

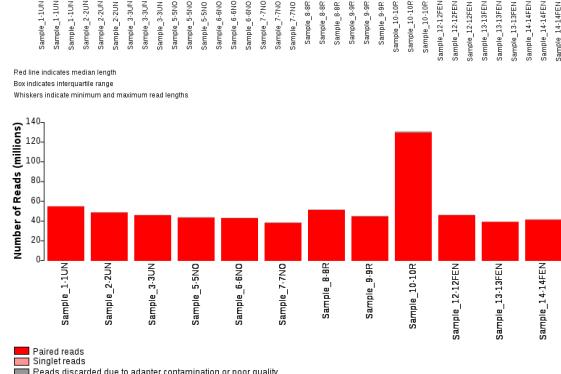
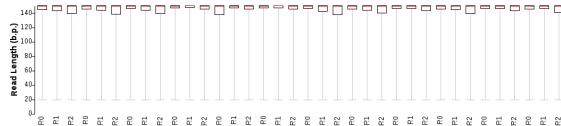
Aedes aegypti transcriptome: resistance to fenitrothion

Xavier Grau-Bové

Experimental setup

12 samples divided in 4 groups (3 identical replicates per group). 6 fenitrothion-resistant samples from Angola (3 have been exposed to fenitrothion, 3 unexposed). 6 susceptible samples from laboratory strains (3 from New Orleans, 3 from Rockefeller). Sample classification:

Sample name	Sample group	Phenotype	Exposure	Location
s01_RUN	ResUnex	Res	Unex	Angola
s02_RUN	ResUnex	Res	Unex	Angola
s03_RUN	ResUnex	Res	Unex	Angola
s05_SNO	SusNew0	Sus	Unex	NewOrl
s06_SNO	SusNew0	Sus	Unex	NewOrl
s07_SNO	SusNew0	Sus	Unex	NewOrl
s08_SRO	SusRock	Sus	Unex	Rockef
s09_SRO	SusRock	Sus	Unex	Rockef
s10_SRO	SusRock	Sus	Unex	Rockef
s12_RFE	ResFeex	Res	Feex	Angola
s13_RFE	ResFeex	Res	Feex	Angola
s14_RFE	ResFeex	Res	Feex	Angola



Excess of reads in s10 due to miscalculation of RNA concentration prior to library preparation. It does not seem to have an effect on final results.

Comparisons for all analyses – these:

- Independent comparisons of susceptible strains (SRO, SNO) to resistant-unexposed (RUN), to uncover genes that are constitutively differentially expressed in the resistant background (independently from specific exposure conditions).
- Analyse overlap of DE genes between SRO-RUN and SNO-RUN – same genes should appear.
- Compare resistant unexposed (RUN) – resistant exposed (RFE), to find genes that are DE upon exposure.

Insecticide info – Fenitrothion is an organophosphate. Insecticide resistance should be due to increased expression of detoxification enzymes (p450, GSTs?), or changes in acetylcholine receptors, etc.

