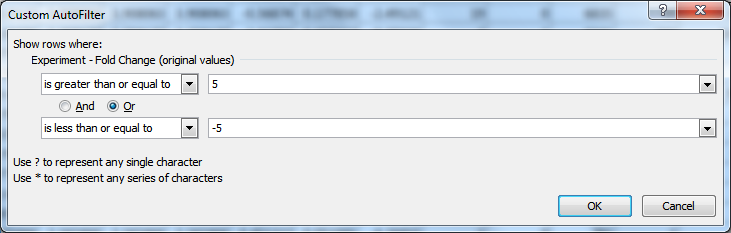
**Onco-Wiki Pathway Analysis**

*This guide uses the Ingenuity Pathway Analysis web-based application which requires a paid license.*

1. **Data Filtration**

Data filtration is required as only <8000 genes can be process through pathway analysis software from our original list of 57,773 spliced variants. This setting is mandated by this software in the interest of time of analysis.

Using a spreadsheet software package (Excel), open your Pair-wise experiment files or your Manually Manipulated patient file and use the ‘Filter’ function (Under Data menu in Excel). Filter using the ‘Experiment – Fold Change (original values)’ column and select ‘Number Filters’ and ‘Custom Filter.’



Filter using data that ‘is greater than or equal to’ OR ‘is less than or equal to,’ with a numerical value between 1 and 10 with its corresponding negative value as above. Export this reduced list (Alt+; ) and save this filtered data in a new spreadsheet.

Since using data filtration cut-offs is quite a crude measure, it is recommended to use a consistent fold change for all your sample data.

It may be worth filtering in reverse where you remove the most positive and most negative fold changes off your gene list but ensure that the genes want to manipulate are still in your reduced list. Then, manually manipulate individual genes and ensure that this list is <8000 genes. In this situation, it is recommended that you manipulate as many relevant genes as possible to ensure that genes involved in tumour pathways are not been excluded from analysis.

1. **Ingenuity Pathway Analysis**

Ingenuity Pathway Analysis (IPA) uses imported differences in gene expression (fold changes) among other quantifiable normalised expression values from datasets to generate hypotheses of how a particular gene interacts with a particular pathway or molecular mechanism. The significance of this interaction depends on the regulation direction and size of the fold change. The knowledge database is entirely based on our existing knowledge and understanding of biological and chemical systems and uses this as a basis of predicting the canonical pathways mapped genes are involved in.

Based on the significance of these predictions, IPA suggests potential drug treatments, classified as upstream regulators. These upstream regulators have putative significance based on defined relationships between transcriptional regulators and target genes. Since genes are often modulated by several upstream regulators, it is not known which will dominate but IPA uses statistical algorithms to make reliable predictions of an activation state.

1. **Data Import**

Login and launch Ingenuity Pathway Analysis ensuring you have Java installed on your computer. Select ‘New’ and ‘New Core Analysis.’ Create a folder for your data and upload your filtered dataset. Allocate ‘ID’ to Column 1 and ‘Observation 1’ to ‘Experiment – Fold Change (original values).’ If only the ‘ID’ column appears, open your filtered dataset file using a spreadsheet or text editing package and remove all columns except for the fold changes and feature ID column and re-import into IPA. Ensure Column Headers are selected and select ‘GenBank’ and ‘Gene symbol – human’ identifier types.

Finally, Analyse/Filter Dataset using Core Analysis. If you have not already done so, set a whole number fold-change cut-off for both upregulated/downregulated genes and recalculate given that Ingenuity Pathway Analysis allows only less than 8000 analysis-ready molecules to be mapped. Take note of the number of ‘Mapped IDs, Analysis Ready IDs and All IDs’ and Run Analysis.

1. **Data Interpretation**

In the Summary tab, you can view the most/least expressed pathways by p-value, regulators, genes, disease associations and toxicity effects. It is worth exporting this as a PDF.

In the Upstream Analysis tab, you can view suggested Upstream Regulators and more precisely drugs and chemicals that are functionally significantly with an interaction pathway (mechanistic network). In the Diseases & Functions tab, there are a series of heatmap analyses which gene functionality in diseases and broad functions of diseases. Relative upregulation is orange and downregulation is blue. You can identify genes and associated genes in relevant areas of interest by clicking on the maps itself. In the Tox functions subtab, you can view toxicity effects induced by upregulation of certain genes and may be of interest to do a ‘Tox analysis.’

In the Regulator Effects tab, IPA may suggest networks through an intuitive algorithm linking activation or inactivation of regulators and an increase or decrease of function. This consists of a ‘consistency score’ which improves on identification of more differentially expressed genes through higher fold changes.

1. **P-value and Z-score**

An overlap p-value is used to identify whether there is a statistically significant overlap between dataset genes and genes regulated by a potential regulator. As a measure of statistical significance, a p-value of 0.05 indicates a false discovery rate of 5% in the data so a lower p-value = more reliable estimation of significance.

Z-score calculations are used to infer the activation state of whether that regulator has significantly more ‘activated’ or ‘inhibited’ predictions based off genes it interacts with.

|  |  |  |
| --- | --- | --- |
| Observation (Fold Change) | Regulation Direction (literature) | Predicted State (z-score) |
| Up | Activating | Activating |
| Down | Activating | Inhibiting |
| Up | Inhibiting | Inhibiting |
| Down | Inhibiting | Activating |

1. **Improving Visualisation of Drug Targets**

Manipulating your desired genes further may accentuate its characterisation and repeat analyses through selecting different fold change cut-offs may help to identify what pathways and networks your desired genes are implicated in.

It is worth noting down the most statistically significant drugs in your analyses as well as the genes which have been found in pathways of interest (not only genes you manipulated) and do a final analysis on these desired genes that you know will be in relevant implicated pathways and disease networks.

In addition, it is possible to do a comparative analysis on all your initial analyses where the most consistently expressed genes / recommended drugs are portrayed. There are also toxicology, metabolomics and biomarker filter analyses to explore.