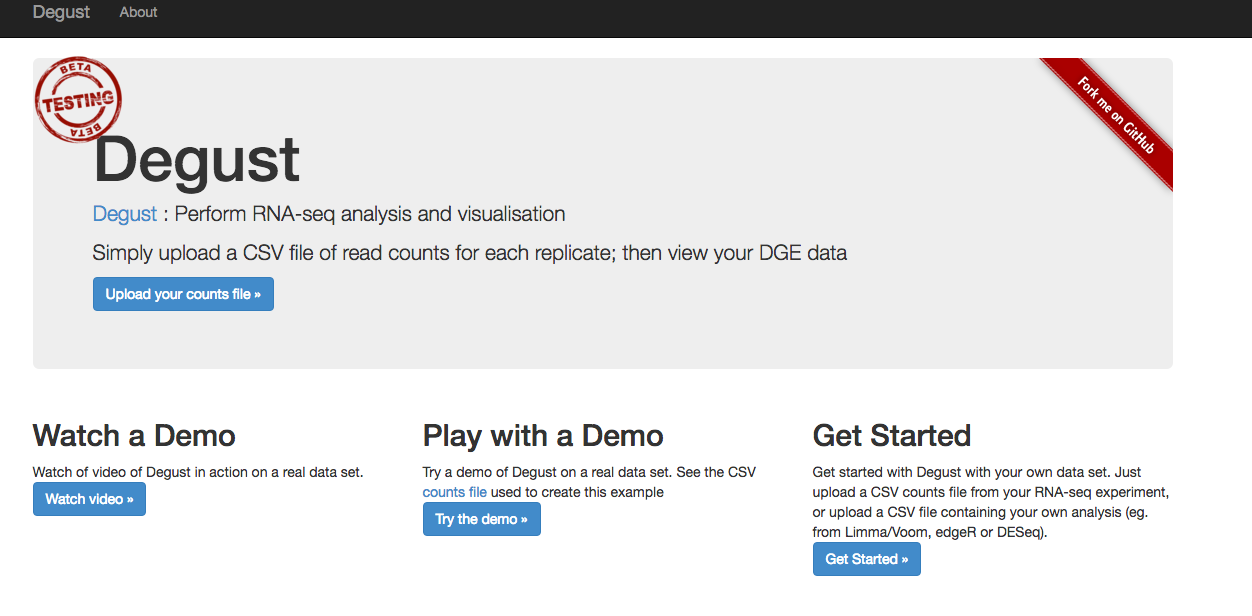
**Degust: Visualize, explore and appreciate RNA-seq differential gene-expression data.**

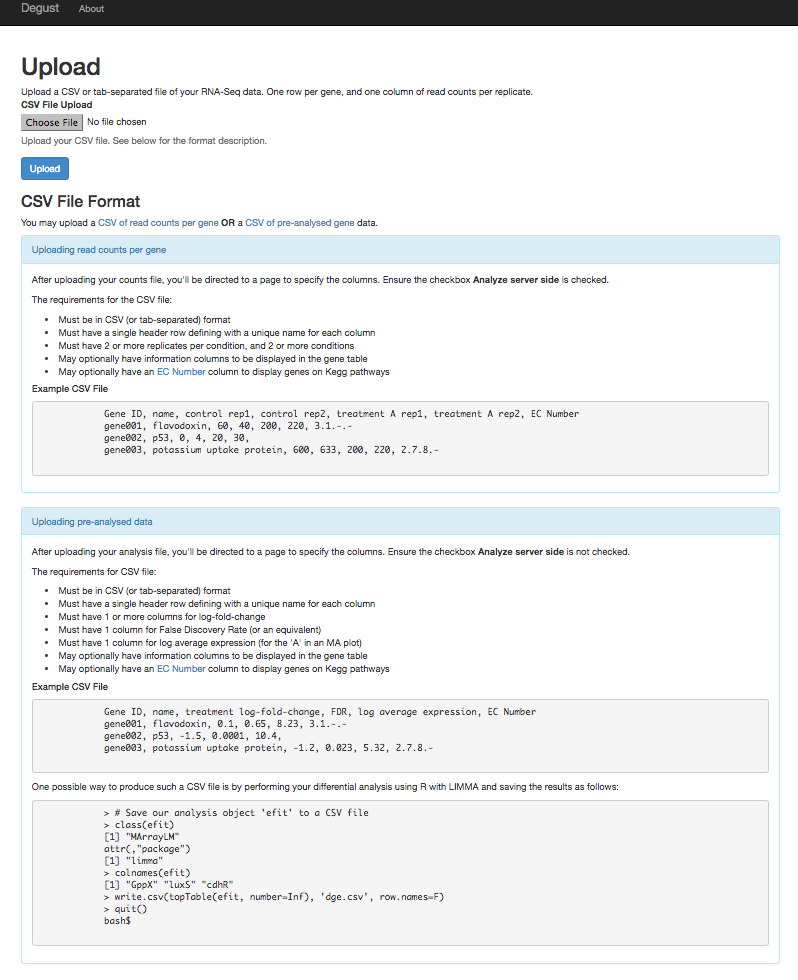
<http://www.vicbioinformatics.com/degust/>



