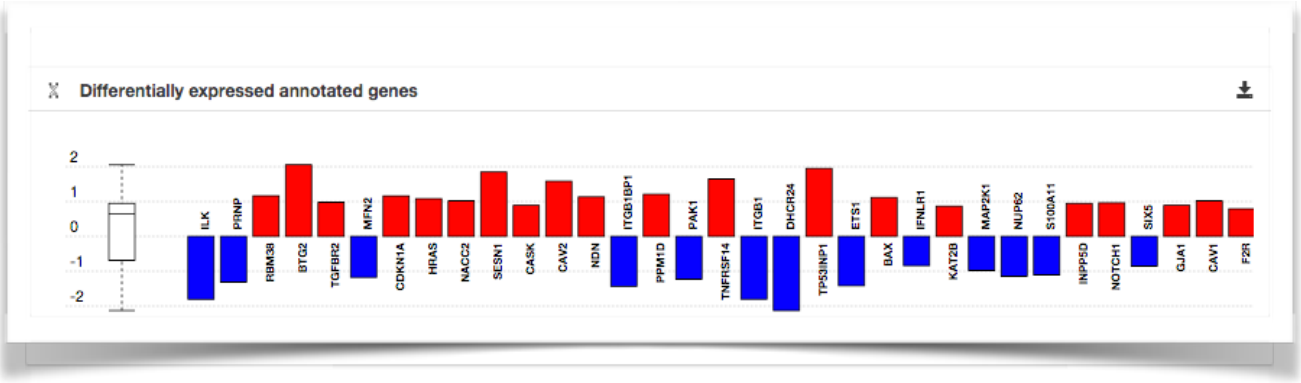
GO Plot

The GO plot is a graphical representation of the GO terms ranked as per their enrichment scores. The top-center part of the circle (magenta block) represents the most significant go term and the significance decreases in clock-wise direction. Magenta blocks are more significant than cyan blocks. By default, only significant terms are displayed. Various correction factors (FDR, Bonferroni, Elim Pruning, Weight Pruning) and thresholds of minimum DE genes can be applied, and the table and plot will automatically update. More information about Elim and Weight Pruning can be found at Alexa et al (2006).



Differentially Expressed Annotated Genes

For each GO term you will see the differentially expressed genes that are annotated for that term with the measured expression change and a box-plot for the overall expression. Clicking on any one of the genes bars will take you directly to the Genes tab with that gene selected.

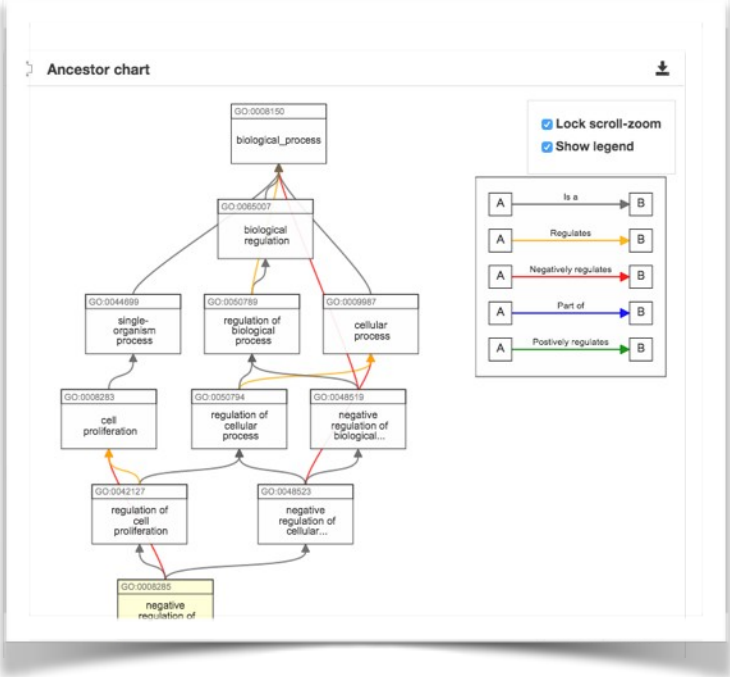


Ancestor Chart

For each GO term, an ancestor chart is provided for further analysis. The chart is interactive and can be dragged, zoomed, and manipulated. Clicking on a GO ID will take you to the chart with that node as the root node.

You can lock scroll-zoom to adjust the graphic.

You can also show the legend



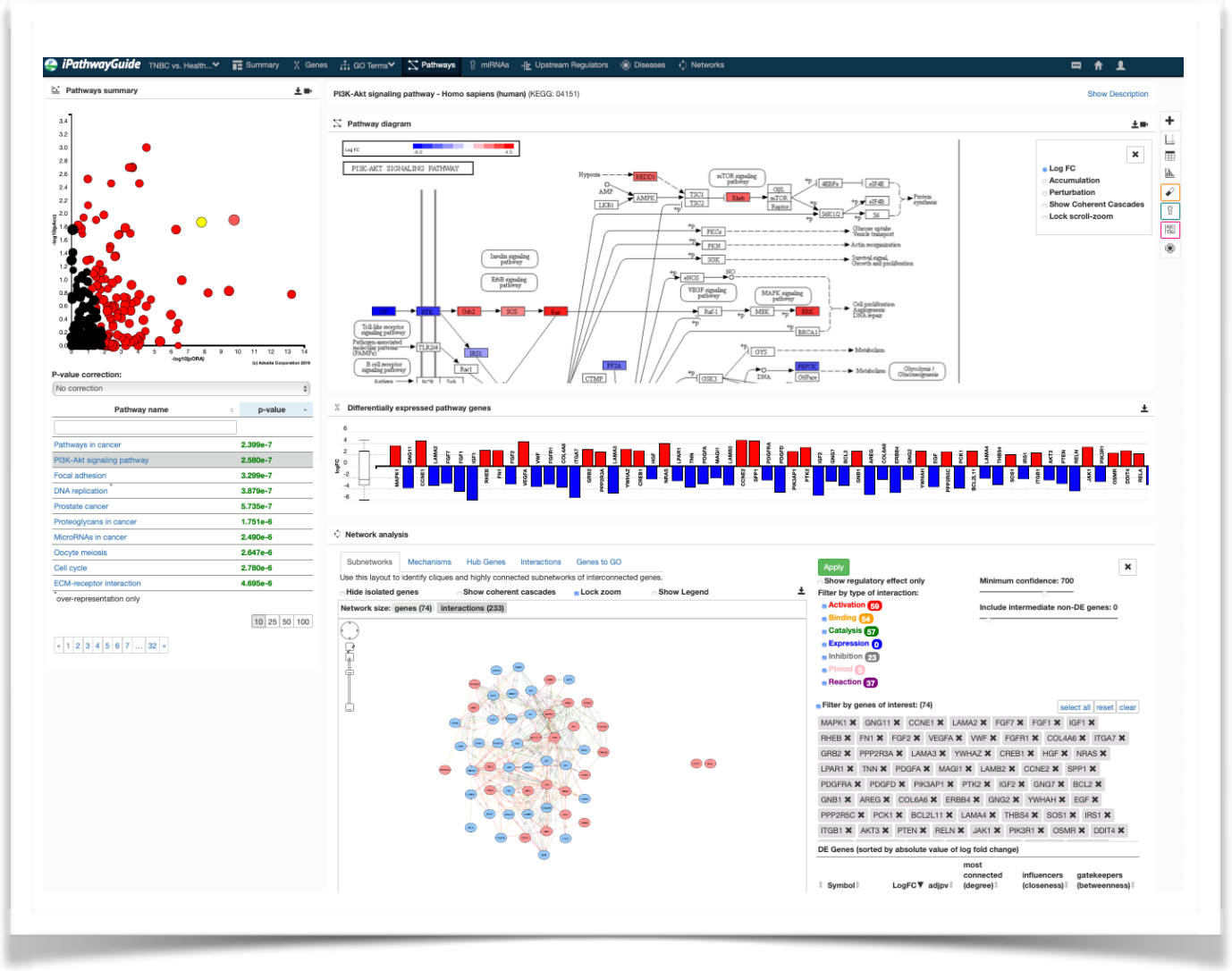
GO Network Analysis

The network analysis included in the GO module is characteristic of the network analysis included in the other modules (e.g. Pathways, miRs, Upstream Regulators, and Diseases). The network analysis in the GO module will automatically display differentially expressed target genes for the selected GO term. More information about how to use this section of the analysis can be found in the Networks section of this guide.

Literature References

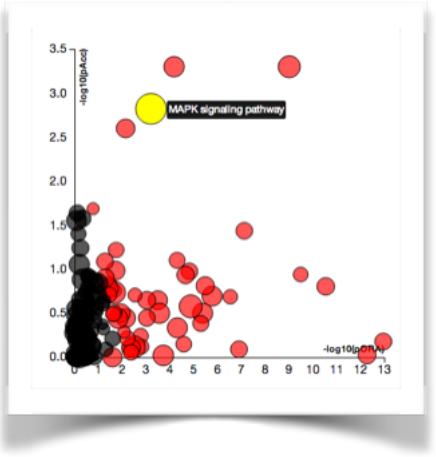
Related literature references are provided for the selected GO Term. These references can be filtered by entering a term into the filter bar.

Pathways



The Pathways section displays the pathway analysis results. On the left is the two-way pathway plot (top) and table (middle) with pathways sorted by p-value. By default, no correction factor has been applied. Both FDR and Bonferroni correction factors can be applied by selecting the correction factor from the spinner between the plot and table. A description of the currently selected pathway can be displayed by clicking on the "show description" label on the top-right of the window. Pathways can be selected by clicking on the dots in the two-way plot or by clicking on the link in the table.

Pathway Plot



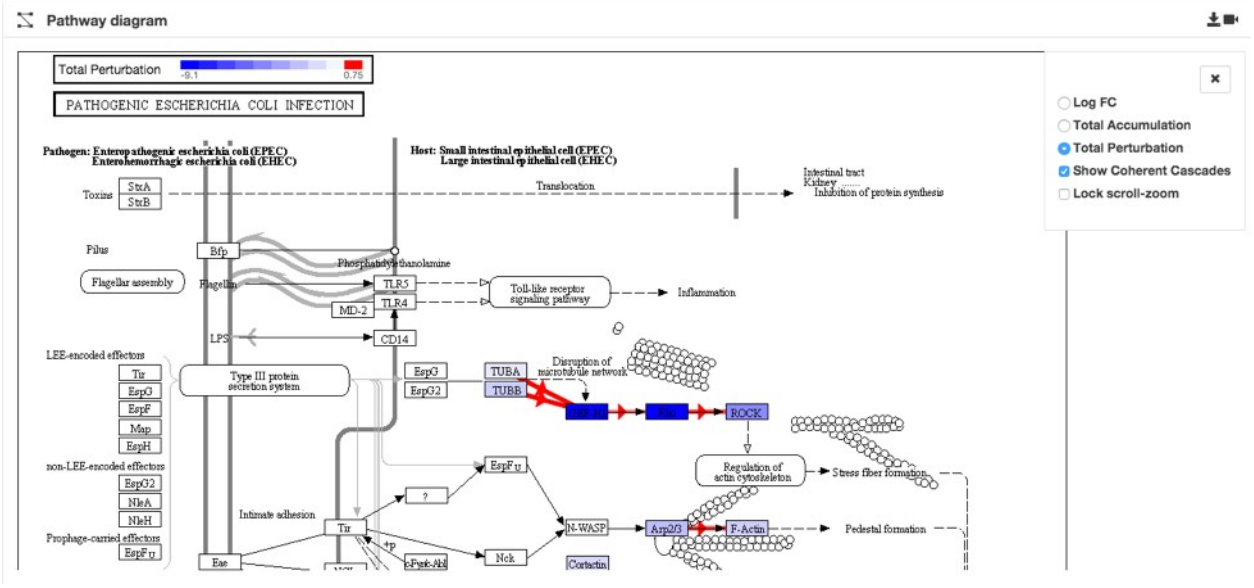
The pathway plot displays the unique character of iPathwayGuide's Impact Analysis, which accounts for both over-representation (enrichment) of DE genes on a pathway AND the topology of the pathway. The x-axis measures the p-value obtained using the classical over-representation analysis (pORA). The y-axis represents the p-value obtained from total perturbation accumulation (pAcc) in the pathway.

Each dot represents a pathway. The size of each dot denotes the number of genes on the pathway. The table directly below the plot shows the combined global p-value for each pathway. More information about this method can be obtained here.

Pathway Diagram

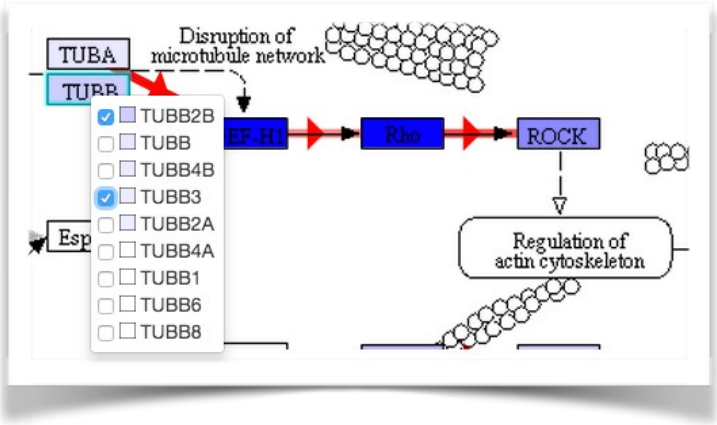
The pathway diagram, by default, displays the measured fold change for each gene on the pathway (LogFC). This is a graphical view of your experimental data applied to the selected pathway.

Total Accumulation and Total Perturbation (Fold Change + Total Accumulation) may be selected with the radio buttons on the top-right of the diagram. "Total Accumulation" represents the propagation of the effect of changes in expression (of upstream genes) on downstream genes, based on documented interactions between genes. "Total Perturbation" for a given pathway is the combination of both, the **measured LogFC** and **Total Accumulation**, showing the **total perturbation** within the pathway.



