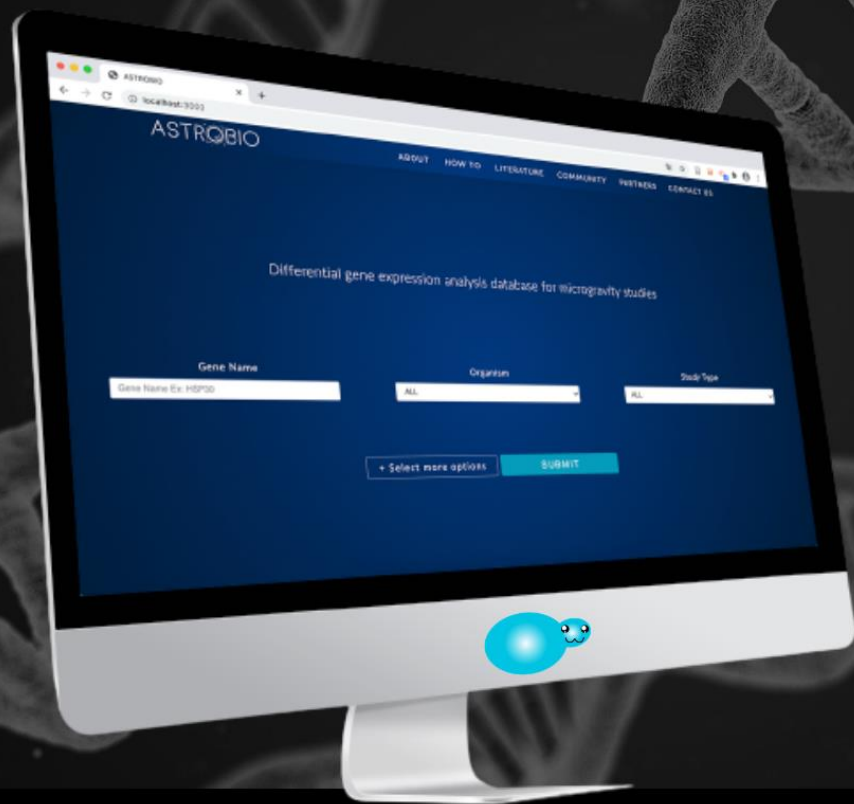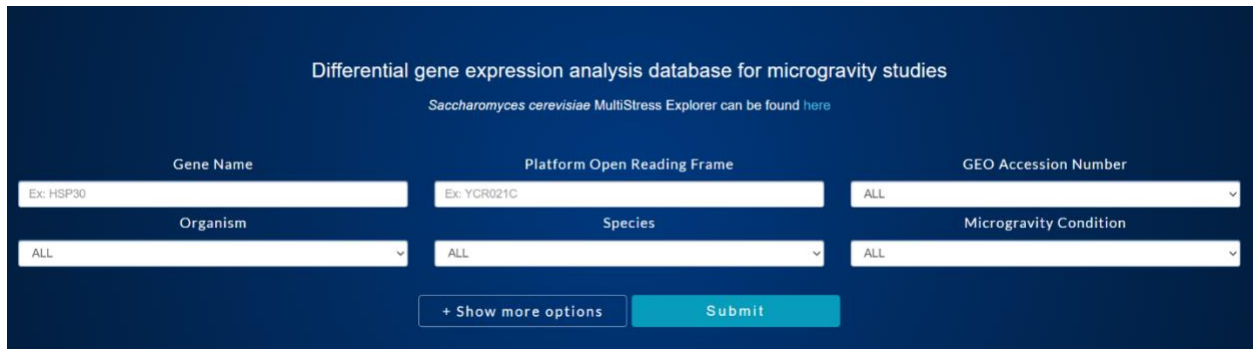# ASTROBIO
USER GUIDE