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

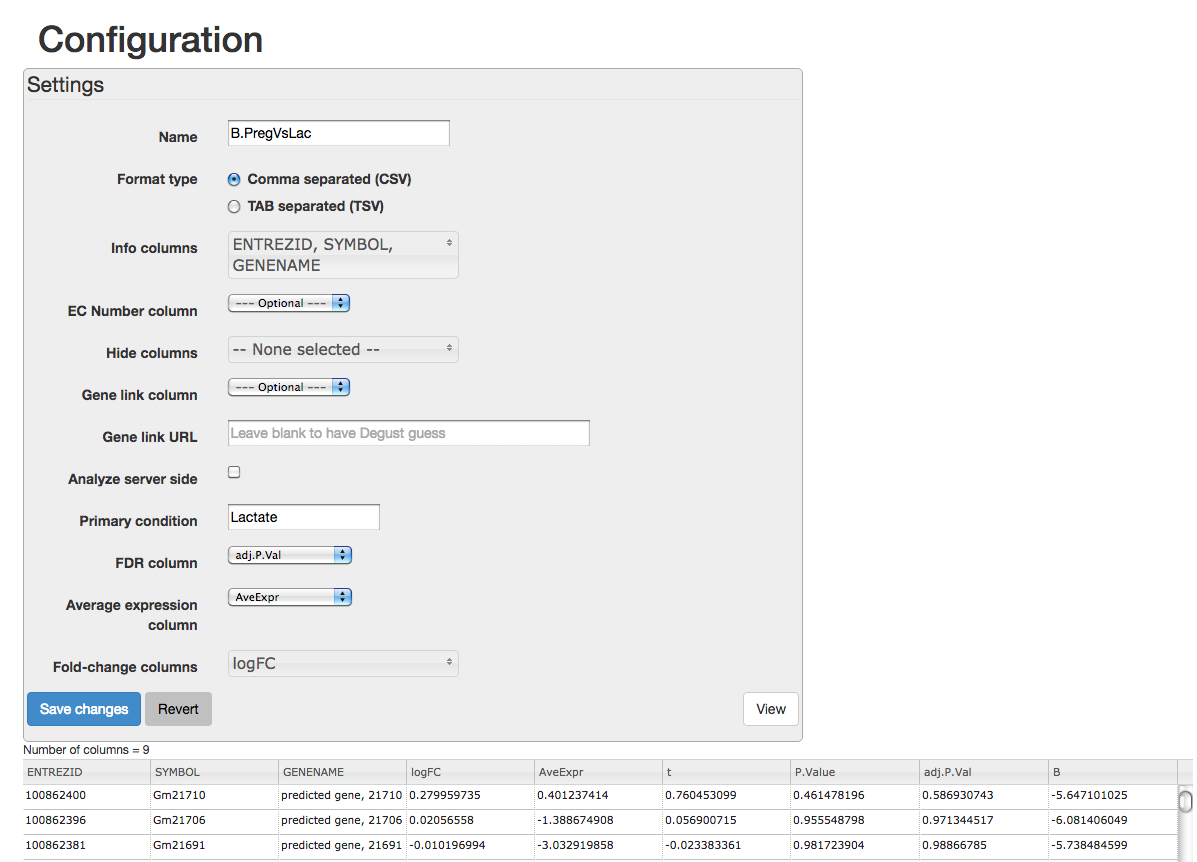
We’ll upload the pre-analysed data we generated in R (B.PregVsLacResults.csv)