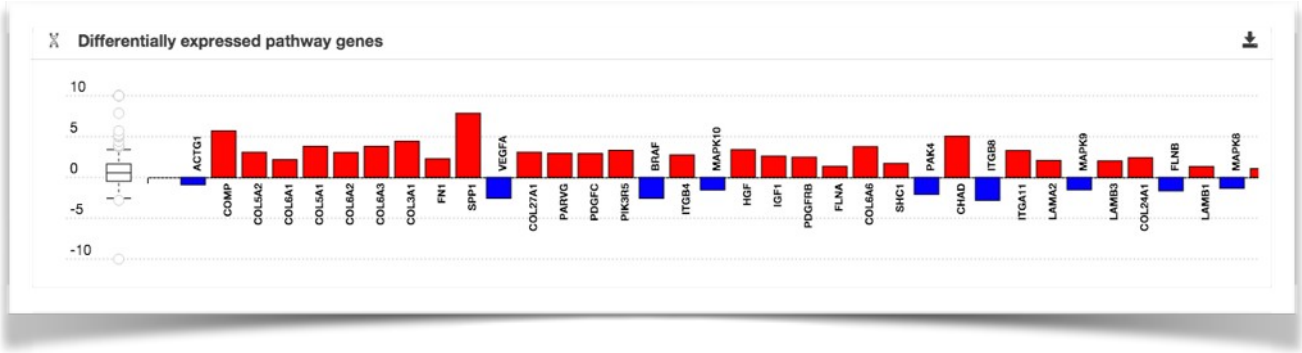
Coherent Cascades can be displayed by clicking the radio button. These are sections of the pathway where the gene expression data and the pathway diagram are coherent (i.e. the pathway diagram shows gene A represses gene B and the data are consistent with that state). These Coherent Cascades can serve as putative mechanisms for all the observed and calculated perturbations in the system.

- Note that some gene icons on the pathway may represent complexes of related genes. Hovering over a desired gene activates a popup with the list of genes contained in the complex with color-coding for fold change, accumulation, or total perturbation.



Differentially Expressed Pathway Genes

For each pathway, you can see the differentially expressed genes that are described by the selected pathway with their measured expression change and a box-plot for the overall expression. Clicking on any one of the genes bars will take you directly to the Genes tab with that gene selected.



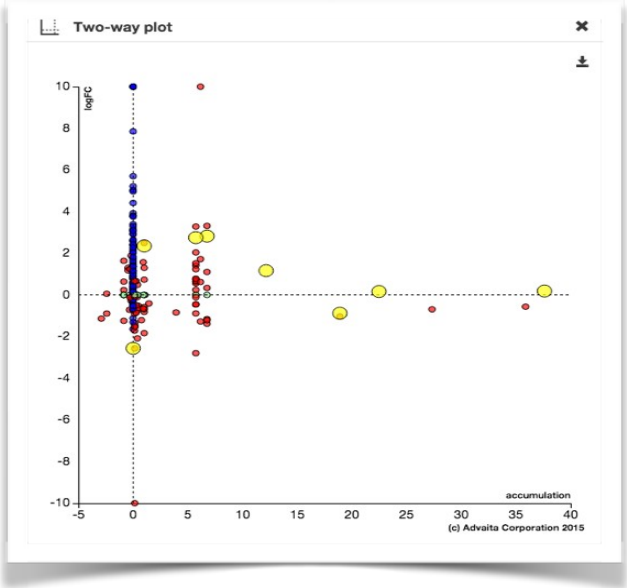
Pathway Details

Detailed information about the pathway can be obtained by clicking any of the buttons on the right hand side of the pathway map. The Pathway Details view allows for the interrogation of details about the specific pathway of interest. Details include:

- Two-Way Plot
- Gene Table
- Bootstrap Distribution (for perturbation)
- Drugs
- miRNAs
- SNPs
- Diseases

Two-Way Plot

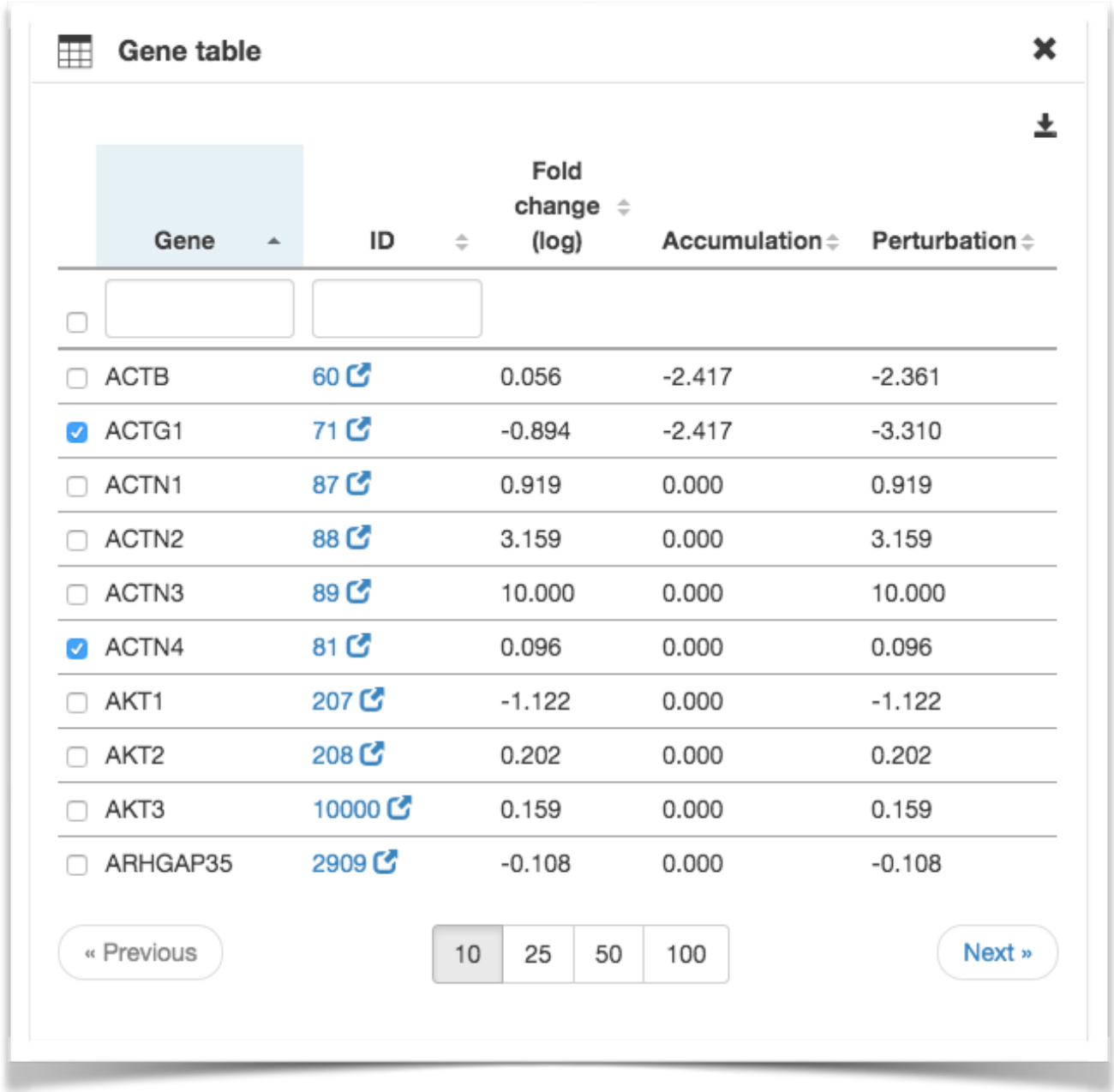
In the two-way plot, the y-axis represents the measured fold-change for each gene, while the x-axis represents the accumulation. Clicking a gene on the pathway will also highlight it in the Two-Way Plot.



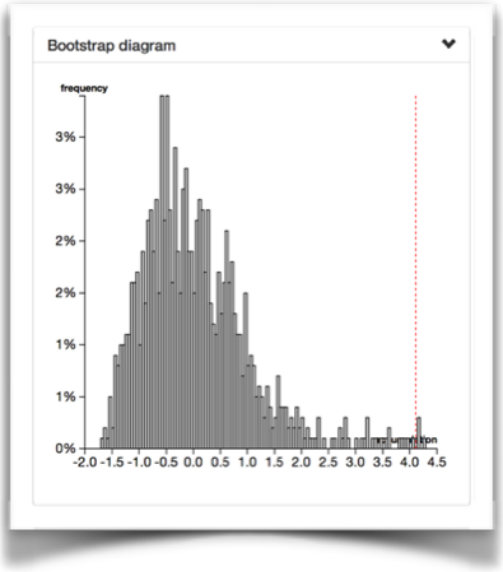
Gene Table

The gene table contains the list of genes represented in the selected pathway. The table displays the gene symbol, Entrez gene id, fold change, accumulation, and total perturbation.

Each column can be sorted by clicking on the headers. Individual genes can be searched by using the search bar at the top of the table and selected either via the checkbox on the left column or in the pathway diagram. Selecting the checkbox in the header will display only selected genes.



Bootstrap Diagram

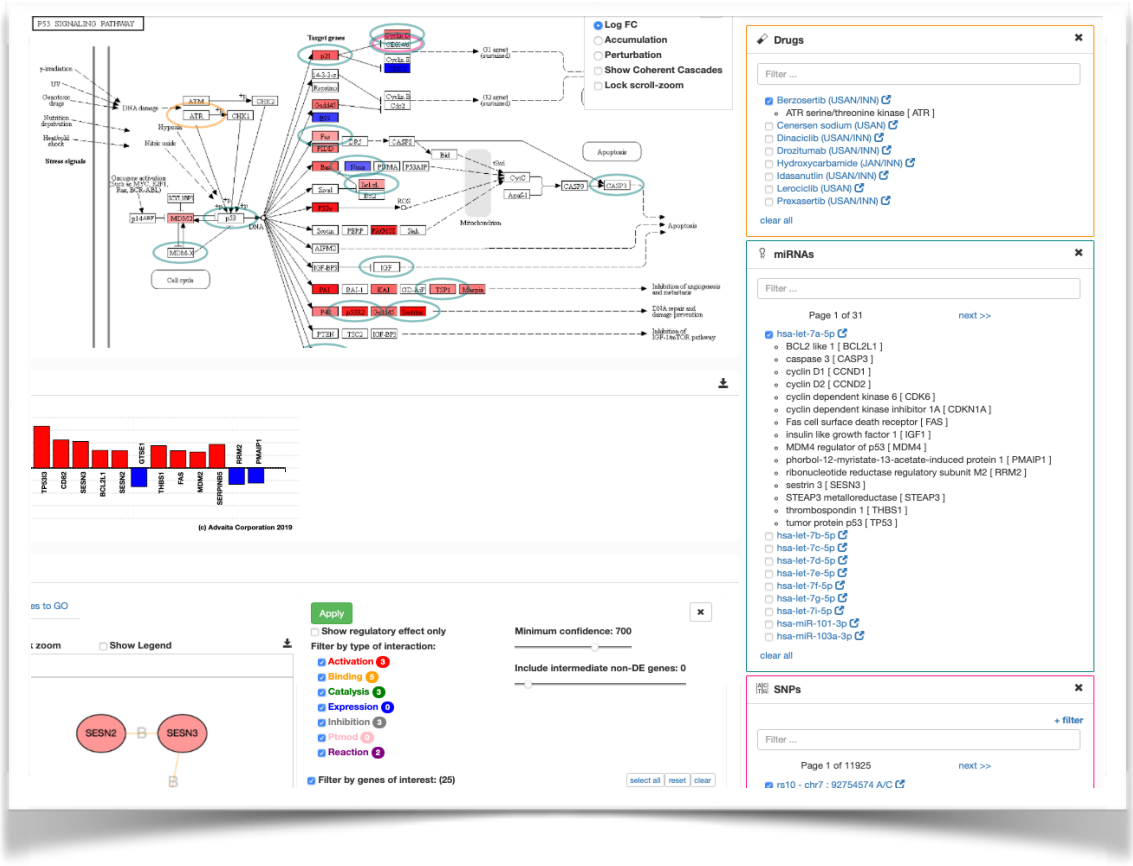


The bootstrap diagram shows the distribution (mean=0, std dev=1) of expected accumulated perturbation values based on the observed data for the selected pathway. 2,000 iterations are used to construct the distribution. The red line indicates where the observed value lies in the distribution of expected results. The further away from the mean, the less likely the observed values are due to random chance. This is the value used for the p-value on the vertical axis in the Pathway Plot.

Drugs, miRNAs, SNPs

The related tables for Drugs, miRNAs, and SNPs list those items known to be associated with the genes on the selected pathway. Selecting a gene filters the list to display those items only related to that gene. Clicking on a check box will highlight the target gene(s) in the pathway. Clicking on the name will bring up additional information about that item.

miRNAs and SNPs have filtering capabilities. For SNPs, the default filter only shows SNPs that are both validated and clinically significant.



Pathways Network Analysis

The network analysis included in the Pathways module is characteristic of the network analysis included in the other modules (e.g. GO, miRs, Upstream Regulators, and Diseases). The network analysis in the Pathways module will automatically display differentially expressed target genes for the selected pathway. More information about how to use this section of the analysis can be found in the Networks section of this guide.