Quantification: DE & AS

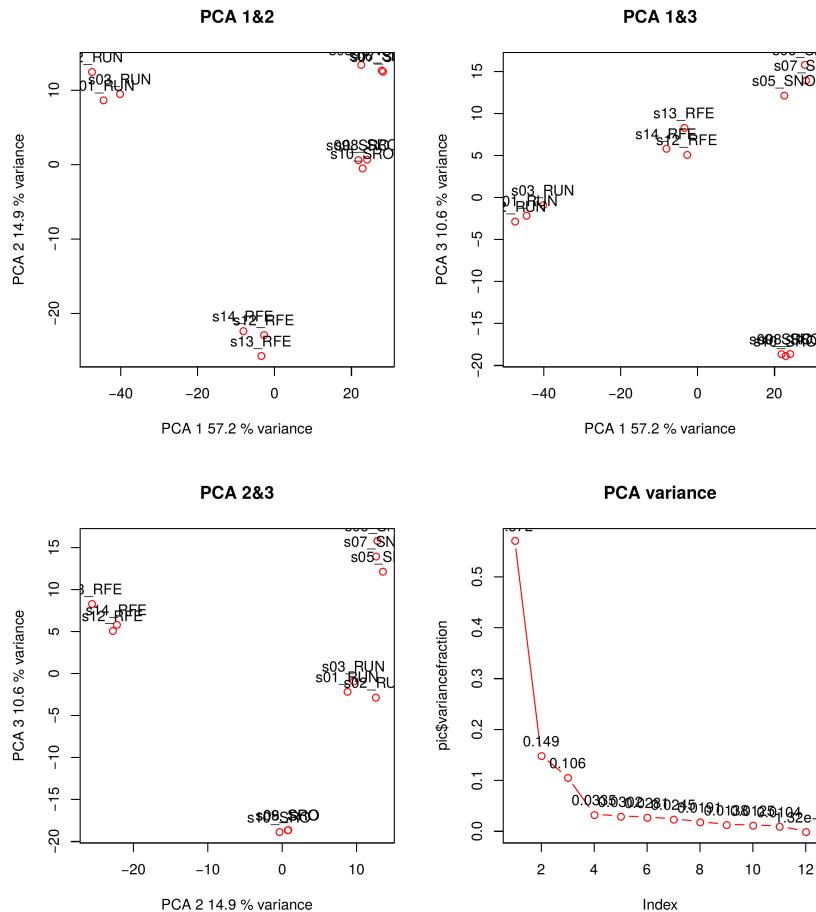
Gene expression – (i) mapping of all RNA-seq samples with *hisat*; and (ii) quantification of expression levels of longest isoform per gene using *salmon* (TPM); then (iii) TPM matrix is loaded onto *Deseq* to produce matrix of normalized transcript counts, log2 fold-change between sample groups.

Alternative splicing – (i) same *hisat* mapping; (ii) assembly of sample-specific transcriptomes with isoforms using *stringtie*, followed by creation of pan-sample consensus assembly; (iii) quantification of expression levels using *salmon* (TPMs) using both all isoforms and longest isoform per gene; (iv) construction of sample-specific isoform and AS event catalogues using *suppa*; and (v) quantification of AS events using *suppa* too.

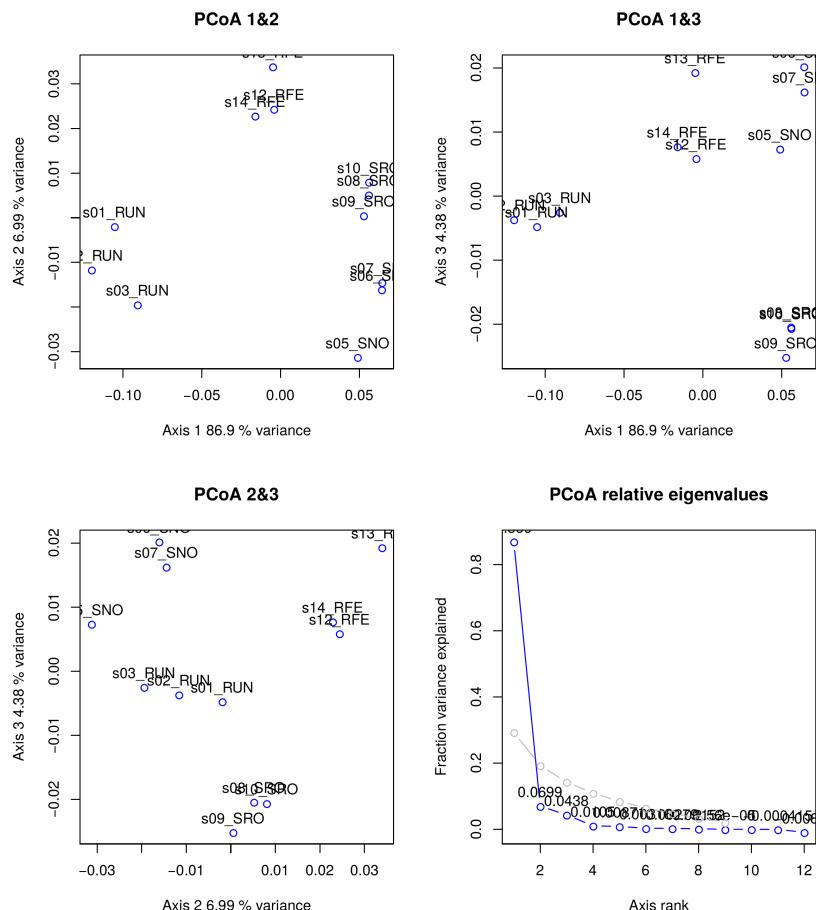
Gene expression: clustering of all samples