**Upload data**



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

**Configure**

****

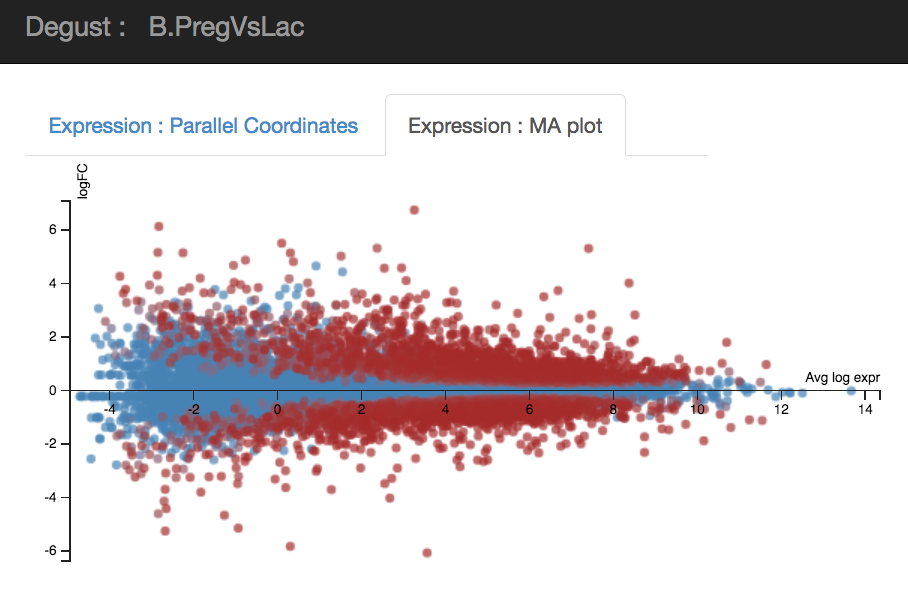
Set the configuration as follows:

* Name: B.PregVsLac (the name of our analysis)
* Format: Comma separated (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: Lactate (the condition that the logFC is relative to)
* FDR column: adj.P.Val
* Average expression column: AveExpr
* Fold-change columns: logFC

Click the blue **Save Changes­** button

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

**Interactive MA plot**

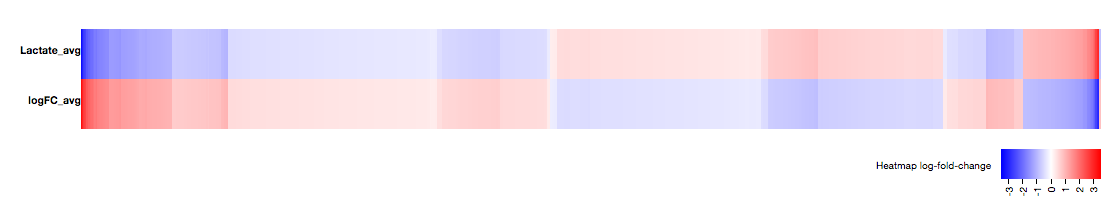


* This is similar to the static MA plot that we created in R and the interactive plot we created in R with Glimma
* Shows expression for 2 conditions
* Average expression is on the x axis, logFC on y, each dot is a gene
* Highly expressed genes are towards the right, lowly expressed towards the left
* 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.

