

DIFFERENTIAL EXPRESSION SEARCHES

Use this search to find genes that are differentially expressed between the samples of a microarray or RNA-Seq experiment. This search offers incredible flexibility in parameter choices as you set up your search, but your parameter choices drastically affect the subset of genes returned by the search. This document explains some parameter choices and how they affect your results.

Consider a microarray or RNA-Seq experiment that measures gene expression at 0, 1, 6, 12, 18 and 24 hours post-infection. Each time point, or sample, contains expression values for thousands of genes. This search looks for genes that are differentially expressed between time points and determines differential expression based on fold-change – which is calculated as a pairwise comparison between gene expression values in a reference and a comparison sample. When you set up the search, you define the differential expression cutoffs, sample groupings and the operations (if any) applied to sample groupings. For each gene, the search calculates differential expression as the ratio of expression in the samples that you want to compare and return genes that meet your parameter choices.

The Basic Search: 1 reference vs. 1 comparison sample

You choose samples, fold change cutoff, and direction of change.

A one-to-one comparison between two samples is the most straightforward comparison that can be made with this search. Suppose you want to find genes that are up-regulated 2-fold or greater between the zero reference and 12 hours post-infection. For each gene, the search calculates fold change as a ratio of two numbers: expression at 12 hours post-infection vs. expression at time zero. The search returns every gene whose expression at 12 hours post-infection is greater than at time zero by at least 2-fold.

Complex search: 1 reference vs. multiple comparison samples

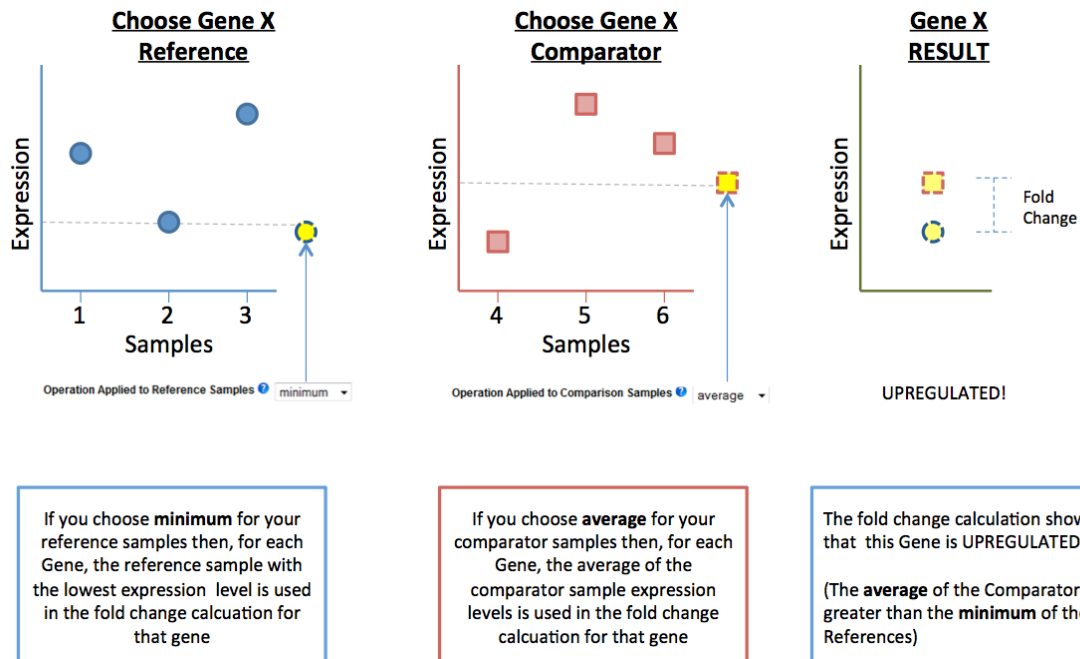
You also choose operation applied to multiple comparison samples.

Suppose you want to know what genes were up-regulated by at least 2-fold in both the 12 and 18 hour time points. In this case there are three expression values for each gene: the zero reference, and two comparisons – 12 and 18 hours. Before calculating fold change, it is necessary to reduce the expression values from the comparison samples to a single number. Single numbers associated with the comparison samples are the maximum, minimum or average value for the 12 and 18 hour time points. When setting up the search, you will be asked to define an additional parameter, the operation applied to comparison. If you choose to apply the minimum operator, the search will return every gene whose lowest expression value among the 12 and 18 hour time points is greater than at time zero by at least 2-fold.

An even more complex search: multiple reference vs. multiple comparison samples

You also choose operation applied to multiple reference and multiple comparison samples.

Suppose you want to compare the early time points (0, 1 and 6 hours post-infection) to the late time points (12, 18 and 24 hours post-infection). In this case, there are six expression values: three from reference samples 0, 1 and 6 hours; and two from comparison samples 12, 18 and 24 hours. Each group of expression values needs to be reduced to a single number using either the maximum, minimum, or average value across the samples in the group. When setting up the search, you will be asked to define the operation applied to the reference and comparison samples. Applying the minimum operator to the reference and the average operator to the comparison will return every gene whose lowest expression value among the early time points is at least 2-fold lower than the average expression in the late time points.



Parameter choices affect the number of genes returned by the search.

The operation parameters define a window of expression values in which to look for genes meeting the fold change cut off. The broadest window for finding up-regulated genes is created using the minimum reference and the maximum comparison value. Choosing to calculate fold change using the maximum reference and the minimum comparison creates the most restrictive window of expression values. Calculating fold change in a broad expression window will return more genes than a restrictive window. (working on a graph for this section)

Global min / max in selected time points

The Global min /max is the minimum/maximum expression value across the entire dataset, not just across your chosen samples. Some of our differential expression searches include this parameter, allowing you to search only those windows that contain the global minimum/maximum values.