PCA analysis – Using the matrix of normalized transcript counts from *Deseq*, calculating Euclidean distances between samples. Good within-replicate clustering.

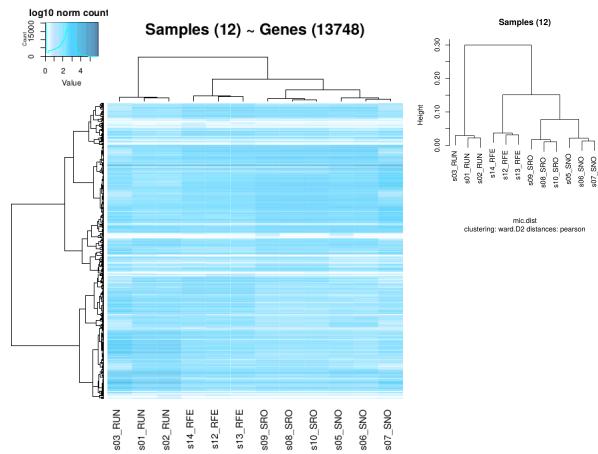
PC1 separates susceptible from resistant samples and is by far the most important. PC2 separates exposed from non-exposed. Oddly, RFE seems to be closer to susceptible samples than RUN, along PC1. PC3 highlights differences between SRO and SNO and it is not very relevant (not the focus of this analysis, although it can help contextualize the overlap differences in SRO/SNO-RUN comparisons),



Analogous PCoA analyses show the same overall clustering (using Pearson correlation-derived distances + Ward clustering), with some more intra-replicate differences.



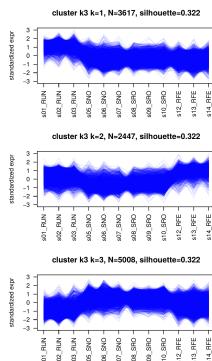
Sample clustering – Heatmap of log10 normalized transcript counts, clustering samples and transcripts by common gene expression profile (Pearson correlation-derived distances + Ward clustering). Similar to PCAs: it recovers RUN samples as being the most distant (it mostly reflects the same tendencies as PC1).



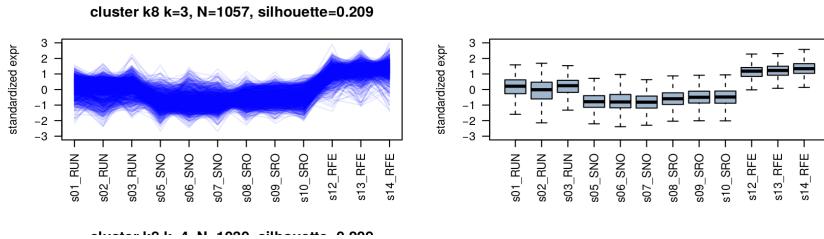
All in all, results make sense. Note that S10 (which had an excess of reads) does not seem to be weird in any analysis.

Gene clustering using k-means – A blind clustering of genes that have similar expression patterns, using various pre-defined number of clusters (k=3, 4, 5, etc.). For this analysis, log10 normalized counts are standardized (mean=0, standard deviation=1).