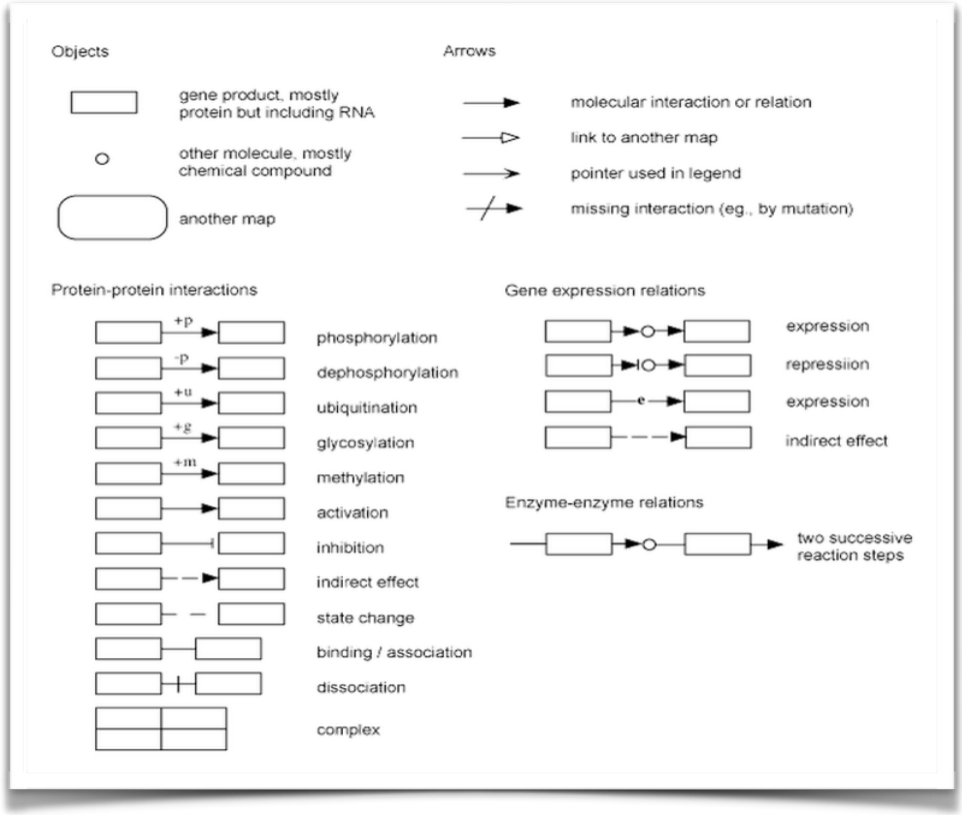
Literature references

Related literature references are provided for the selected Pathway. These references can be filtered by entering a term into the filter bar.

Pathway Legend

See below for a legend of symbols for KEGG pathway diagrams.

Additional information on how to read a KEGG pathway image can be found at https://www.genome.jp/kegg/document/help_pathway.html

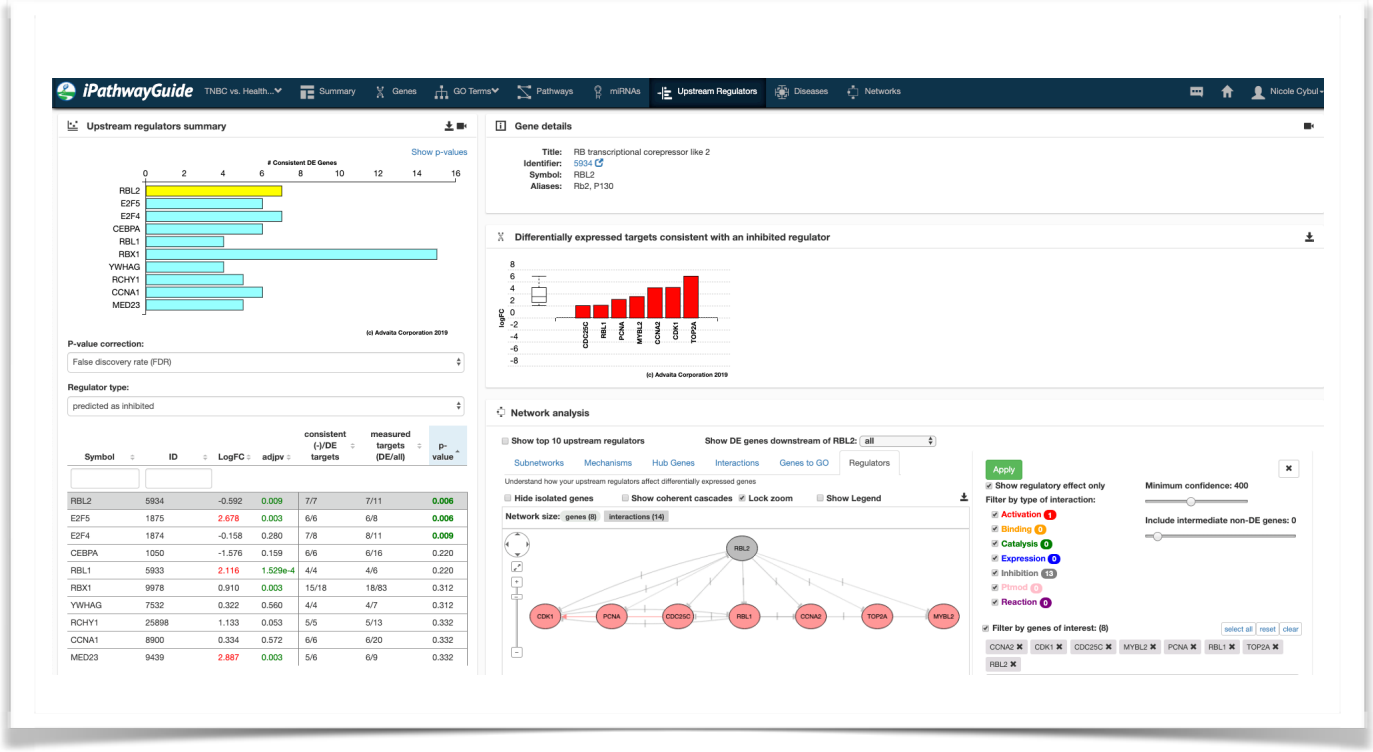


Upstream Regulators

The Upstream Regulators module uses the differential expression of genes from the experimental data to predict upstream regulators that may be activated or inhibited in the experiment. This information can be useful in understanding which upstream components may be influencing a certain condition or phenotype. These relationships may also represent novel hypotheses for mechanisms active in your experiment.

The following sections are included in the Upstream Regulators module:

- Upstream regulators summary
- Gene details
- Differentially expressed target genes
- Network analysis
- Literature references



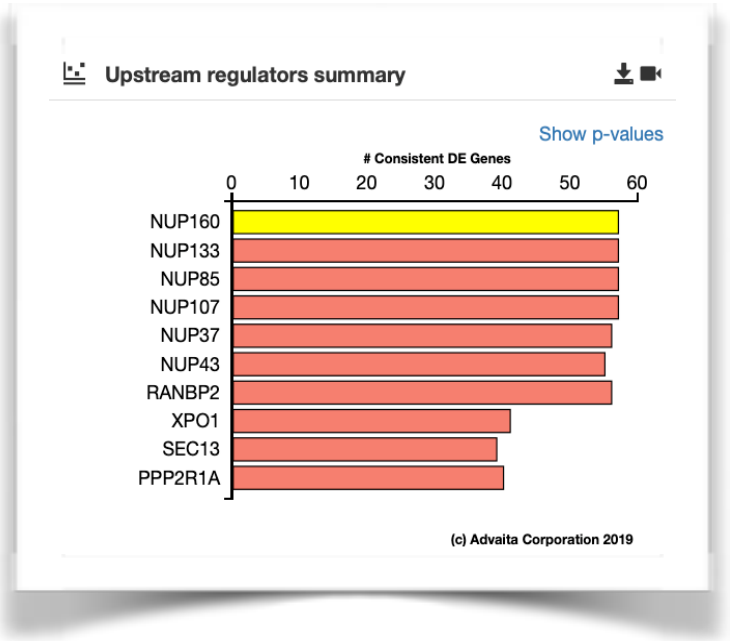
Upstream regulators summary

Regulator Type:

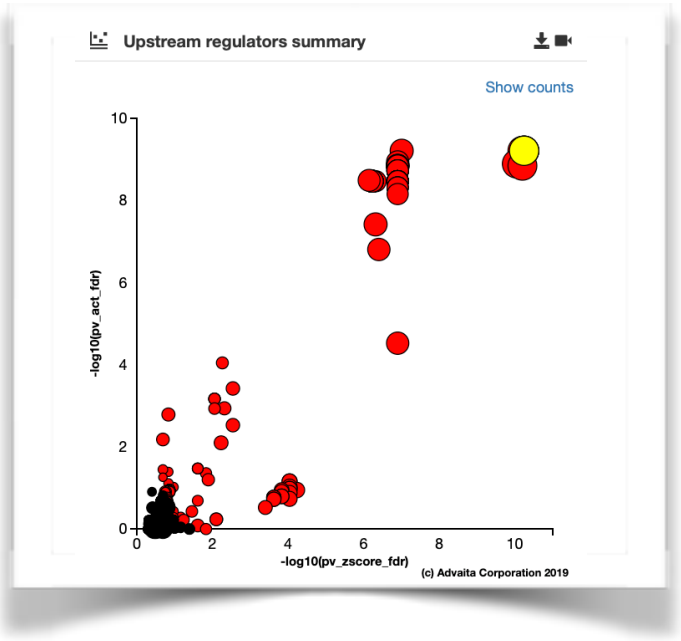
iPathwayGuide predicts both activated and inhibited Upstream Regulators, based on the differential gene expression observed downstream. You can switch between activated and inhibited upstream regulators by selecting from the options below “Regulator type.”

In the upstream regulators summary section, you may choose between two different views by clicking on the blue text in the top-right corner.

Counts View



p-value View



Counts view: The "Counts View" is a bar graph that shows the count of consistent DE target genes downstream of the regulator. Clicking on one of the bars will select that upstream regulator, changing the color of the bar from orange or aqua to yellow. By default the upstream regulators are ordered by p-value (most significant to least significant).

P-value view: The "P-value view" shows each Upstream Regulator as a dot. The x-axis position is the log of the p-value based on the Z-score while the y-axis position is the log of the overrepresentation p-value based on the number of consistent DE target genes. The size of each dot represents the number of consistent DE genes for each regulator. The selected upstream regulator appears in yellow.

Below the plots, you can select which p-value correction you would like to apply. By default, an FDR correction is selected, but you may also choose to apply a Bonferroni correction or no correction.

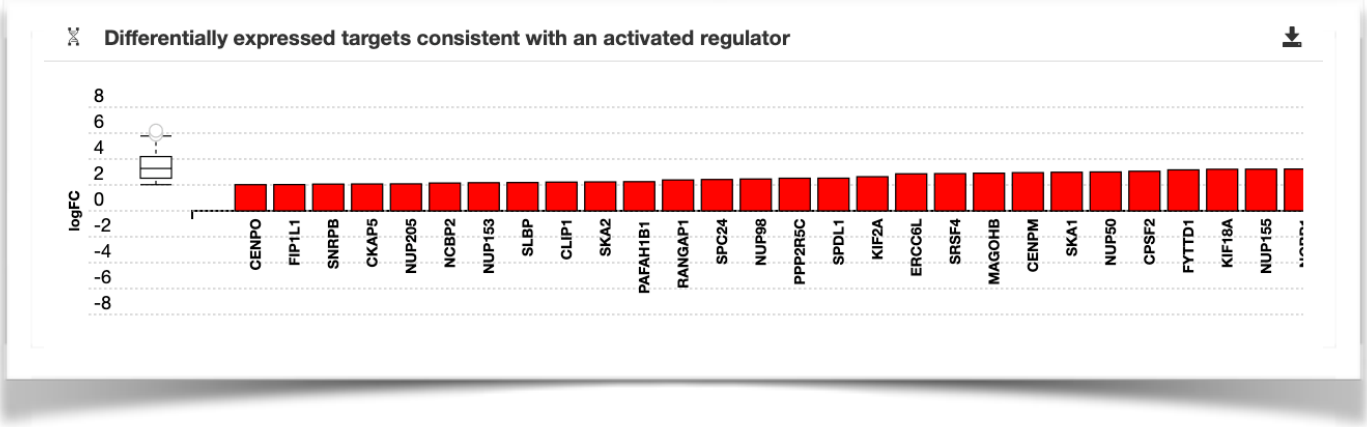
A table listing the upstream regulators is displayed below the summary plot. From left to right, this table includes columns for the gene symbol, the gene ID, the LogFC measured in the experiment, the adjusted p-value of the upstream regulator determined from the experiment (for any genes that were measured), the number of downstream targets consistent with either a predicted or inhibited upstream regulator divided by the total number of DE targets, the number of DE targets divided by the total number of measured downstream targets, and the p-value associated with the prediction that the gene is either an activated or inhibited upstream regulator. By default, the upstream regulators are ordered from lowest p-value to highest p-value. You can also sort by any other column. Clicking the arrows next to one of the column titles will sort the table based on that column.

Gene Details

The Gene details section provides specific information about the selected upstream regulator, including its title, identifier, symbol, aliases, and summary. This information is the same information found in the genes module.

Differentially expressed target genes

The differentially expressed targets genes plot displays target DE genes that are consistent with either an activated or inhibited upstream regulator, depending on the chosen regulator type, with the relative measured fold change (FC) from the experiment. The DE targets are ordered from greatest negative FC to greatest positive FC. Depending on the number of DE targets consistent with the regulator, you may need to scroll the table to the right or left to view all the genes. Clicking on any one of the bar plots for the genes will take you to the genes tab for additional information about that gene.



Network analysis

The network analysis in the Upstream Regulators module will automatically display differentially expressed genes associated with the selected upstream regulator. Choosing whether to show all, none, or only the consistent DE genes downstream of the regulator can be achieved by selecting one of the choices below the “Show DE genes downstream” label and clicking “Apply.”

It is also important to note that the Regulators mode for Network Analysis is only available within the Upstream Regulators module. Here, you can also show the top 10 upstream regulators by checking the box labelled “Show top 10 upstream regulators” and then clicking “Apply.” This will show you if and how the top ten upstream regulators interact with the genes downstream of the selected Upstream Regulator.