iGEM CONCORDIA

This is a User Guide for AstroBio, an open-source, differential gene expression analysis software and database for microgravity researchers.
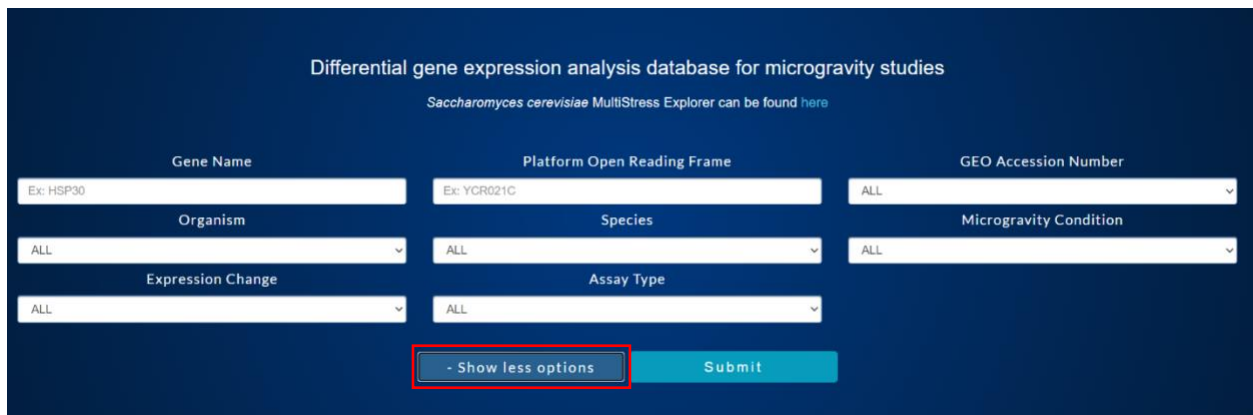
Step 1. Enter your desired search options:

Figure 1: AstroBio Search Parameters

- *Gene Name*: HSP30
- *Platform Open Reading Frame*: YCR021C
- *GEO Accession Number*: GSE4136
- *Organism*: Yeast
- *Species*: *Saccharomyces cerevisiae*
- *Microgravity Condition*: HARV

Step 2. Click on **+ Show more options** if you want to refine tour search paramaters. You can also opt to just **- Show less options**.

Figure 2: AstroBio Search Parameters - Options

- *Expression Change*: Up Regulation
- *Assay Type*: Microarray

Step 3. Once you are satisfied with your search criteria, click on **Submit**.

The following **Search Results** should be available to you – in this example, we looked for the HSP30 Gene using the Gene Name parameter

Figure 3: AstroBio Search Results

Step 4. If your research results multiple results, you can choose how many studies are displayed by clicking on **Results per page**. You can also sort the results by clicking on **Sort by**. This will allow you to sort sutdies by *Adjusted P Value*, *P Value*, *Descending Expression Change*, as well as *Ascending Expression Change*.
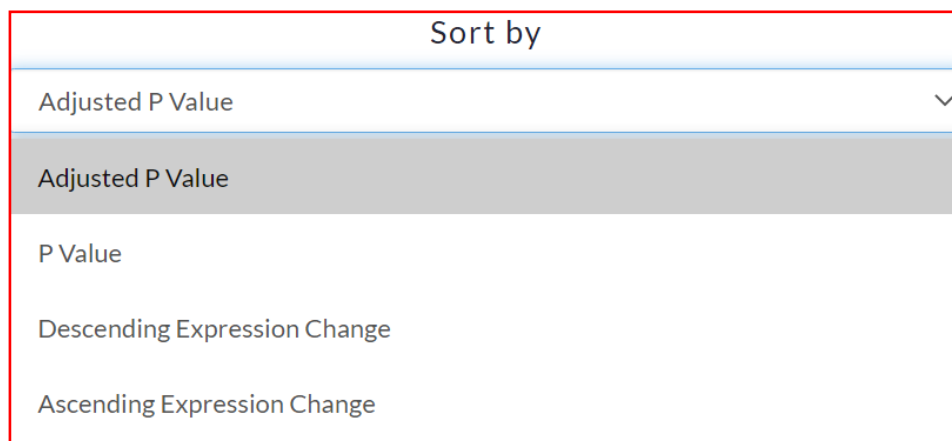

Figure 4: Sort by Options

Step 5. To obtain more information on a particular study, simply click on **More Info**.