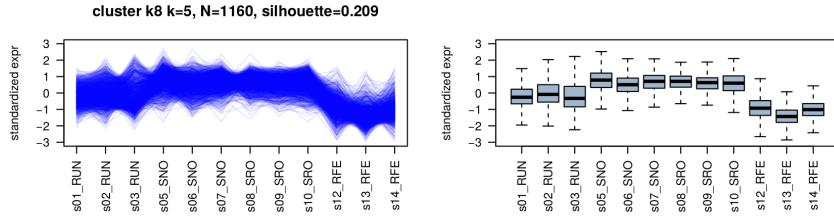
At low values of k, most genes are aggregated in meaningless groups, but with higher k some interesting groups may appear.



For example: at k=8, there is a cluster (3) of 1057 genes that are over-expressed in resistant samples, and RFE>RUN too:



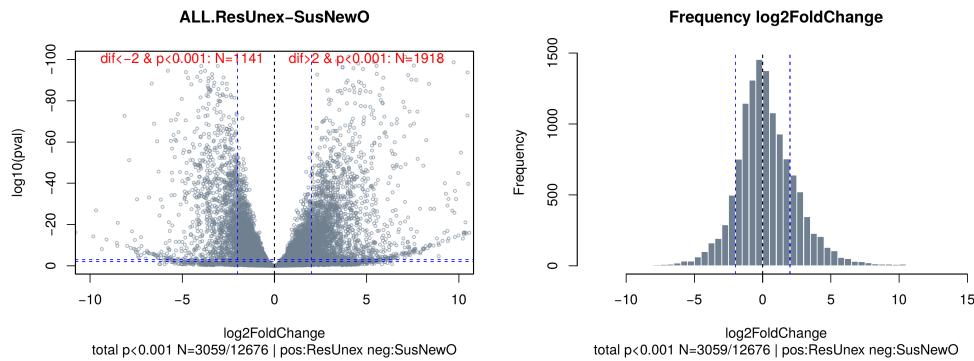
This other cluster represents genes that are under-expressed in RUN and RFE (and more under-expressed in RFE than in RUN):

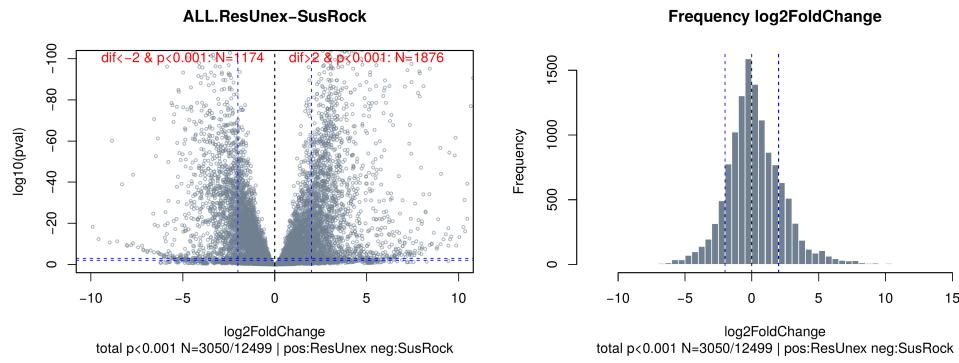


This sort of analysis is *not particularly useful* for gene expression analysis (formal DE is better here) but it helps visualize and quantify the amount of dynamically regulated genes. In other types of analyses such as AS, this can help to identify clusters where normal differential comparisons may fail.

