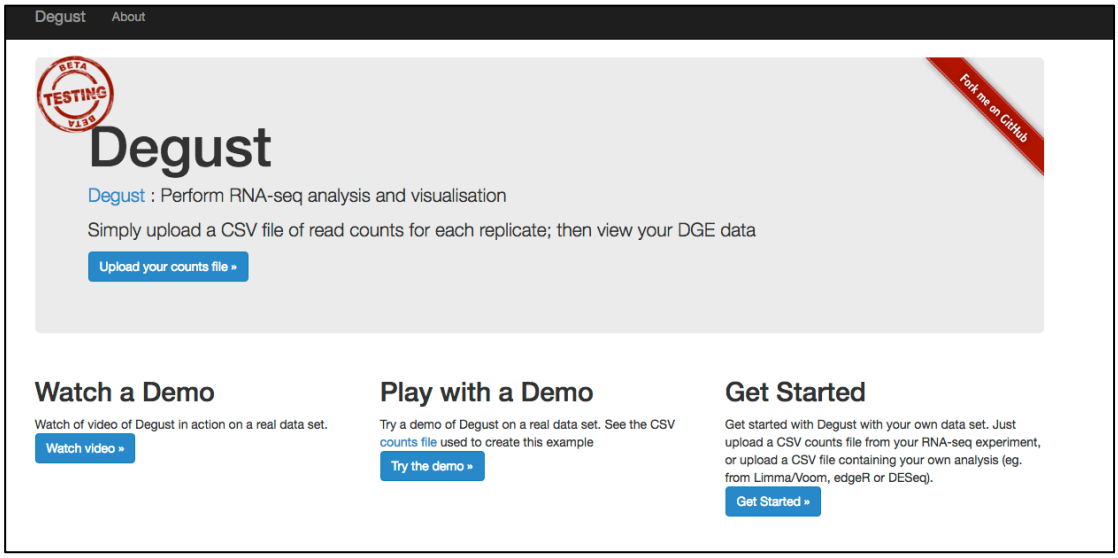
Degust: Visualize, explore and appreciate RNA-seq differential gene-expression data.

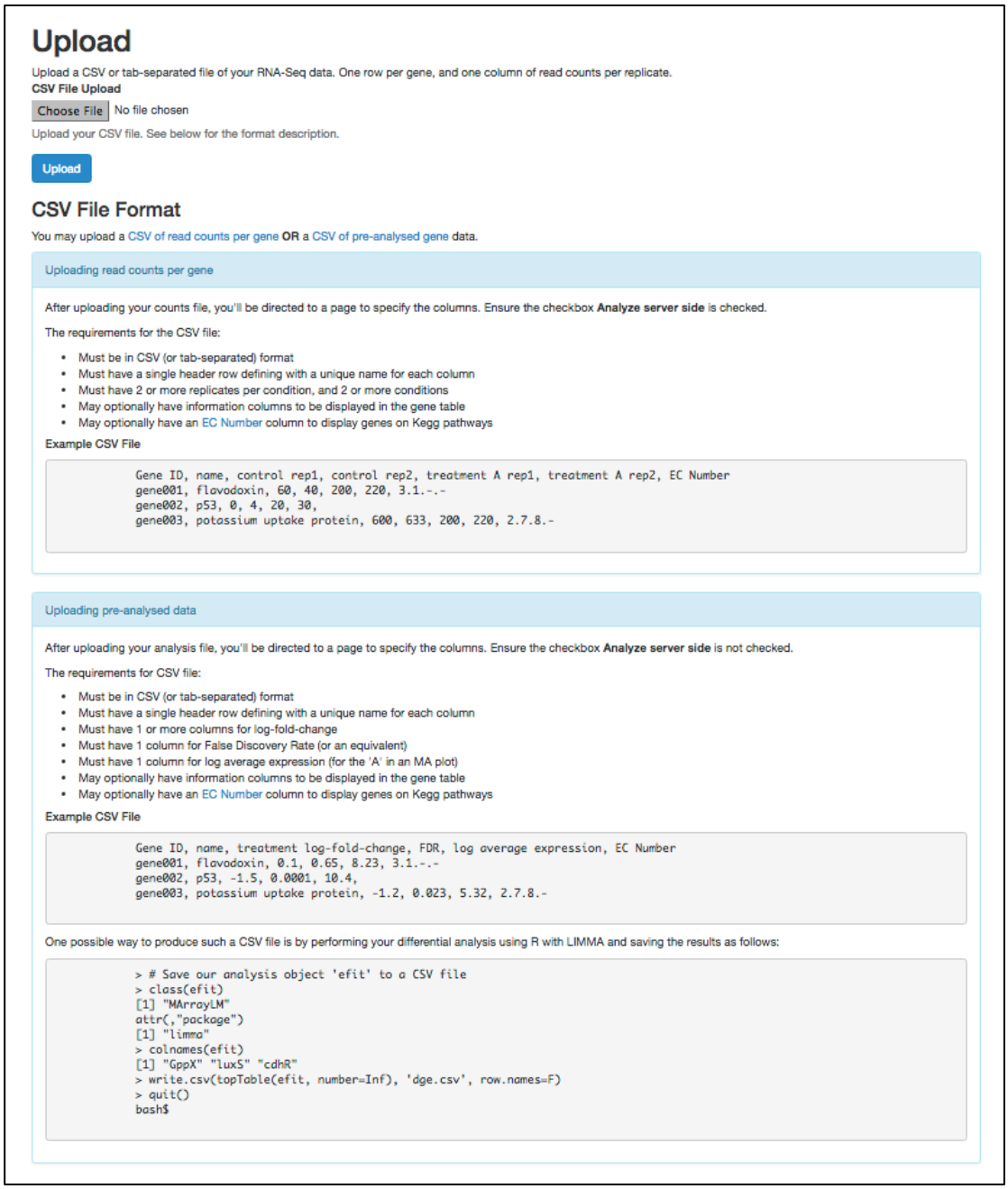
http://www.vicbioinformatics.com/degust/