Figure 5: More Info Option

The following window will appear:

Figure 6: Overview Tab

*Note: Click on **AstroBio MultiStress Explorer** for Saccharomyces cerevisiae*

AstroBio MultiStress Explorer

About   Graphs

Choose a Metastudy
HeatShock

Gene Forest Plot
HSP30

Choose threshold values to download genes and/or display on a pca biplot

Summary Log2-Fold Change   **Number of Top Ranking Genes**   **Sign Consistency**

Enter common gene names from S.cerevisiae genome. Multiple entries can be entered using ',' as a delimiter with no spaces between and in all caps. Ex : ADR1,UPC2

Plot Specific Gene Values:

Show IDs on PCA-Biplot   Reset PCA-Biplot

Number of Queried Genes:

⬇ Download PCA-Biplot Data

Summary

AstroBio MultiStress Explorer is a small webapp intended for exploring results of a series of pairwise comparisons of *Saccharomyces cerevisiae* transcriptomics analyses in simulated microgravity (HARV bioreactor) with various other environmental stressors. Prior to the meta-analysis, differential expression analysis was performed using linear mixed models provided the by the R package limma for each condition. The adjusted p-values for each experiment were combined usng a random effects model in addition to summarizing the fold change across conditions. The environmental stresses include:

- Oxidative Stress from 0.19mM Cumene hydroperoxide. RNA was extracted after 6 minutes of treatment. (GSE26169, Sha et al. 2013)
- Heat Shock at 37C for 7 minutes. (GSE132186, Mühlhofer et al. 2019)
- Hyperosmotic Shock from 0.4M of NaCl. RNA was extracted after 6 minutes of treatment.(GSE13097, Romero-Santacreu et al. 2009)

All of the above datasets were then combined both pairwise and pooled with data derived from the condition:

- Simulated low shear modeled microgravity (LSMMG) from growth in a High Aspect Ratio Vessel (HARV). RNA extracted after 5 generations of growth.(GSE4136, Sheehan et al.)

Guide

The MultiStress explorer provides three different plots to review the results of the meta-analysis. The first is an interactive **Volcano Plot** which displays the combined p-values versus the summarized log-fold change in expression for each gene. Colour intensity is derived from sign consistency across conditions, a sign consistency of 2 for instance would indicate that the gene is upregulated across two conditions. Users can hover over the datapoints to view information concerning the gene or click on them to change the forest plot display below.

The **Gene Forest Plot** shows the expression of particular gene across conditions when a name is typed in the window or when clicked on the volcano plot. The input expects a common gene name in most cases or an ORF identifier when none are available. All genes displayed in the volcano plot are valid entries for this view.

The **PCA-Biplot** is the results of a principal component analysis of the log-fold difference in expression prior to meta-analysis. Eigenvectors are plotted with their direction indicative of their similarity to other groups and their length corresponding to the magnitude of variance the group contributes. Individual genes can be plotted and labeled by using the three query variables which select genes on the basis of their inclusion in the random effects model. These query variables are:

- **Log2 Summary Fold Change**: the threshold for summary fold change values as shown in the volcano plot. The genes are queried using an *at least* heuristic of selection. This means that an input of 1 will select all genes at one or above (>= 1) while an input of -1 will select all genes at -1 or *below* (<= -1).
- **Number of Top Ranking Genes**: Genes are ranked from smallest to largest based on their summary p-values. A input of 10 for example would select the ten top-most genes in the volcano plot.
- **Sign Consistency**: The direction of expression across stress conditions are combined into a single metric by using the sign of the fold change value. A gene which is upregulate in one study contributes +1 to the sum and vice versa. The formal definition is:

$$\sum_{i=1}^{n} \frac{logFC_i}{|logFC_i|}$$

where *n* is the number of studies and *logFC* is the log2 fold change for a particular gene.

- **Plot Specific Gene Values**: A gene name or a series of gene names can be entered to be shown on the biplot. If multiple entries are desired then the user must use a comma as separator with no spaces included. Gene names are expected in all-caps and are not present if one or more name is not found in a study.

The **Heatmap** is also generated from the result of the query input variables showing the intensity of log2 Fold-Change from each dataset. Genes are only drawn if the gene is found in every study.

The data which generates the PCA-Biplot and the heatmap can be downloaded in csv format. All datapoints are log2 Fold-Change values.

Figure 7: AstroBio MultiStress Explorer for *Saccharomyces cerevisiae*

Step 6. Browse through the different tabs to obtain more information.