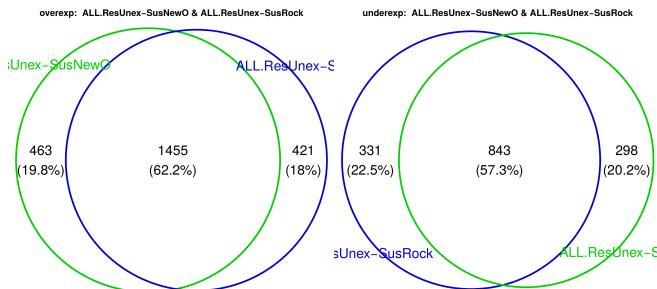
Comparison of RUN (resistant unexposed) to SNO/SRO (susceptible)

In comparisons of unexposed resistant (RUN) ~ unexposed susceptible (SNO & SRO), results are quantitatively quite similar: almost 2,000 genes are overexpressed in the resistant samples, and ~1,100 are underexpressed (thresholds at $p=1e-3$ and FC =4). There are thousands of genes with less lower DE levels, so tuning FC looks key here.



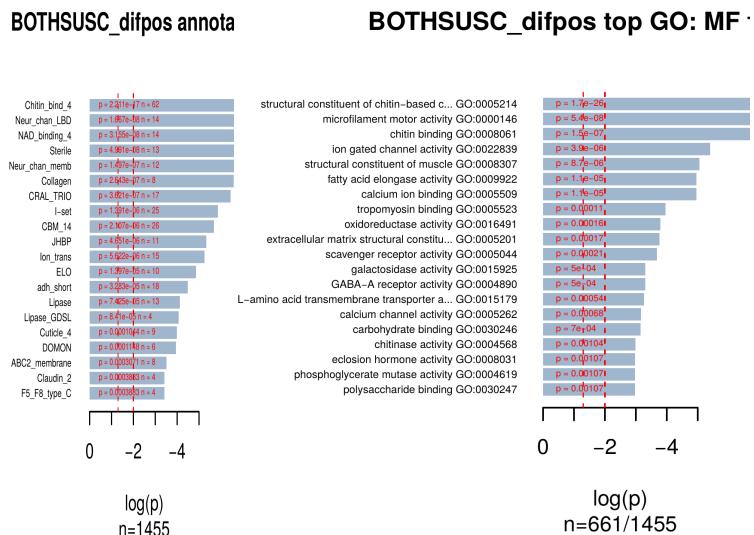


Intersection – Overlap of DE genes in both analyses is pretty good, at ~60% for both overexpressed and underexpressed genes.

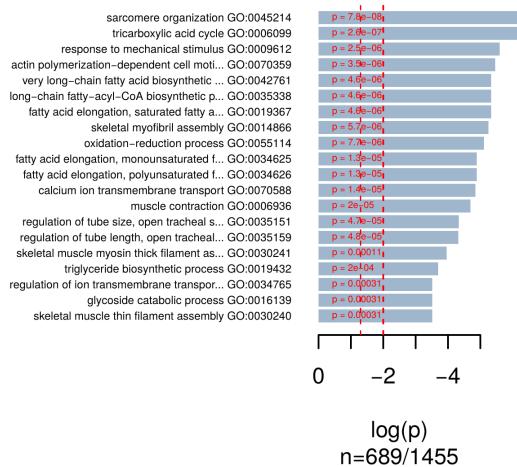


Functional enrichments – Genes that are **overexpressed in resistant when compared to both susceptible strains** shows some protein families involved in:

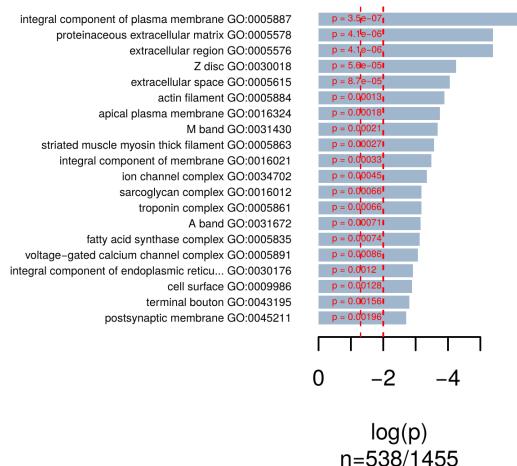
- chitin binding (Chitin_bind_4, CBM_14; confirmed in MF GO enrichments, second panel)
- cuticle thickening (Cuticle_4)
- ion transport (Neur_chan_LBD, Ion_trans),
- fatty acid metabolism (ELO, Sterile, confirmed in BP GO enrichments, third panel);
- signalling pathways: small G-proteins signalling (CRAL_TRIO), juvenile hormone (JHBP)
- various receptors: DOMON, I-set
- ...