You can upload raw counts OR pre-analysed data to Degust.

We'll upload the pre-analysed data we generated in R (B.PregVsLacResults.csv)

Upload data



Upload

Upload a CSV or tab-separated file of your RNA-Seq data. One row per gene, and one column of read counts per replicate.

CSV File Upload

Choose File | No file chosen

Upload your CSV file. See below for the format description.

Upload

CSV File Format

You may upload a CSV of read counts per gene **OR** a CSV of pre-analysed gene data.

Uploading read counts per gene

After uploading your counts file, you'll be directed to a page to specify the columns. Ensure the checkbox **Analyze server side** is checked.

The requirements for the CSV file:

- Must be in CSV (or tab-separated) format
- Must have a single header row defining with a unique name for each column
- Must have 2 or more replicates per condition, and 2 or more conditions
- May optionally have information columns to be displayed in the gene table
- May optionally have an EC Number column to display genes on Kegg pathways

Example CSV File

```
Gene ID, name, control rep1, control rep2, treatment A rep1, treatment A rep2, EC Number
gene001, flavodoxin, 60, 40, 200, 220, 3.1.-.-
gene002, p53, 0, 4, 20, 30,
gene003, potassium uptake protein, 600, 633, 200, 220, 2.7.8.-
```

Uploading pre-analysed data

After uploading your analysis file, you'll be directed to a page to specify the columns. Ensure the checkbox **Analyze server side** is not checked.

The requirements for CSV file:

- Must be in CSV (or tab-separated) format
- Must have a single header row defining with a unique name for each column
- Must have 1 or more columns for log-fold-change
- Must have 1 column for False Discovery Rate (or an equivalent)
- Must have 1 column for log average expression (for the 'A' in an MA plot)
- May optionally have information columns to be displayed in the gene table
- May optionally have an EC Number column to display genes on Kegg pathways