More information about how to use this section of the analysis can be found in the Networks section of this guide.

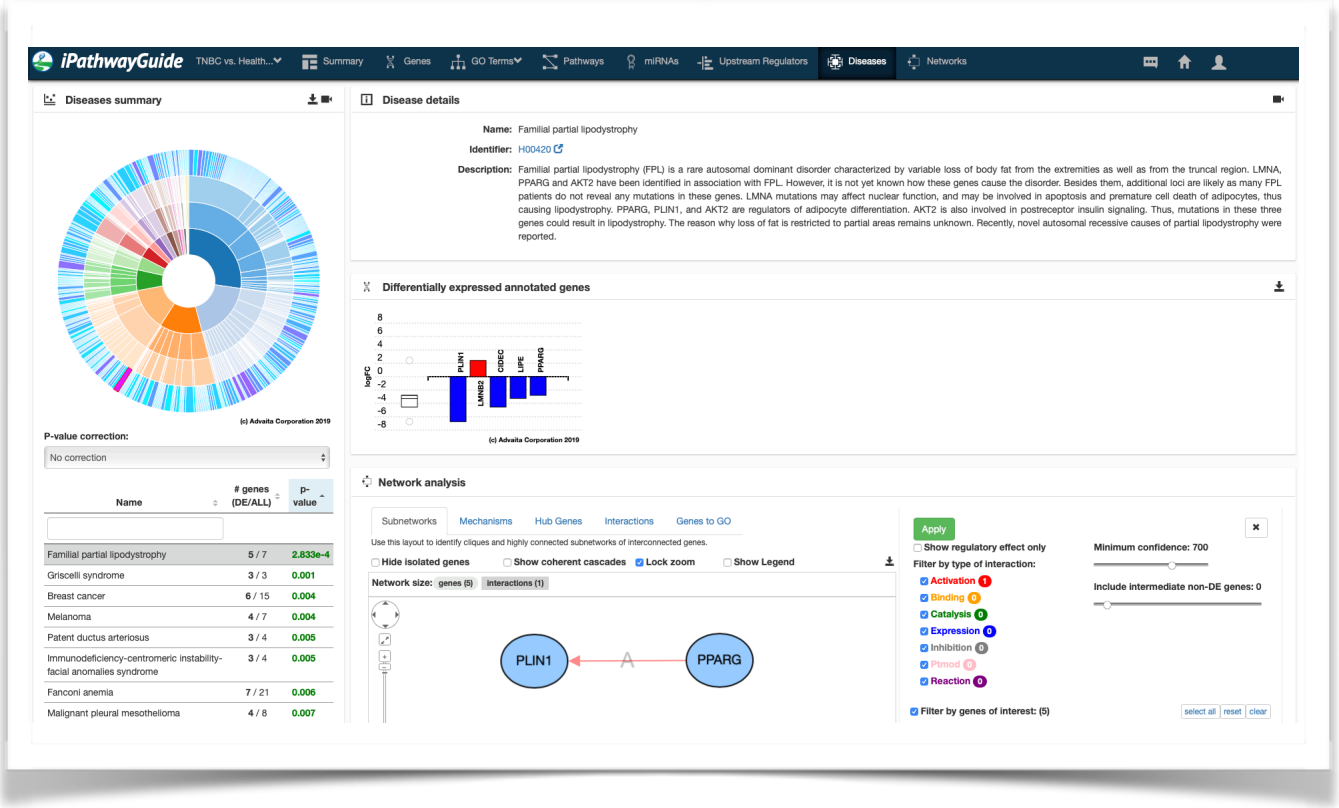
Literature references

Related literature references are provided for the selected upstream regulator. These references can be filtered by entering an author, key word, etc into the search bar.

Diseases

The Diseases tab displays the over-represented diseases based on the DE genes found in the analysis. The diseases are categorized by the International Classification of Disease – 2010 (ICD-10). As with the GO analysis, magenta blocks are more significant than cyan blocks.

Working from the interior of the graph to the exterior, users may narrow the list of diseases based on groups and subgroups of diseases. For example, clicking on the "Neoplasms" groups will display those disease subgroups over-represented in the data associated with Neoplasms.



For each disease term you will see the differentially expressed genes that are directly annotated for that term, with the measured expression change and a box-plot for the overall expression. Clicking on any one of the genes bars will take you directly to the Genes tab with that gene selected. The network analysis section of this module will automatically display differentially expressed genes related to the selected disease.

Diseases Network Analysis

The network analysis included in the Diseases module is characteristic of the network analysis included in the other modules (e.g. GO, miRs, Upstream Regulators, and Upstream Regulators). The network analysis in the Diseases module will automatically display differentially expressed target genes for the selected pathway. More information about how to use this section of the analysis can be found in the Networks section of this guide.

Literature references

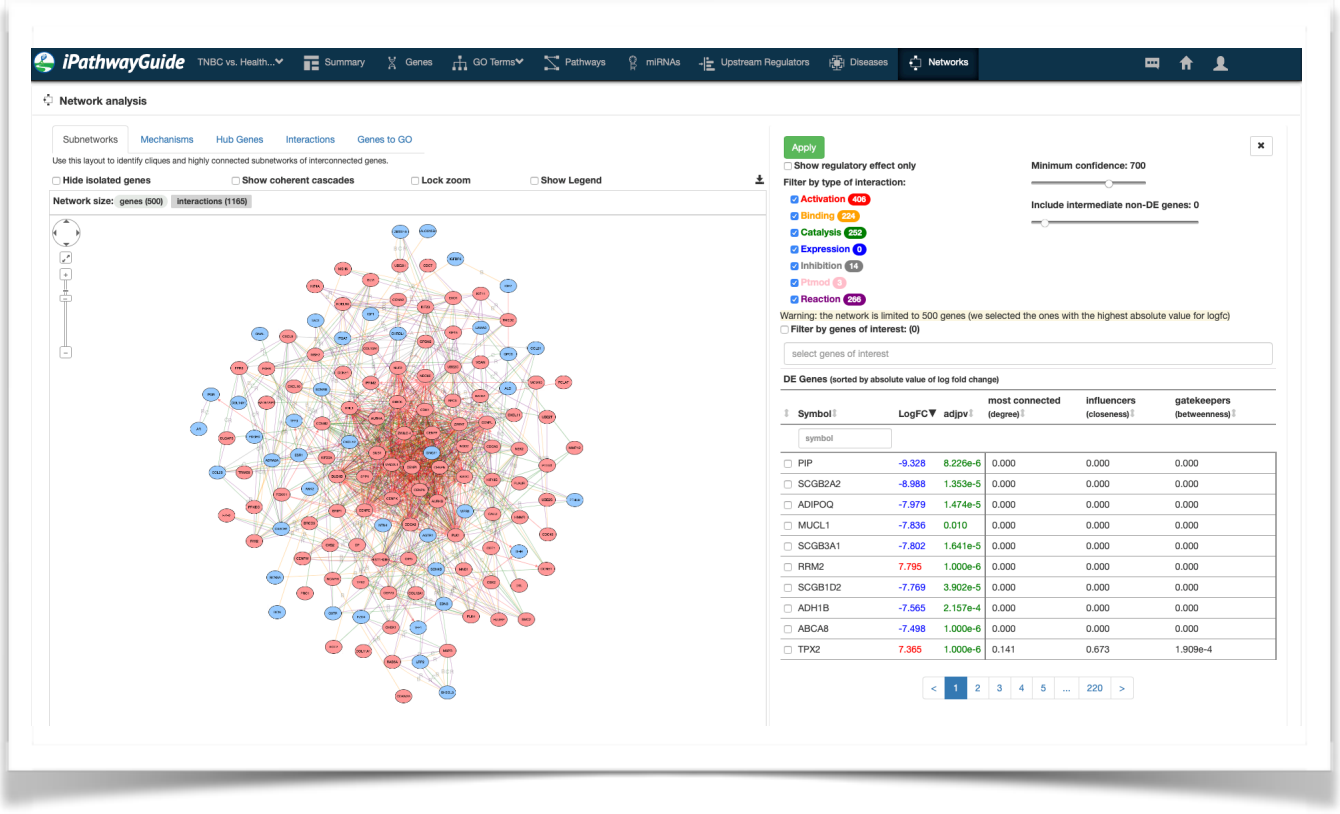
Related literature references are provided for the selected Disease. These references can be filtered by entering a term into the filter bar.

Networks

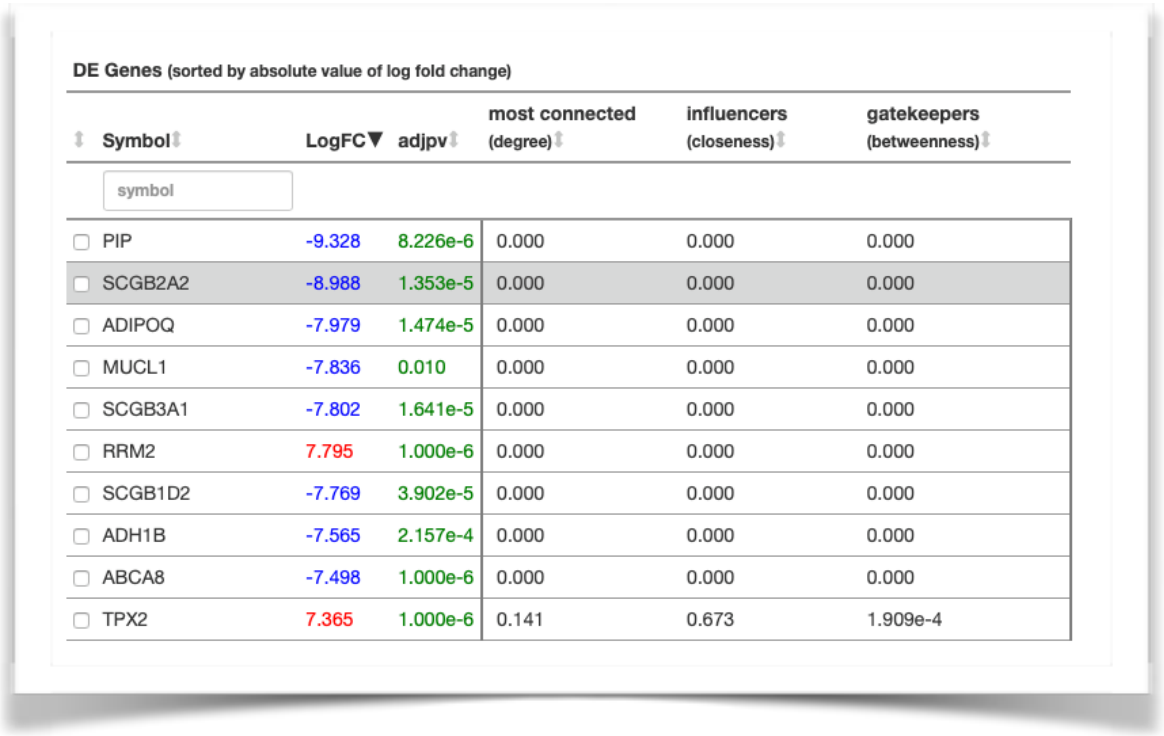
The networks module integrates the differential expression data from your experiment with known gene-gene interactions annotation to generate novel networks that may be important in understanding your experimental data. It allows you to identify putative mechanisms, analyze novel networks of interest, and relate genes to GO terms.

DE Genes

The central component of the Networks page is the table of DE genes in the bottom-right corner. By default the DE genes with the highest absolute value for LogFC are included in the diagram. Only 500 DE genes can be analysed at once, but you can choose exactly which genes you want to include. Selecting the box next to each gene in the table will add it to your list of genes of interest. Once you have selected all the genes you want to include in the network, click “Apply.” This will update the network to reflect your selections.



The table also includes important measures of centrality that you can use to sort genes and quickly find those that are significant. Genes that are “most connected” share a relatively high number of edges with other genes; genes that are relatively close to a large number of other genes are known as “influencers”; and genes that lie between a large number of genes on one side and a small number of genes on the other are known as “gatekeepers.”