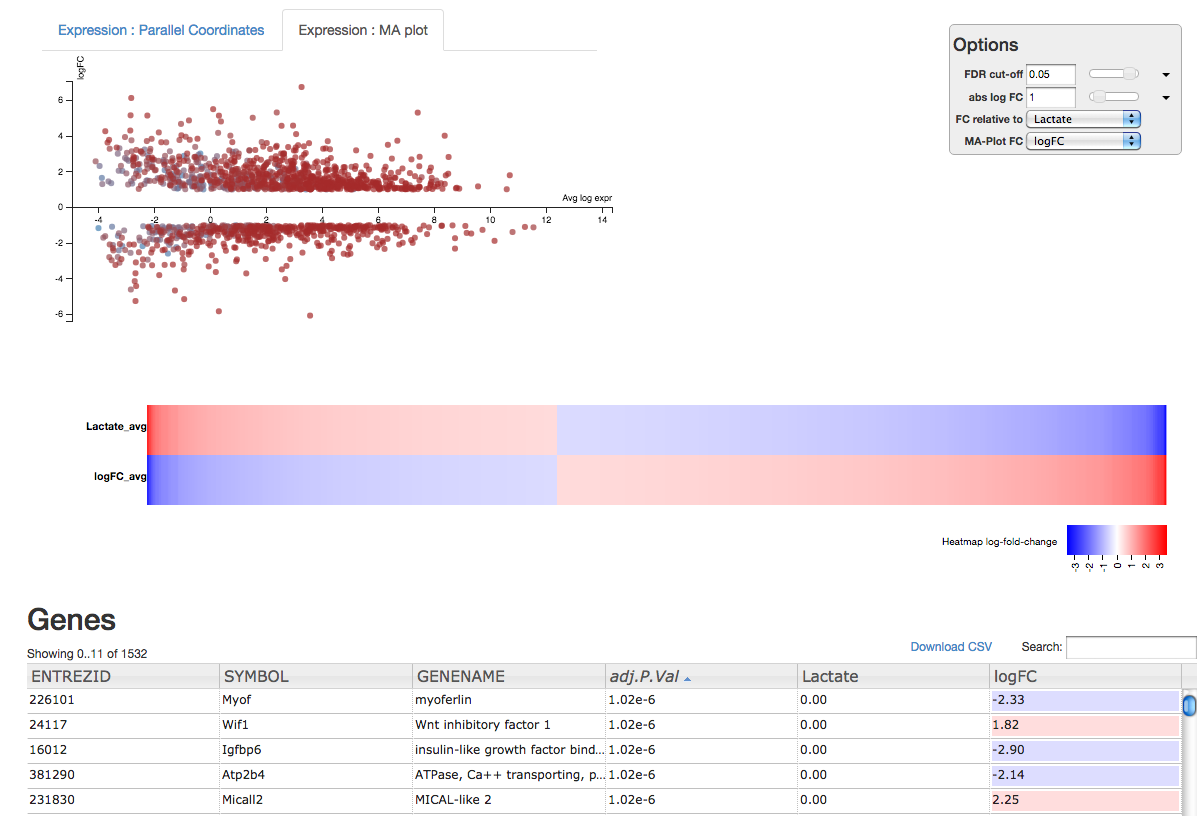
**Interactive heatmap**



* 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 cut-off” = 0.05 and “abs log FC” = 1

Notice

* The total number of genes that meet these thresholds can be seen above the table under “Genes”
* 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 csv 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 and

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).

**Answers:**

Note: If Degust reports 15805 genes for L.PregVsLac (or  B.PregVsLac) then it's including a blank line. If it happens to correctly identify that there are 15804 genes for L.PregVsLac then subtract 1 from the no LogFC numbers below.

1) With **B.PregVsLac**

no logFC cutoff and

FDR 1: 15804 genes

FDR 0.05: 5340 genes (note that 5340 is the number we got with decideTests in R post limma-voom, see day 1 html)

FDR 0.01: 2995 genes

FDR 0.001:  1092 genes

FDR 0.0001: 232 genes

logFC set to 1 and

FDR 1:  2179 genes

FDR 0.05:  1532 genes

FDR 0.01:  1068 genes

FDR 0.001:  586 genes

FDR 0.0001:  210 genes

2) With **L.PregVsLac** (there's a blank line NaN included in the numbers below when no logFC cutoff is used)

no logFC cutoff and

FDR 1: 15805 genes

FDR 0.05: 7354 genes (note that 7353 is the number we got with decideTests in R post limma-voom, see day 1 html)

FDR 0.01: 5325 genes

FDR 0.001: 3183 genes

FDR 0.0001: 1782 genes

logFC set to 1 and

FDR 1: 4300 genes

FDR 0.05: 3257 genes

FDR 0.01: 2683 genes

FDR 0.001: 1978 genes

FDR 0.0001: 1378 genes