Example CSV File

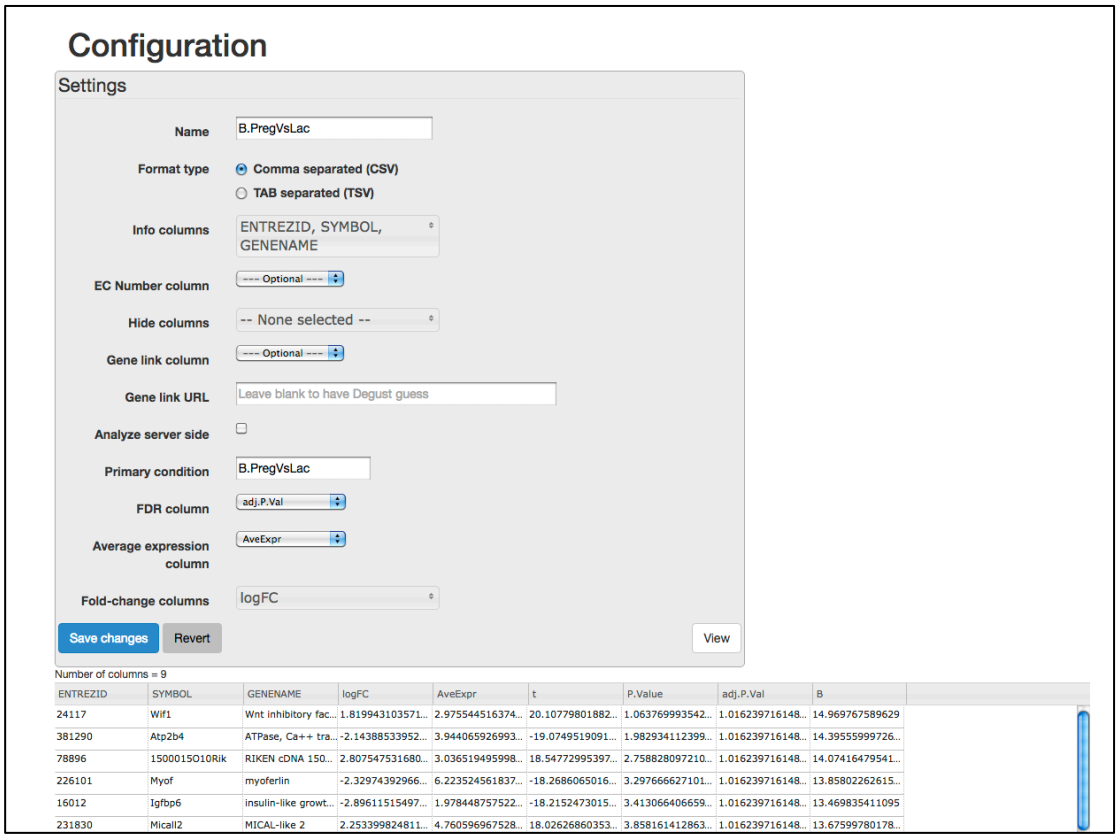
```
Gene ID, name, treatment log-fold-change, FDR, log average expression, EC Number
gene001, flavodoxin, 0.1, 0.65, 8.23, 3.1.-.-
gene002, p53, -1.5, 0.0001, 10.4,
gene003, potassium uptake protein, -1.2, 0.023, 5.32, 2.7.8.-
```

One possible way to produce such a CSV file is by performing your differential analysis using R with LIMMA and saving the results as follows:

```
> # Save our analysis object 'efit' to a CSV file
> class(efit)
[1] "MArrayLM"
attr(,"package")
[1] "limma"
> colnames(efit)
[1] "GppX" "luxS" "cdhR"
> write.csv(topTable(efit, number=Inf), 'dge.csv', row.names=F)
> quit()
bash$
```

- Click the grey **Choose File** button
- Select the B.PregVsLacResults.csv file created in the workshop
- Click the blue **Upload** button

