This is a demonstration of HitWalker2 using data from the Cancer Cell Line Encyclopedia.

The main page allows the user to choose a subject of interest, in this case a cell line. After typing in the desired subject name or alias in this case ‘Hep G2’ The relationships between the subject and any experimental samples derived from them. As multiple samples can be derived from a given subject for a given datatype, at this point the choice of samples to use for a given data type can be adjusted.

We can also adjust various parameters of the prioritization algorithm as well as the thresholds for determining whether a particular datatype is a hit for a given subject/subjects.

These parameters can be changed, saved and reset as well.

Assuming there are appropriate datatypes available, a prioritization procedure can be carried out.

Alternatively the subject can be examined further in terms of its relationships to other subjects and genes.

First lets look at the prioritization path.

For our Hep G2 sample there is both variant data as well as drug data which are the two datatypes necessary for prioritization as it is currently configured.

Clicking the ‘Prioritize’ button will bring back an HTML table of the prioritization results. This table can be exported as a CSV file and imported into Excel. This excel file contains the parameters used by the algorithm as well as the current sample assignments.

Clicking ‘Display’ brings up a sorted list of variants and Drug gene Hits. You can choose up to 5 of each. A subnetwork is generated using the chosen genes and set as the first panel in the ‘panel’ view.

There are several things to mention about the panel view. On the very left hand side, the legend depicts the relationships within a given node as well as between nodes.

Within node relationships such as variants or Hits seen in a gene for a given sample are represented as smaller nodes within the gene or sample nodes. Here, the solid purplish nodes represent that a variant exists in a given gene, while the tannish nodes represent that the node is seen to have a gene score hit which in this context means that the particular gene is seen to be important with respect to the drug data.

Mousing over each panel gives a rough idea of the lineage of the panel.

Interaction with the panels is mainly accomplished through mouse events. The nodes can be repositioned in a panel singly or in a group.

Right-clicking on a panel allows the choice of several options. A new node can be added to the nodes already in a panel. This new node can be a gene, subject or pathway. When adding in a node, a new panel is created and all of the nodes are grouped by their relationship types. We can see this by Adding in JAK2. Notice that the special grouping node, termed a MetaNode is formed to represent distinct groups of genes in this case delineated by the types of relationships with the cell line.

The data contained within these MetaNodes (as well as the regular gene or subject nodes) can be examined by right clicking on them. In the case of a MetaNode, a new panel can be created containing only the relevant subset by clicking on the count label. These summaries can also be exported to a CSV document.

Panels can be deleted (deleting the 3rd panel)

Many times, it is examine the relationships between the subjects and the genes in a particular pathway. There are several mechanisms to do that via the panel right click.

First, by clicking the ‘Pathway context’ button 5 of the genes in the panel are searched agains HitWalker2’s pathway database. In this case no pathway has all 5 genes but genes can be added or removed until an interesting pathway is found.

Clicking OK directs the user to a new window where the genes from the pathway are displayed in a network context.

Additionally pathways can be grouped by relationship type (Hit/no Hit expression, expression) when specified in the ‘Add Node’ -> ‘Pathway’

Each panel can be exported to a PDF or to a CSV where the node/edge information is described in a modified adjacency matrix format, commonly used for graph data.

The entire panel view can also be exported as a PDF by right clicking on the image outside of a panel

Left-clicking, with/without the shift button provides a way to select one or more nodes. When a node is right clicked after being selected it can be used in a predefined query. A default set of these queries are automatically generated, however new ones can be added to the configuration file relatively easily.

By carrying out the ‘GeneScore’ query we get back a MetaNode containing all the subjects who have a ‘GeneScore’ hit for MAP2K7.

We can examine this subset of cell lines in more detail by either adding in gene or by dragging existing genes from one panel to the next. Or similarly by selecting both MAP3K13 and IKBKB.

Similar queries can be asked of the subjects. We know the Answer for genescore, and there are no genes with variants or expression hits for all 93.

We can however subset the metanode by say whether they were derived from the liver and ask similar questions. Such as whether there are genes that are mutated in both cell lines. MetaNodes can be ungrouped to retrieve info on the genes within. These results can be exported or used for other queries such as overlaying these data on a pathway containing some of these genes.