DE Genes (sorted by absolute value of log fold change)

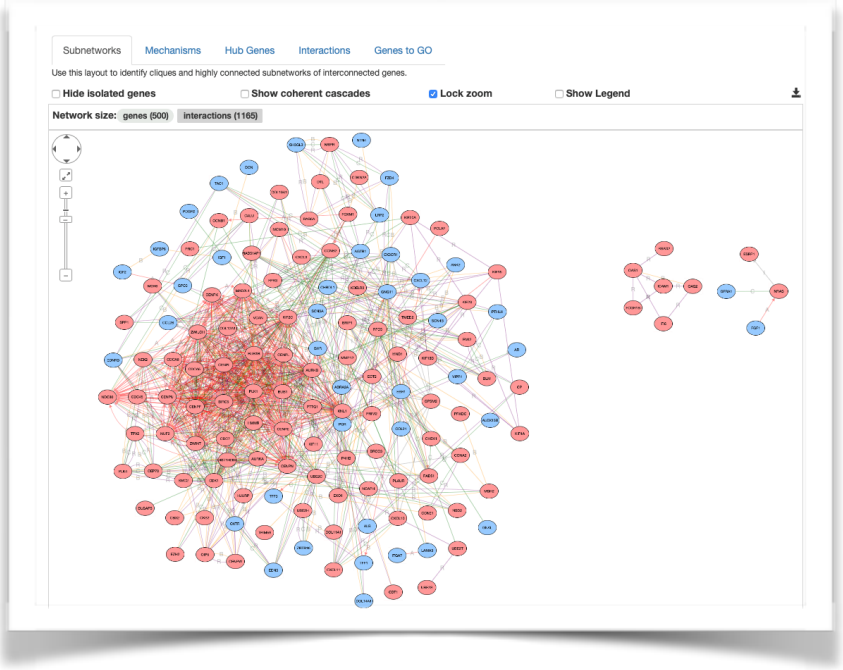
Symbol	LogFC	adjpv	most connected (degree)	influencers (closeness)	gatekeepers (betweenness)
symbol					
PIP	-9.328	8.226e-6	0.000	0.000	0.000
SCGB2A2	-8.988	1.353e-5	0.000	0.000	0.000
ADIPOQ	-7.979	1.474e-5	0.000	0.000	0.000
MUCL1	-7.836	0.010	0.000	0.000	0.000
SCGB3A1	-7.802	1.641e-5	0.000	0.000	0.000
RRM2	7.795	1.000e-6	0.000	0.000	0.000
SCGB1D2	-7.769	3.902e-5	0.000	0.000	0.000
ADH1B	-7.565	2.157e-4	0.000	0.000	0.000
ABCA8	-7.498	1.000e-6	0.000	0.000	0.000
TPX2	7.365	1.000e-6	0.141	0.673	1.909e-4

There are 5 different modes offered in the Networks module which are designed to help you answer specific questions about your experiment:

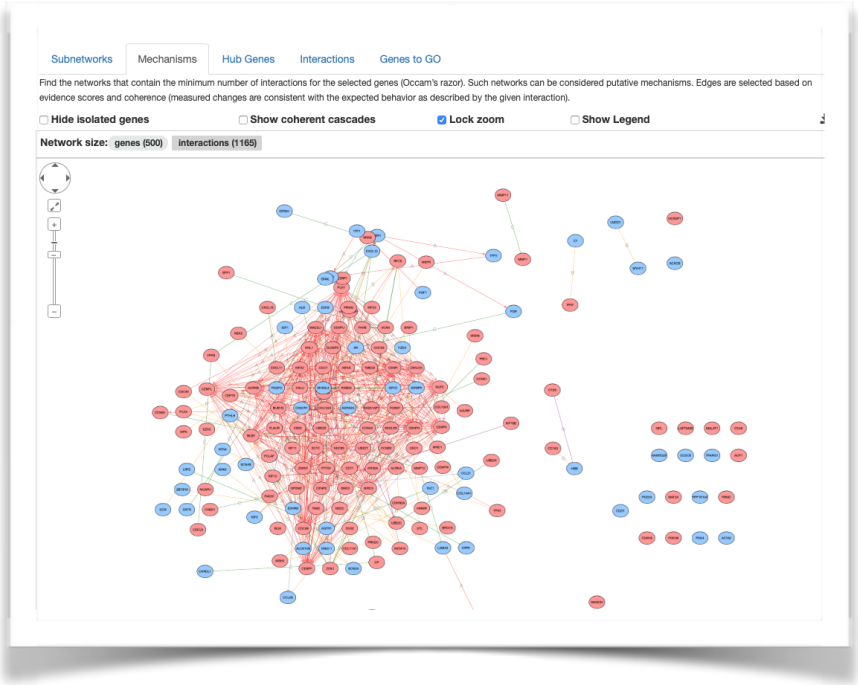
- Subnetworks
- Mechanisms
- Hub Genes
- Interactions
- Genes to GO

In each mode, there are many features under your control. Some of these options are listed right above the network diagram. You can ***hide isolated genes,*** which are genes that are not involved in any interactions. You can ***show coherent cascades***: interactions where the data are consistent with the experimental expression levels from your experiment. You can ***lock*** the ***zoom***, which prevents you from zooming in and out of the diagram while scrolling. You can also show a detailed legend by checking the ***"Show legend"*** option. The specific features of each mode are described in detail below.

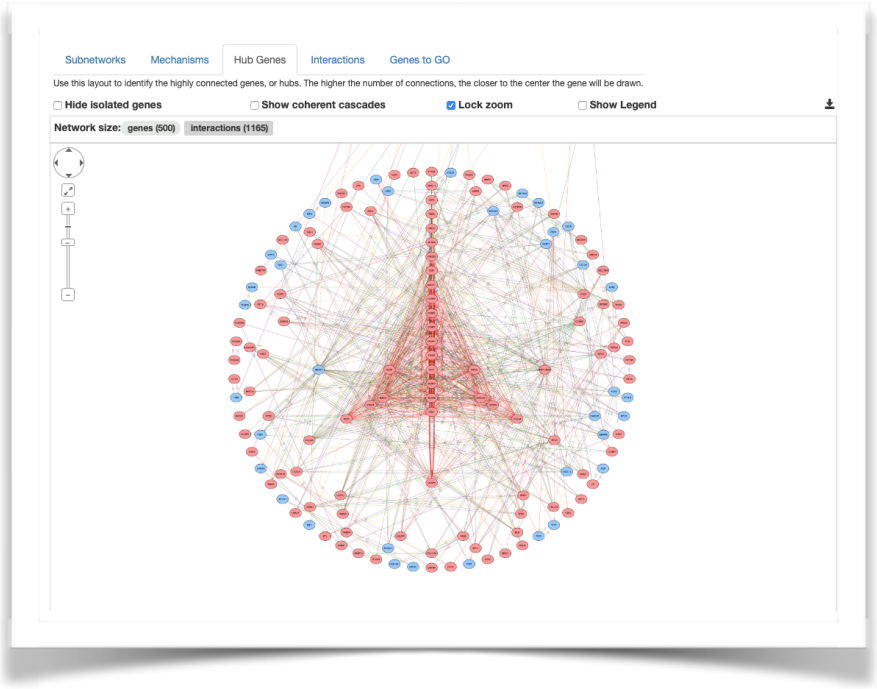
Subnetworks: This mode allows you to analyze the relationships between interconnected subsets of genes. Genes will be divided into groups that represent independent subnetworks. Here, you can see which genes are closely related.



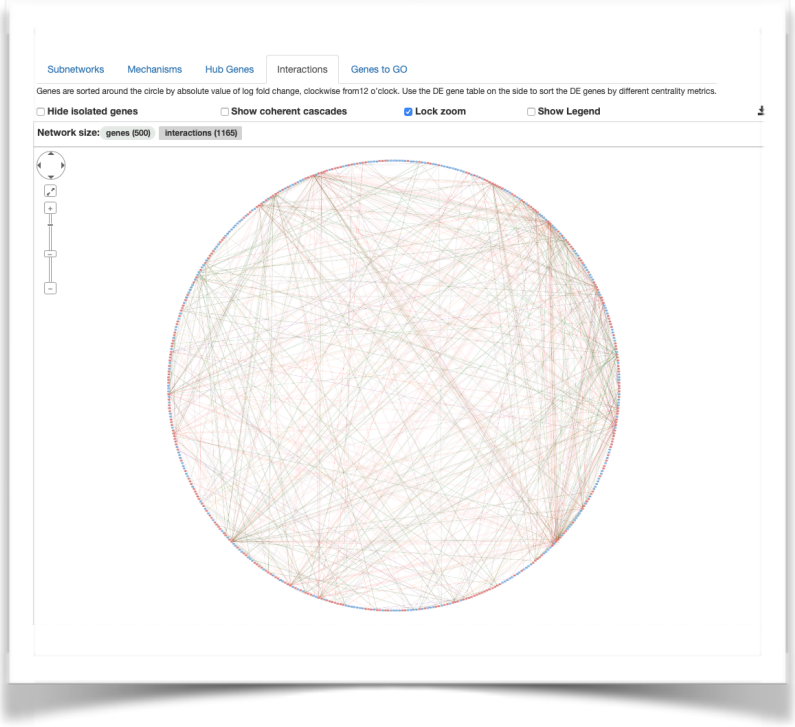
Mechanisms: This mode allows you to find putative mechanisms by displaying the interactions between small sets of genes with the fewest edges.



Hub Genes: This mode displays genes with the highest degree in the center, allowing you to visualize heavily connected networks in your set of differentially expressed genes. These "hubs" can be excellent places to formulate new hypotheses or get a better sense of the important nodes in your experiment.



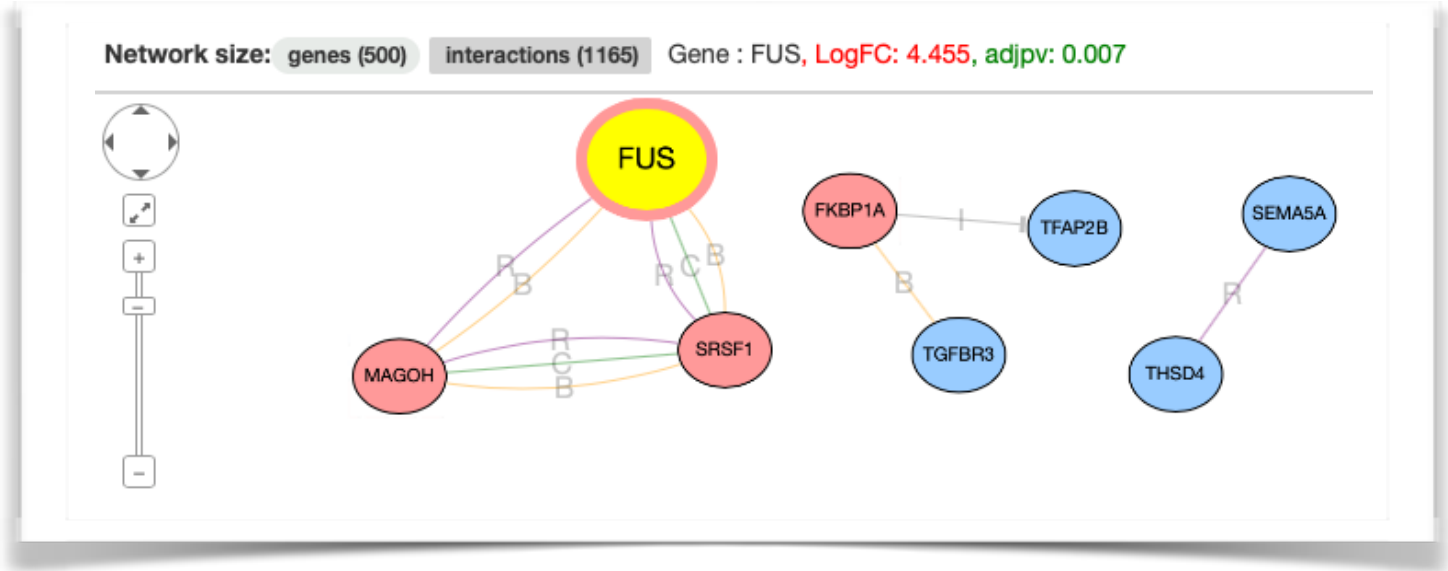
Interactions: This mode arranges genes based on significance and allows you to see how the number of interactions is distributed based on the differential expression of genes.



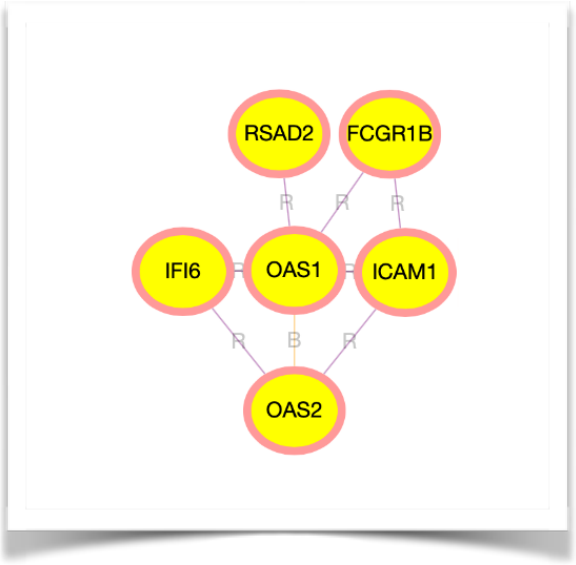
Genes to GO: This mode allows you to assess how your network of interest is related to a certain biological process, molecular function, or cellular component. Select a GO category then begin typing the GO term of interest. A drop-down list will populate. Select the GO term of interest and "Apply" the term to your network. This will show how the GO term is associated with the novel network displayed.

Navigating the Network Analysis

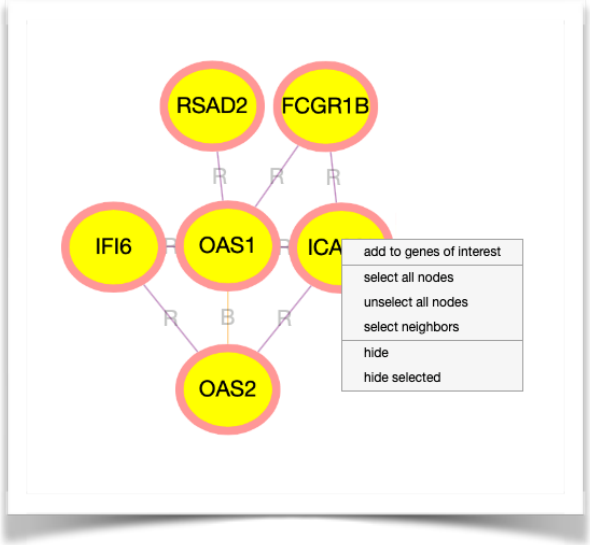
Clicking an empty space in the diagram and dragging the mouse will move the frame of the screen, so you can focus on the genes of interest. Clicking on a gene in the network displays its gene symbol, its LogFC and its adjusted p-value right above the diagram.



To select and drag a group of genes, hold down the shift key and drag over the desired group of genes. This will highlight the selected genes in yellow. Then, you can move the all the selected genes as one group.



Double clicking on any gene while it is highlighted will display a pop-up menu with several options. You can add these genes to your genes of interest, select all nodes, unselect all nodes, select all neighbours, hide one gene, or hide all the selected genes. Hiding will remove those genes from the network.



Navigation can be found on the upper left-hand side of the diagram. Moving the slider icon up will zoom in while moving the slider icon down will zoom out. Clicking the expansion button will return you to the default view of the network. You can also shift the screen in any direction by using the pan in the top-left corner.