Configure

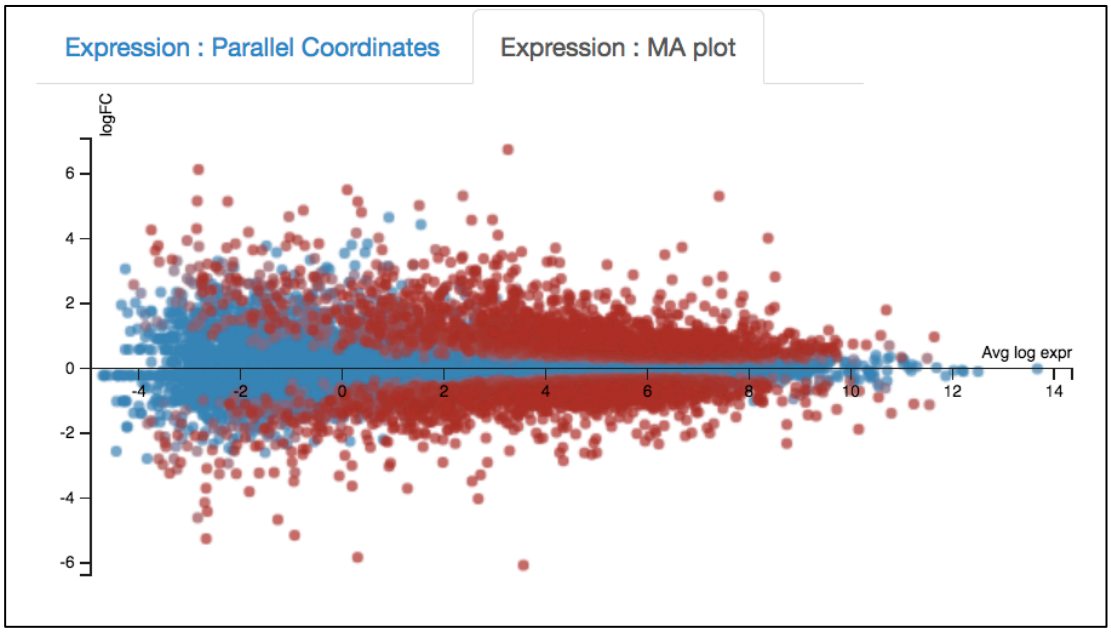


- Name: B.PregVsLac
- Format: CSV (should be already selected)
- Info columns: Select the columns in our csv file that provide info on the genes - Entrezid, Symbol, Genename
- Untick “Analyze server side” as we’re working with pre-analysed data not raw counts. This causes more boxes to appear. Set the columns as below:
- Primary condition as our Contrast:
- FDR column: adj.P.Val
- Average expression column: AveExpr

Click the blue **Save Changes** button

A box should pop up with “Saving settings” click **View**

Interactive MA plot

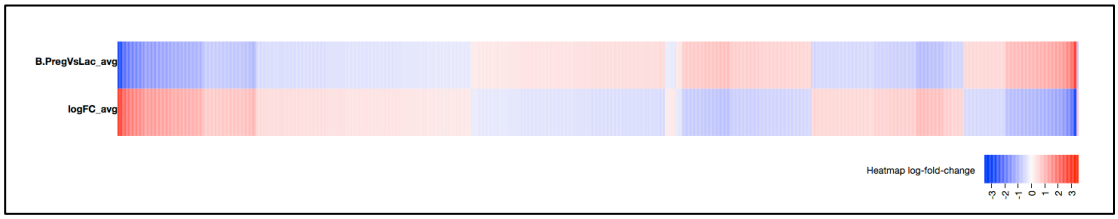


- This is like the static MA plot that we created in R, but interactive
- Shows expression for 2 conditions
- Average expression is on the x axis, logFC on y, each dot is a gene
- Highly expressed genes are towards the right, lowly expressed towards the left
- Upregulated genes are above the horizontal line, downregulated below
- Red means more significant FDR, blue means less significant

Hover over the dots in the plot to see gene info: Entrez id, Symbol, Gene Name, A, M, FDR.

Click and Drag on the plot to select genes. The heatmap and table will be filtered to show just those genes. Click anywhere on the plot to remove the rectangle.