GENE ONTOLOGY

Overview

Gene Ontology

Gene Location

Spaceflight Essential Genes

Study Overview

Study Information

Molecular Function

ion channel activity,molecular_function

Biological Processes

cellular response to DNA damage stimulus
cellular response to ethanol
cellular response to heat
cellular response to hydrogen peroxide
cellular response to osmotic stress
ion transport
negative regulation of ATPase activity

Cellular Component

endoplasmic reticulum
integral component of membrane
integral component of membrane
membrane
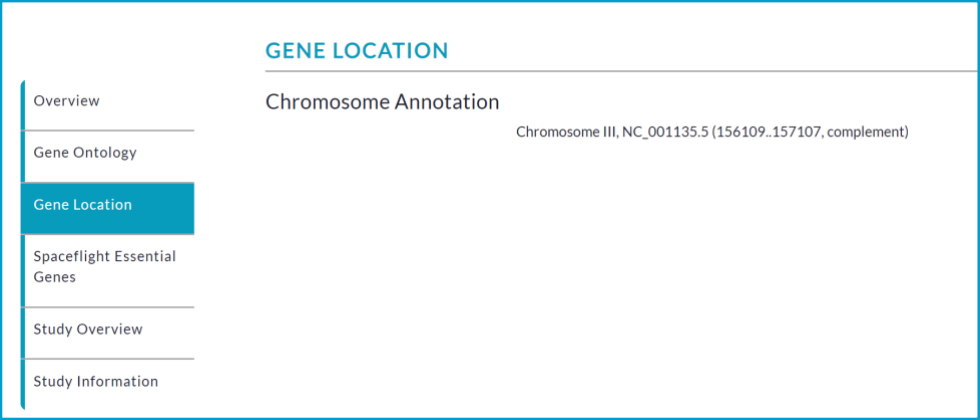plasma membrane
plasma membrane
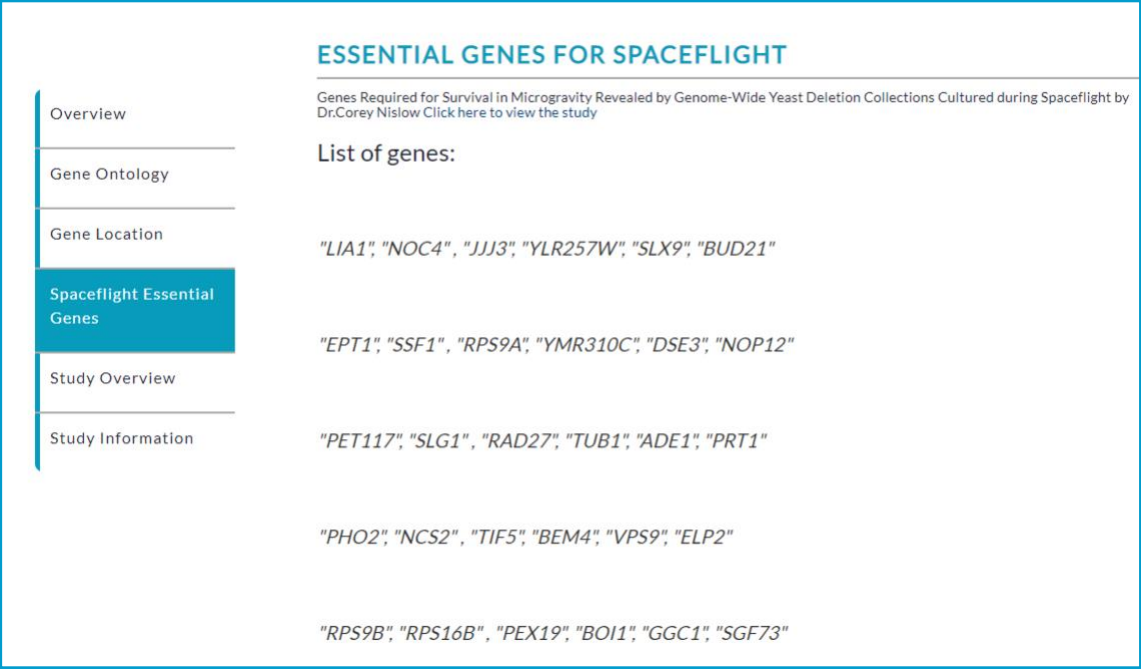
Figure 8: Gene Ontology Tab

Figure 9: Gene Location Tab

Figure 10: Spaceflight Essential Genes

Figure 11: Study Overview Tab

Figure 7: Study Information Tab
*To access the study directly, go to the Link provided in the **Study Information** tab.*

ASTROBIO

iGEM CONCORDIA

Igem.concordia@gmail.com