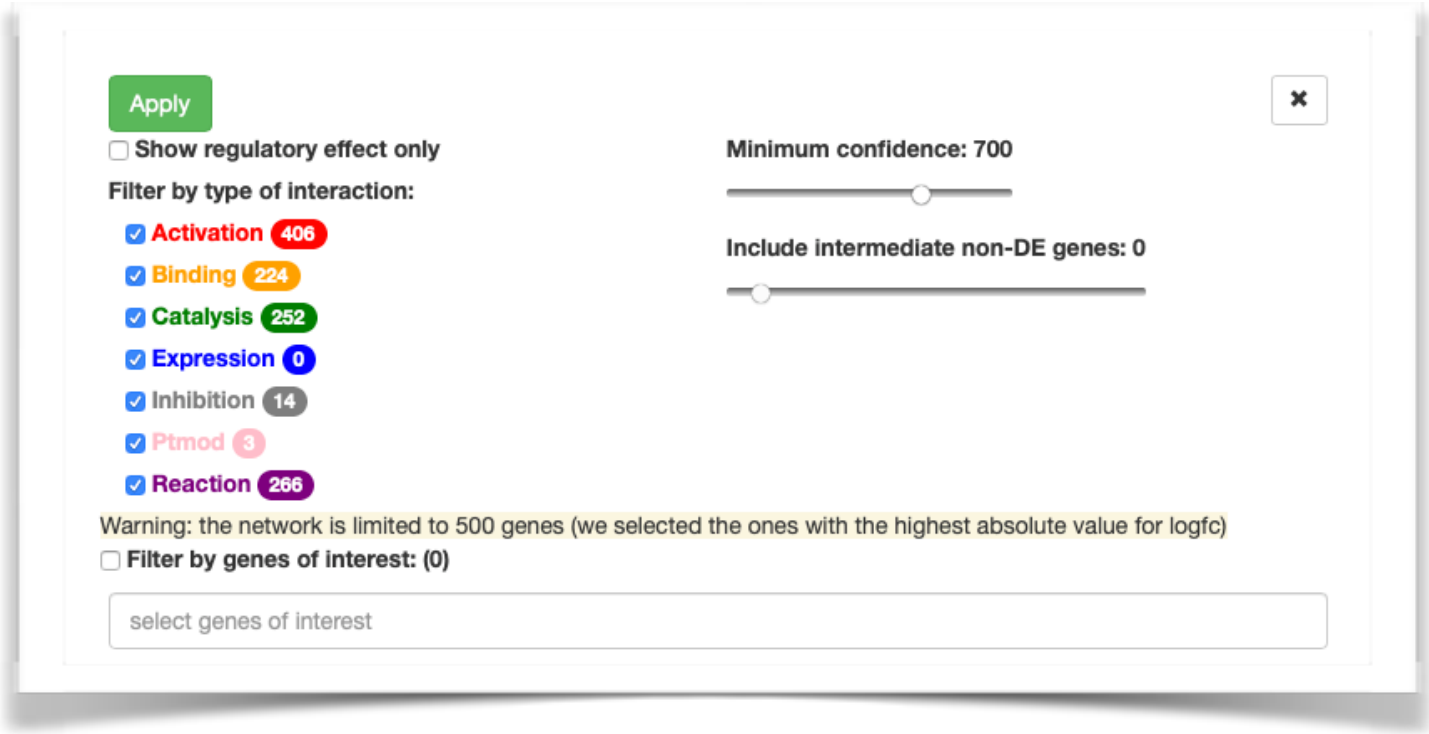
To the left of the network analysis diagram is the legend.

The legend details the different interactions that can be used to annotate the network diagram. The following interactions are available: activation, binding, catalysis, expression, inhibition, post-translational modification, and reaction. You can select or deselect any type of interaction to show in the diagram.

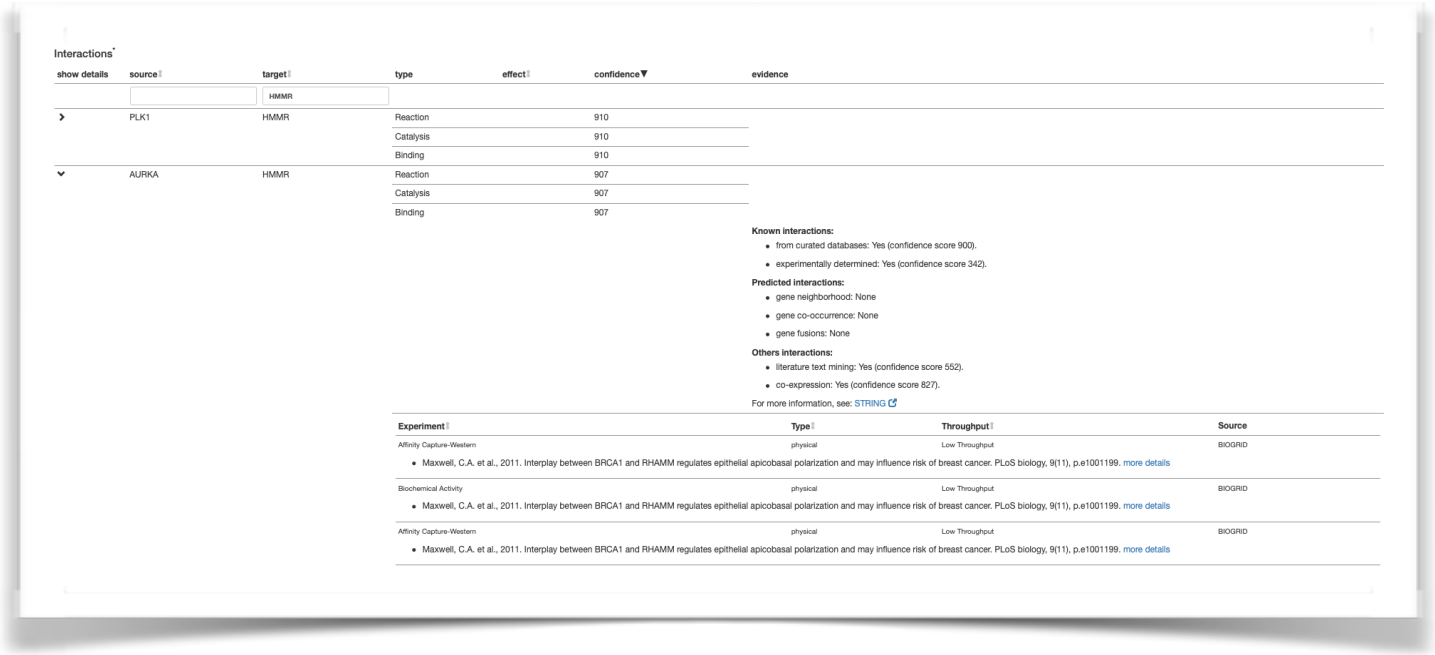
Likewise, you can choose to only show regulatory effects provided by the STRING database.

Beside the legend, you can choose to increase the minimum confidence level of the included interactions. Increasing the confidence level will eliminate interactions with less evidence supporting them. If you have a network of 50 genes or less, you can choose to include intermediate non-DE genes in the diagram as well. Filtering by genes of interest means that when you are filtering, by a GO term or another method, only the genes of interest will be considered.

Once you have made all of your selections, click “Apply” to update the network.



The table of interactions below the diagram provides more information about the evidence and experiments supporting specific interactions. You can search for interactions by entering a gene symbol into the search box under source and/or under target.



Interactions*

show details	source	target	type	effect	confidence▼	evidence
		HMMR				
›	PLK1	HMMR	Reaction		910	
			Catalysis		910	
			Binding		910	
⌄	AURKA	HMMR	Reaction		907	
			Catalysis		907	
			Binding		907	

Known interactions:
- from curated databases: Yes (confidence score 900).
- experimentally determined: Yes (confidence score 342).

Predicted interactions:
- gene neighborhood: None
- gene co-occurrence: None
- gene fusions: None

Others interactions:
- literature text mining: Yes (confidence score 552).
- co-expression: Yes (confidence score 827).

For more information, see: STRING

Experiment	Type	Throughput	Source
Affinity Capture-Western	physical	Low Throughput	BIOGRID
• Maxwell, C.A. et al., 2011. Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. PLoS biology, 9(11), p.e1001199. more details			
Biochemical Activity	physical	Low Throughput	BIOGRID
• Maxwell, C.A. et al., 2011. Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. PLoS biology, 9(11), p.e1001199. more details			
Affinity Capture-Western	physical	Low Throughput	BIOGRID
• Maxwell, C.A. et al., 2011. Interplay between BRCA1 and RHAMM regulates epithelial apicobasal polarization and may influence risk of breast cancer. PLoS biology, 9(11), p.e1001199. more details			

Meta-Analysis

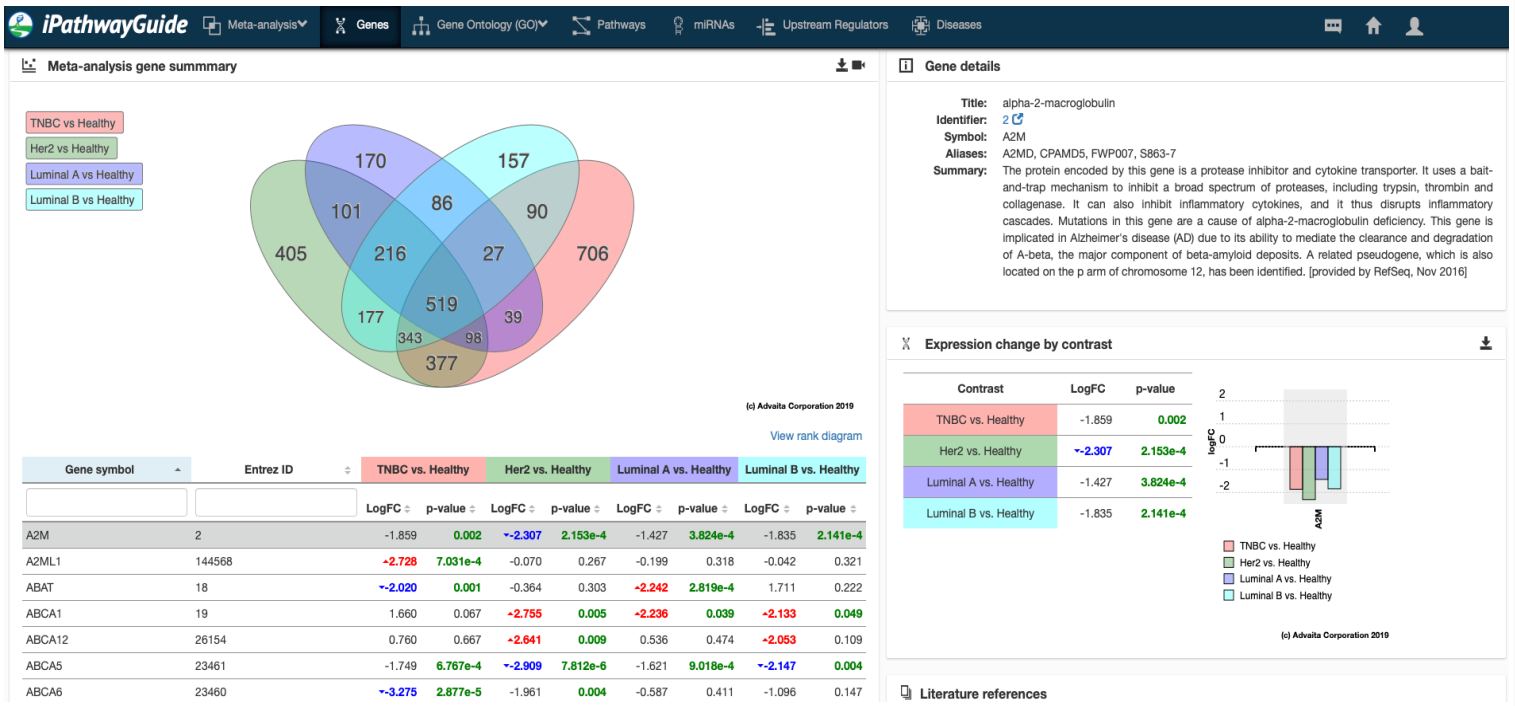
iPathwayGuide has the unique capability to compare conditions or contrasts within an set of related experiments, or to build a new meta-analysis from existing contrasts across several experiments. This display makes it easy to see responses that are the same and responses that are different across a set of related experiments. This module is also excellent for integrating across omics technologies (e.g. integrating experimental data from expression, proteomics, ChIP, epigenomics data sets).

Currently, iPathwayGuide supports analysis of up to 5 contrasts in a single meta-analysis. iPathwayGuide can also recognize input files with multiple contrasts, such as CuffDiff, Sciex SWATH data, nSolver data, and JMP Genomics, and generate meta-analyses from these files.