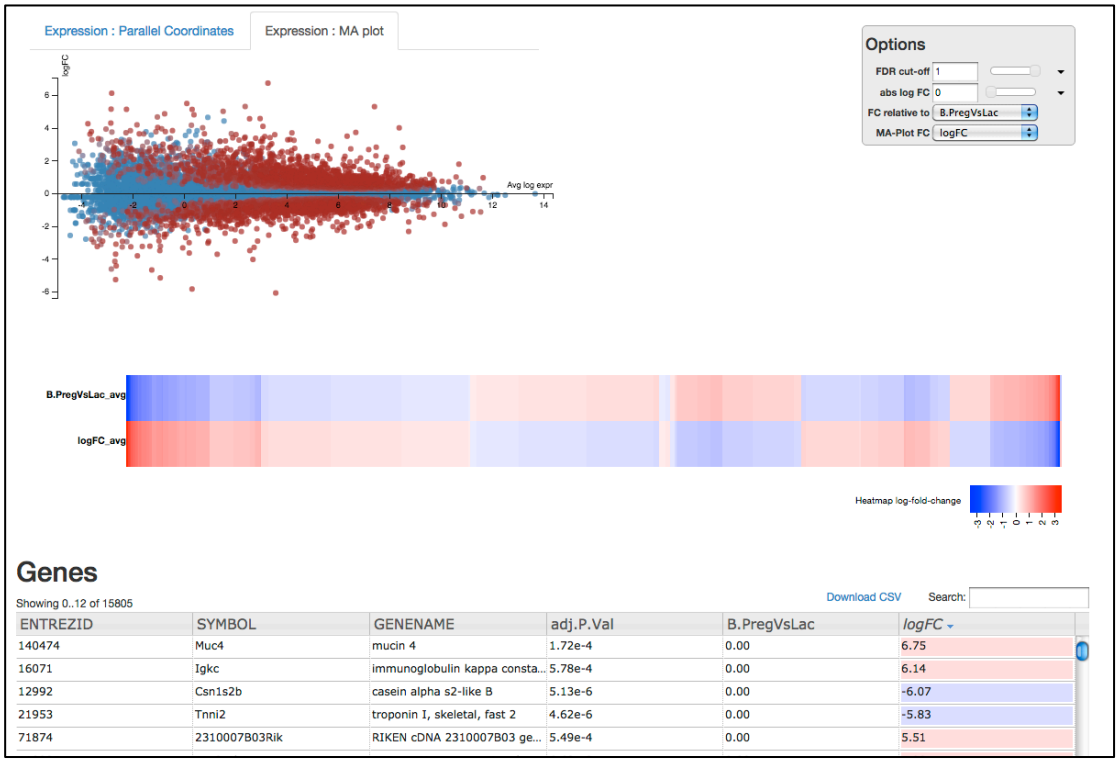
Interactive heatmap



- Displays all genes loaded
- Upregulated genes are red in the logFC panel, downregulated genes are blue

Hover over heatmap to see gene info and the gene highlighted in the MA plot

Dynamic filtering of plots and table



- Can easily filter genes and get dynamic recalculation of results
- Filter by FDR
    - Can type desired threshold,
    - Or use slider,
    - Or use dropdown arrow to select from commonly used thresholds (0.05, 0.01 etc)
- Filter by logFC
    - Can type desired threshold,
    - Or use slider,
    - Or use dropdown arrow to select from commonly used thresholds (1.5x, 2x etc), notice that it shows e.g. logFC of 1 is equivalent to a 2x fold change.

Apply the filters FDR 0.05 and logFC 1

Notice

- The total number of genes that meet these thresholds can be seen
- How the MA plot and heatmap have changed; the gap in the MA where the genes have been filtered out and that there are now less genes in the heatmap,

Table:

- Can sort the table of genes by column headers
- Can download a file of filtered genes

- Can search for a specific gene e.g. your favourite gene.

Let's have a look at the gene we visualised with the stripchart in R, Wif1 (Wnt inhibitory factor 1).
Type Wif1 in the search box, then in the Table doubleclick on Wif1 to open the NCBI Gene page for that gene.

Exercises:

1) Load in the B.PregVsLac csv file that we created in R into Degust and explore

How many genes are differentially expressed:
At logFC 0 i.e. no threshold for logFC
FDR 1
FDR 0.05
FDR 0.01
FDR 0.001
FDR 0.0001

If you set the logFC to 1 and repeat with the FDR thresholds above, now how many genes are there?

2) Create a csv file in R for the L.PregVsLac contrast, load that into Degust and repeat 1).