BOTHsUSC_difpos top GO: BP 1



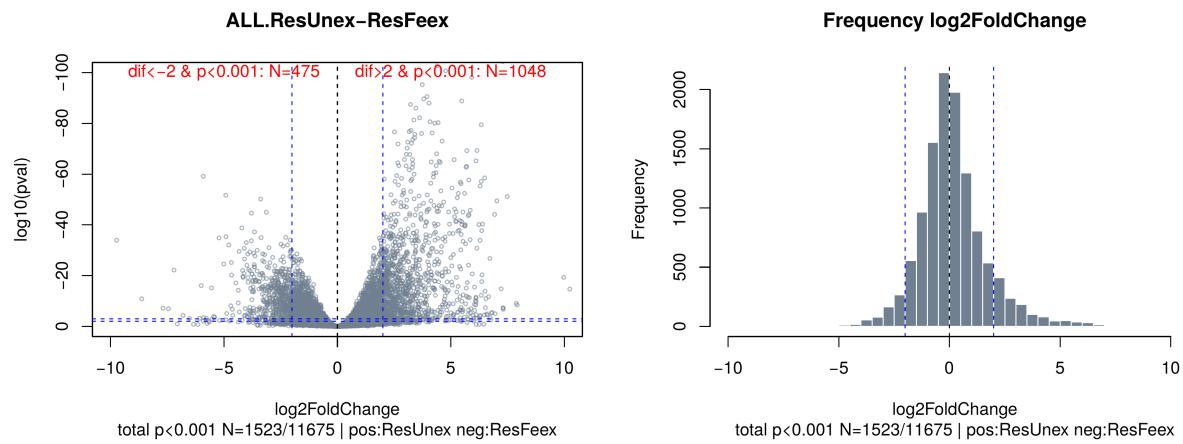
BOTHsUSC_difpos top GO: CC



The functional enrichments in the list of overexpressed genes in the resistant compared to individual susceptible strains looks pretty similar.

Comparison of RUN (resistant unexposed) RFE (resistant exposed)

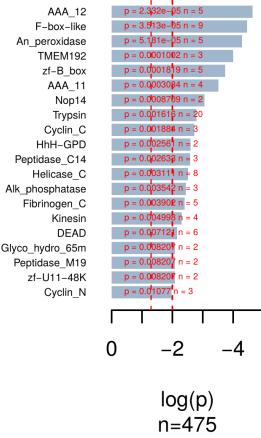
The overall differences between resistant exposed and unexposed are less strong than the differences between resistant and susceptible, with just ~1500 DE genes (instead of >3,000 in the previous comparisons). The **exposed individuals seem to exhibit a slight down-regulation of gene expression**, with twice as much underexpressed genes (~1000) compared to overexpressed (~500).



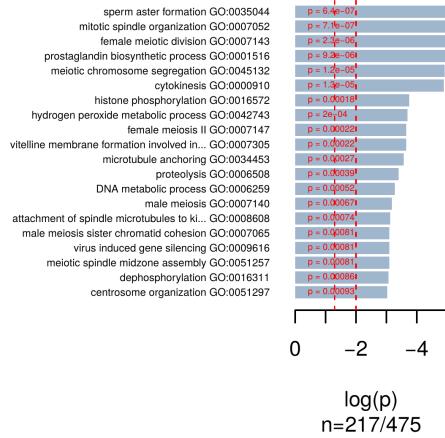
Functional enrichments – the functional enrichments of **genes overexpressed in exposed** compared to unexposed shows a conspicuous lack of interesting annotations (detoxification), but some stress response. Putatively interesting things are:

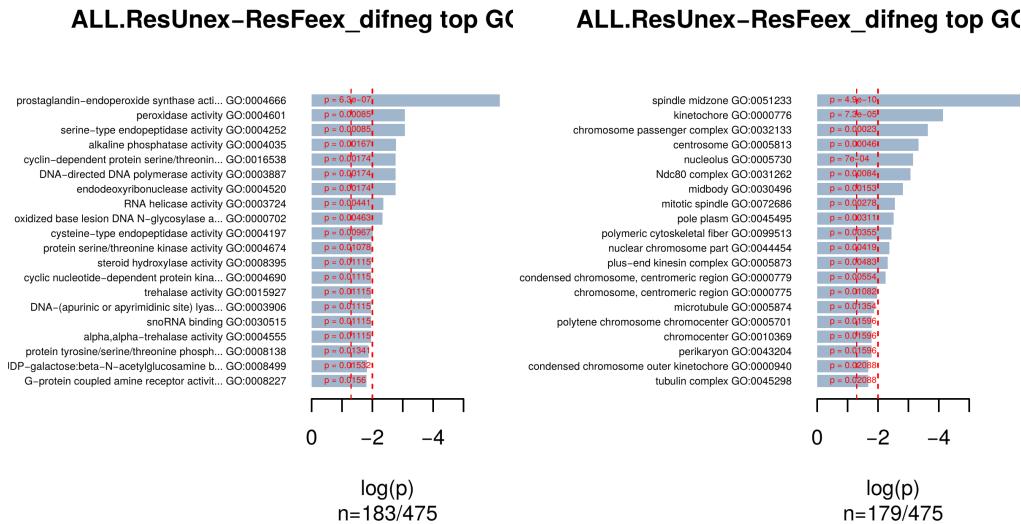
- A general stress response could be indicated by overexpressed Trypsins and Peptidases (C14/M19), also possibly ubiquitination enzymes (F-box-like, zf-B_box?)
- An_peroxidase?
- AAA_12? ATPases associated with diverse cellular activities

ALL.ResUnex-ResFeeex_difneg a



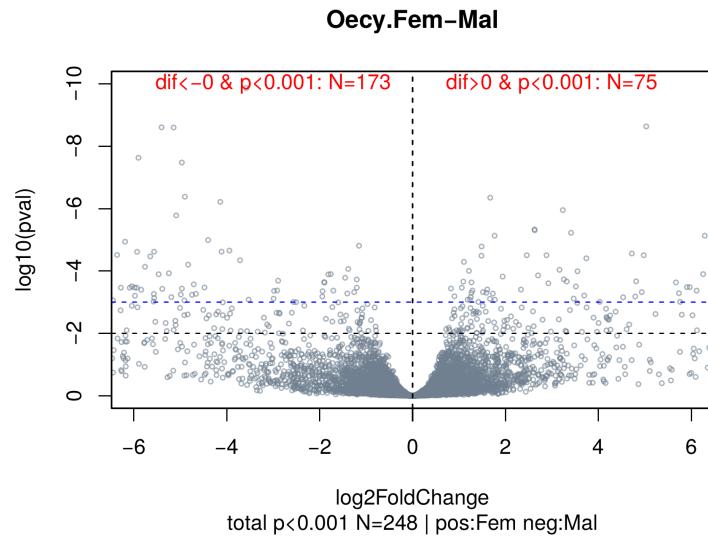
ALL.ResUnex-ResFeeex_difneg top GO





Comparison of female oenocytes v. male oenocytes

Sex-specific signal within oenocyte samples only became visible in PC3, so it is understandable that differences are not as strong as in other comparisons. In this case, there are **248 differentially expressed genes**, with a ~2x bias towards over-expression in male oenocytes.



Is this important at all?

Alternative splicing patterns

Globally, AS patterns se