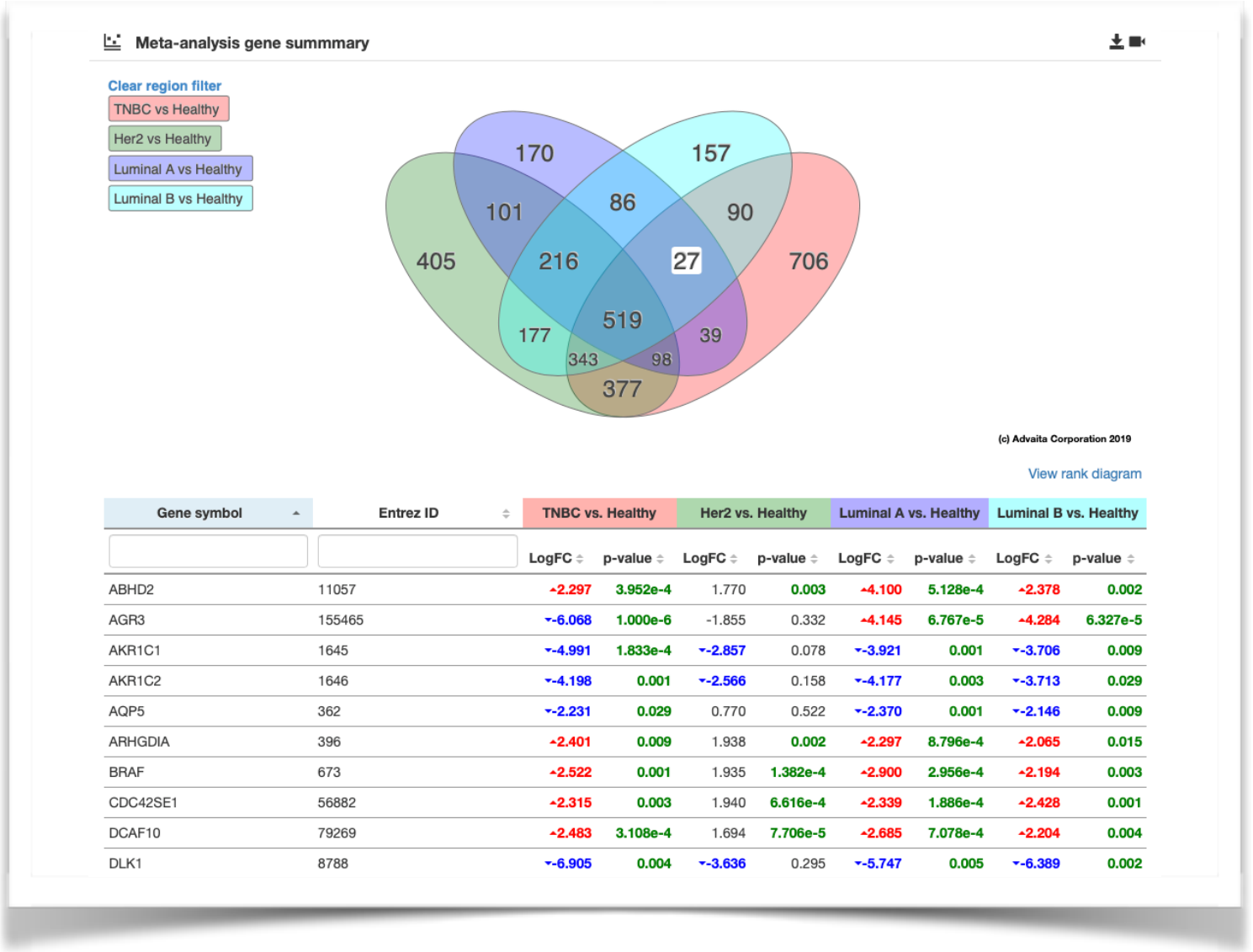
Viewing Meta Analysis from a Multi-contrast report

To compare contrasts in a set of contrasts, select Meta-Analysis from the drop-down menu.

The default view for the meta-analysis is genes. The Venn diagram shows the intersections of the contrasts with the number of entities that are in common in that region. Selecting the genes tab shows the number of genes for each region. Selecting other tabs changes the view to that data type (e.g. selecting the miRNAs tab will show predicted miRNAs).

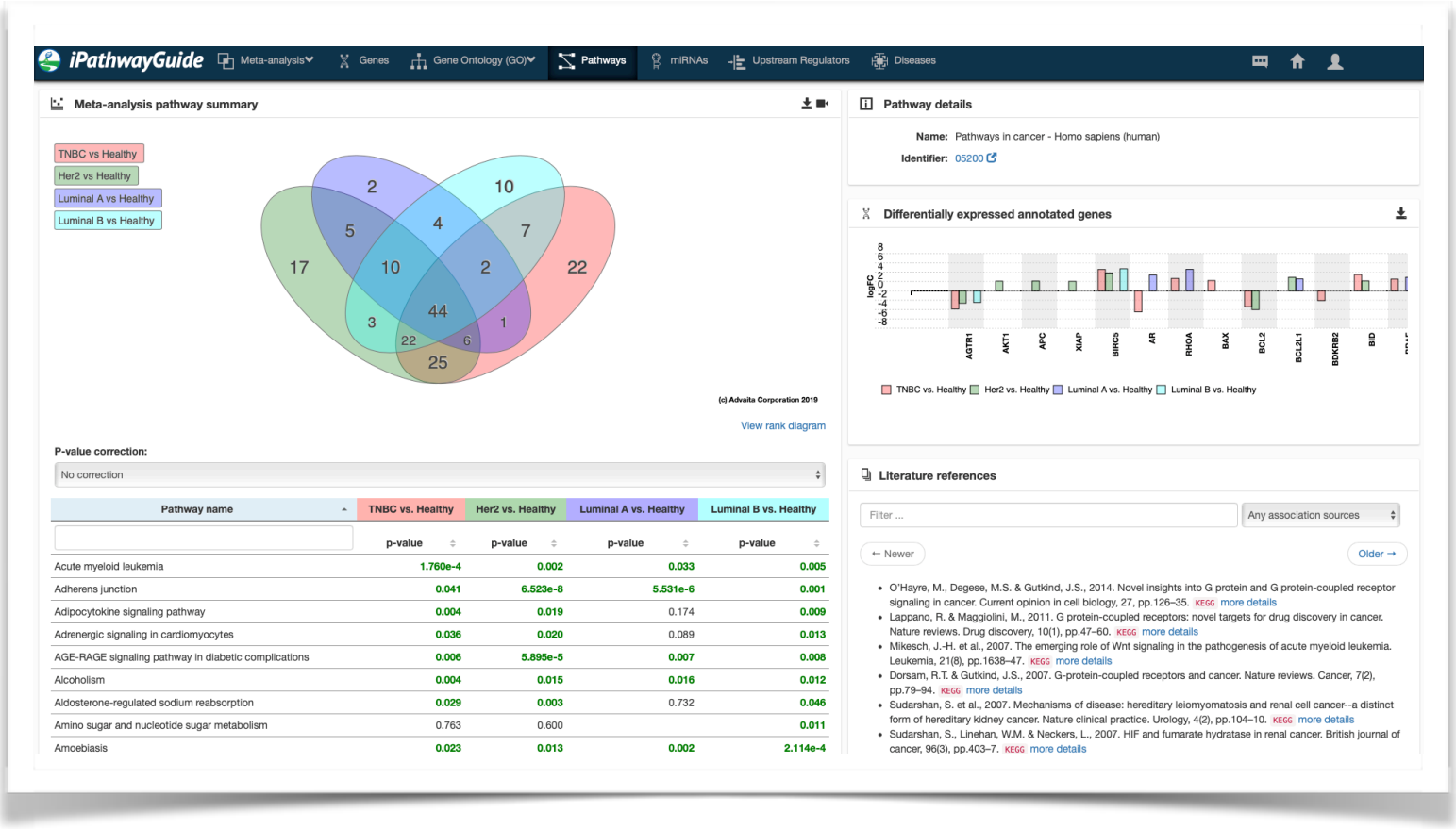


Selecting one or more regions will filter the list of results to the entities for that selection.



Gene symbol	Entrez ID	TNBC vs. Healthy		Her2 vs. Healthy		Luminal A vs. Healthy		Luminal B vs. Healthy	
		LogFC	p-value	LogFC	p-value	LogFC	p-value	LogFC	p-value
ABHD2	11057	2.297	3.952e-4	1.770	0.003	4.100	5.128e-4	2.378	0.002
AGR3	155465	-6.068	1.000e-6	-1.855	0.332	4.145	6.767e-5	4.284	6.327e-5
AKR1C1	1645	-4.991	1.833e-4	-2.857	0.078	-3.921	0.001	-3.706	0.009
AKR1C2	1646	-4.198	0.001	-2.566	0.158	-4.177	0.003	-3.713	0.029
AQP5	362	-2.231	0.029	0.770	0.522	-2.370	0.001	-2.146	0.009
ARHGDIA	396	2.401	0.009	1.938	0.002	2.297	8.796e-4	2.065	0.015
BRAF	673	2.522	0.001	1.935	1.382e-4	2.900	2.956e-4	2.194	0.003
CDC42SE1	56882	2.315	0.003	1.940	6.616e-4	2.339	1.886e-4	2.428	0.001
DCAF10	79269	2.483	3.108e-4	1.694	7.706e-5	2.685	7.078e-4	2.204	0.004
DLK1	8788	-6.905	0.004	-3.636	0.295	-5.747	0.005	-6.389	0.002

When analyzing entities with multiple annotated genes, such as pathways, on the right hand side, information about the selected entity is provided including the relative expression values for the annotated genes, color-coded to the respective contrast(s).

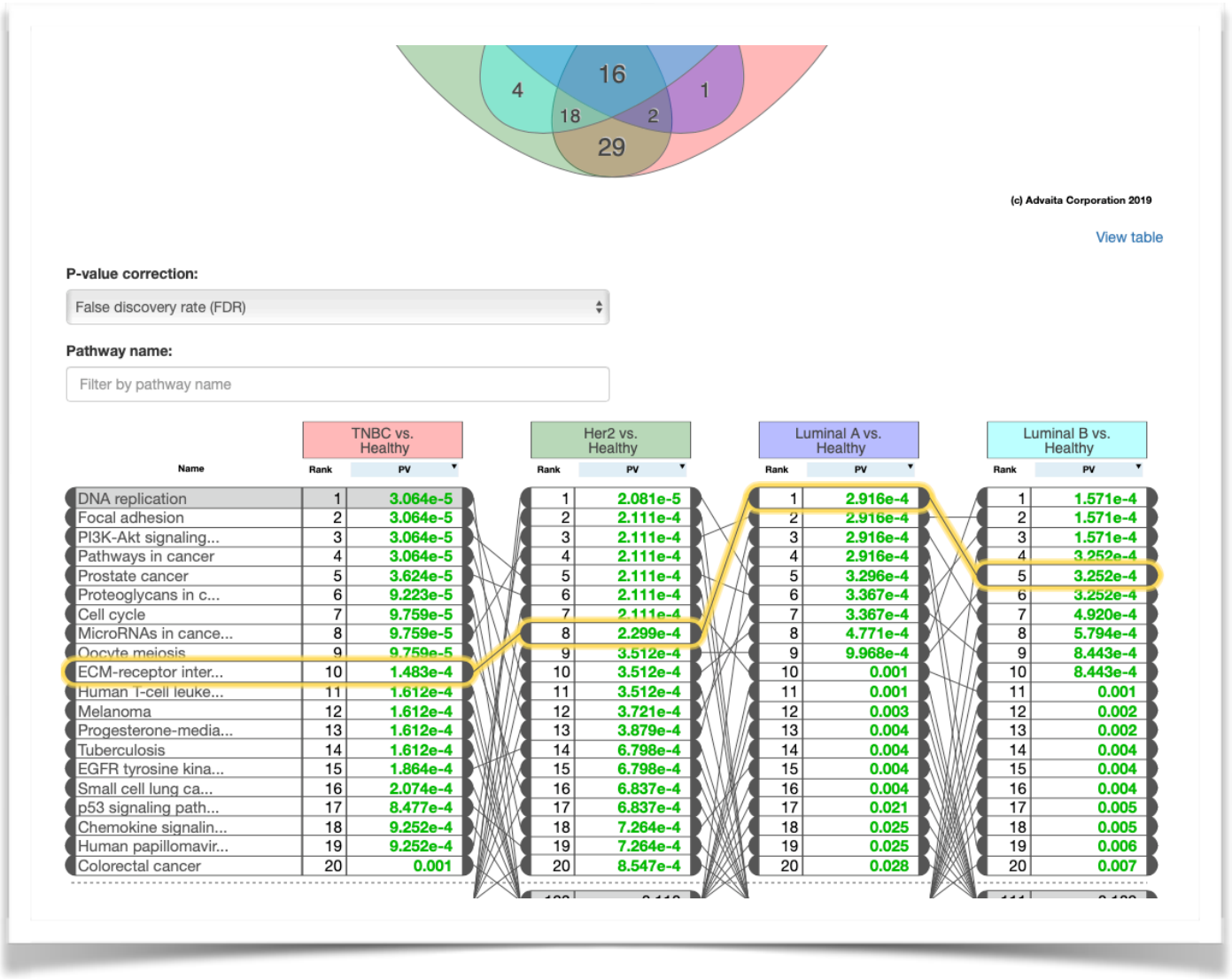


Pathway name	TNBC vs. Healthy	Her2 vs. Healthy	Luminal A vs. Healthy	Luminal B vs. Healthy
	p-value	p-value	p-value	p-value
Acute myeloid leukemia	1.760e-4	0.002	0.033	0.005
Adherens junction	0.041	6.523e-8	5.531e-6	0.001
Adipocytokine signaling pathway	0.004	0.019	0.174	0.009
Adrenergic signaling in cardiomyocytes	0.036	0.020	0.089	0.013
AGE-RAGE signaling pathway in diabetic complications	0.006	5.895e-5	0.007	0.008
Alcoholism	0.004	0.015	0.016	0.012
Aldosterone-regulated sodium reabsorption	0.029	0.003	0.732	0.046
Amino sugar and nucleotide sugar metabolism	0.763	0.600		0.011
Amoebiasis	0.023	0.013	0.002	2.114e-4

Rank View Layout

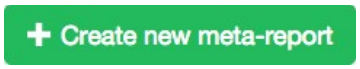
The Meta-analysis allows for the relative rank comparison between the contrasts for Genes, miRNAs, GO terms, Pathways, Upstream Regulators and Diseases. Select 'View rank diagram' in the lower right corner below the Venn diagram. The table will be replaced with a relative ranking of the results between the contrasts contained in the analysis.

Clicking the header listing the contrast will shift the relative rank focus to that contrast.

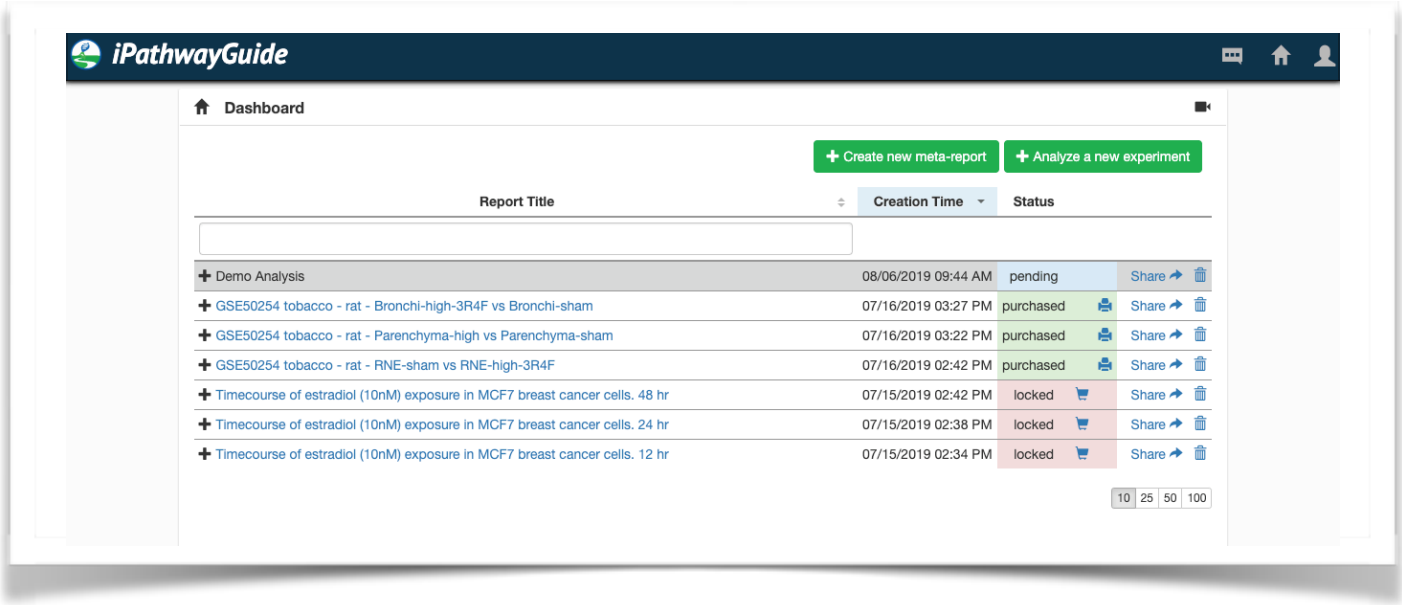


Constructing a New Meta-Analysis Report*

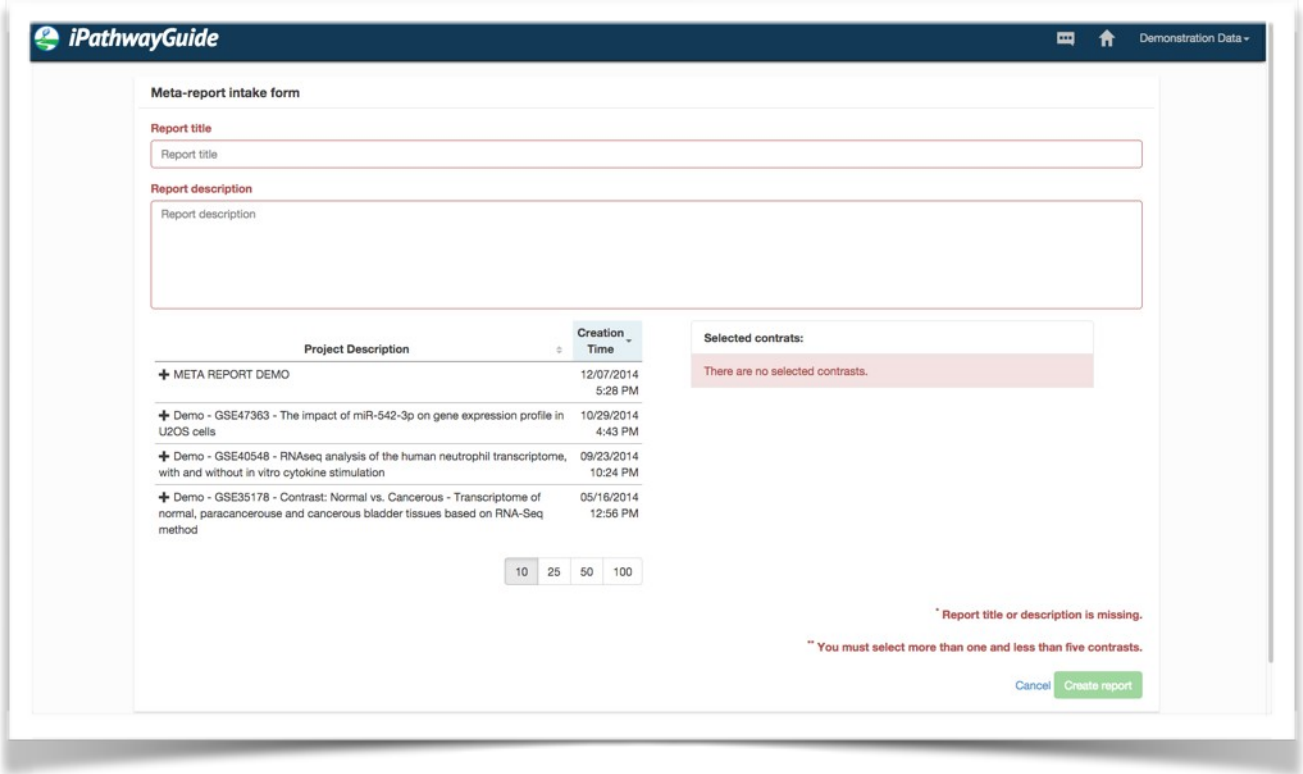
To generate a new meta-analysis based on existing reports:

To start creating a new meta-analysis click the 

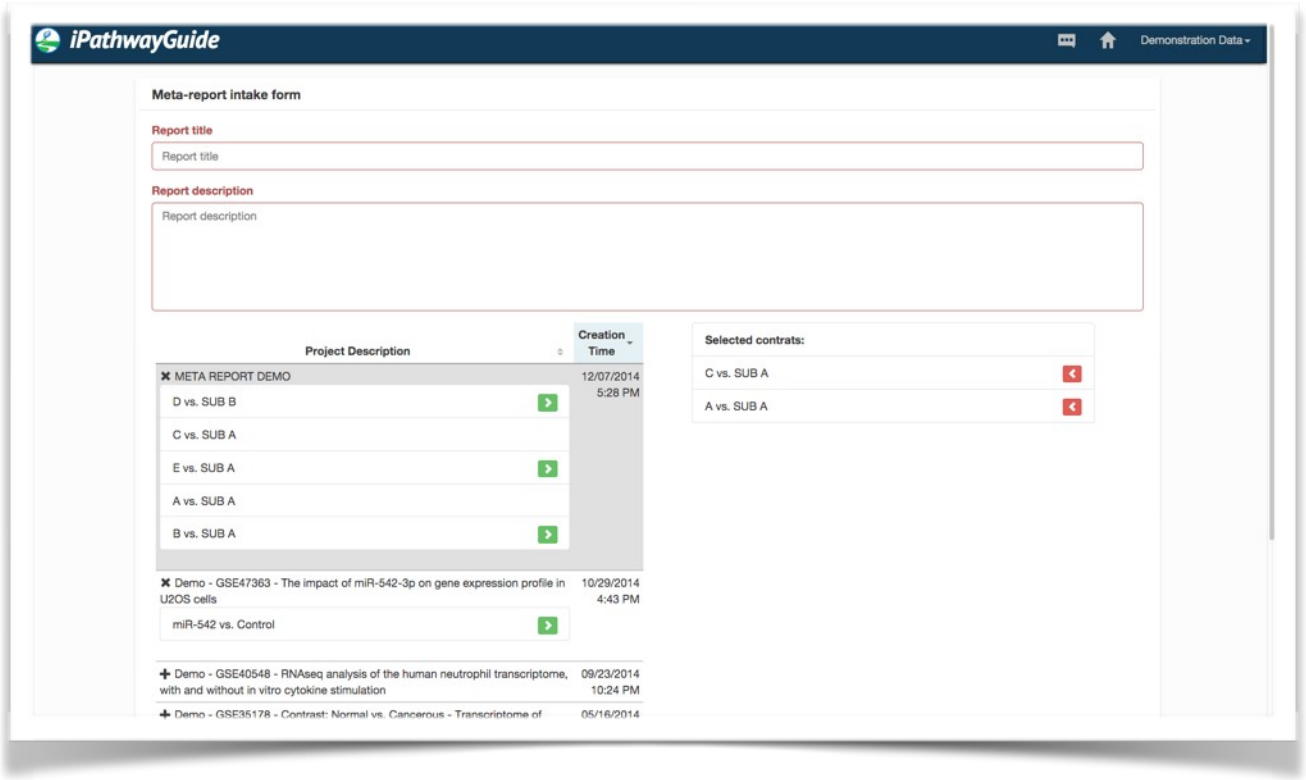
button on the dashboard.

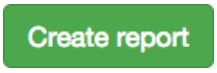


This button will take you to the Meta-report Intake form. Here, you may provide a title, description, and select which contrasts to include.



Only contrasts from purchased reports can be included for building a meta-report. Select up to 5 contrasts to include by expanding each report and selecting the desired contrast.



Once you have made your selections, you may create the report. 

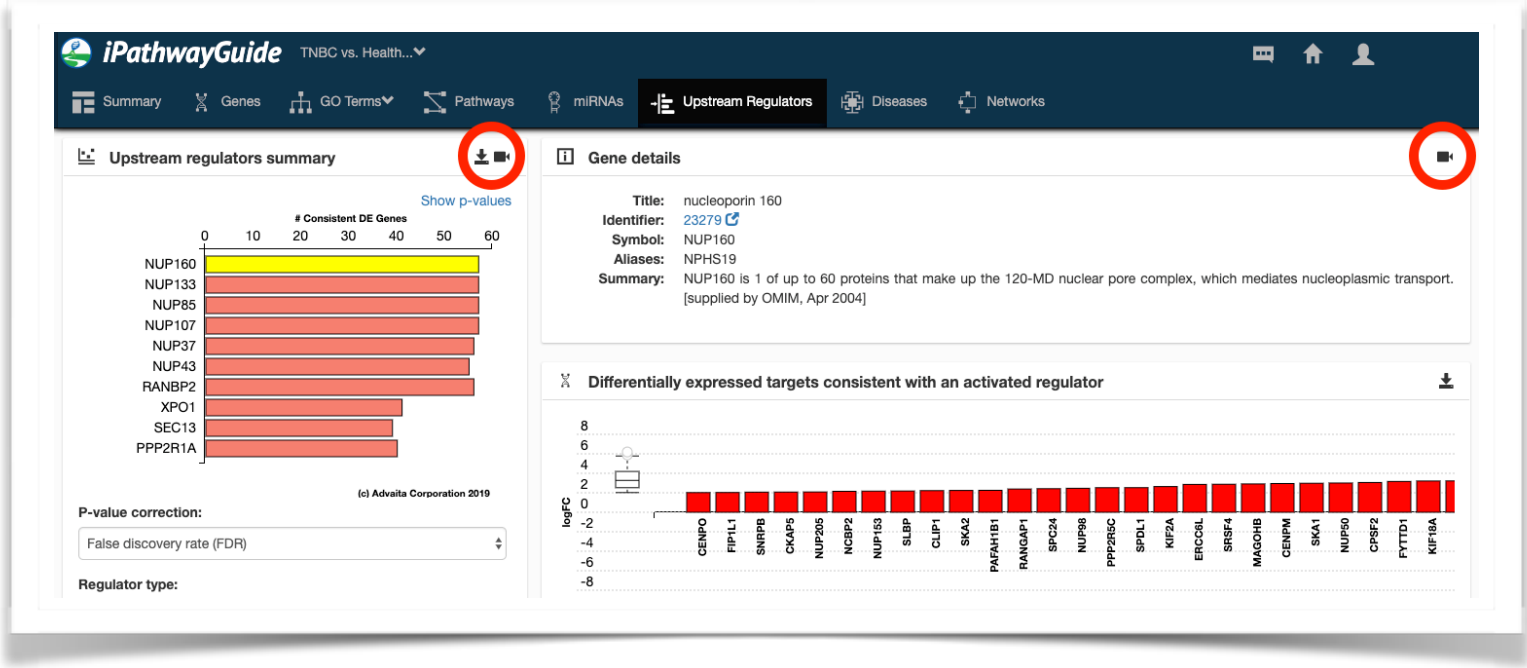
Meta-reports appear in your dashboard along with single-contrast reports.

Exporting/Downloading

Images and tables can be exported or downloaded by clicking the download button.

Help and Video Tutorials

In each section, there are a number of resources available to answer questions you may have. Simply click on the “camera” icon in the section you are interested in to get a pop up window with a linked video or additional information.



Citing iPathwayGuide and Help

Using Advaita Bio's products or content for any form of publication (e.g. print, electronic, etc.) requires citation. Please use one of the options below for citations:

"The Data (significantly impacted pathways, biological processes, molecular interactions, miRNAs, SNPs, etc.) were analyzed using Advaita Bio's iPathwayGuide (https://www.advaitabio.com/ipathwayguide). This software analysis tool implements the 'Impact Analysis' approach that takes into consideration the direction and type of all signals on a pathway, the position, role and type of every gene, etc., as described in (Draghici, 2007, Ahsan 2018)." [1]

[1] Identifying significantly impacted pathways and putative mechanisms with iPathwayGuide
S Ahsan, S Drăghici, Current protocols in bioinformatics 57 (1), 7.15. 1-7.15. 30
A systems biology approach for pathway level analysis
S Draghici, P Khatri, AL Tarca, K Amin, A Done, C Voichita, C Georgescu, ...
Genome research 17 (10), 1537-1545

For additional guidance, please contact info@AdvaitaBio.com or visit the FAQs page of our website at www.AdvaitaBio.com