Differential Expression on the Web

**DEW is a platform that allows users to explore RNA-Seq data using a Graphical User Interface.** Users select two or more RNA-Seq readsets and provide a dataset of genes to be used for ascertaining their expression profiles.

**Administrators** can rapidly setup the server using the Drupal Content Management System (<http://drupal.org>) and can specify which RNA-Seq readsets (either Single- or Paired-end) are used by providing the program with the files (as .bam or fastq.bz2) and a user-friendly name.

**End-users** can then provide their own list of genes and perform two types of analyses: ‘survey’ and ‘contextual’. In the survey mode, each gene is investigated independently against all reads and the results are stored in a database. If a gene has already been checked against the same readsets (this is validated using a unique checksum signature derived from the actual sequence rather than any user-provided gene names), the database is used to rapidly retrieve alignment statistics so that users need not to wait for alignments to be completed. In the contextual mode, alignments of reads are considered in relation to all of the genes that have been provided and certain corrections are made to account for isoforms and close paralogues. Further, DEW checks if significant differential expression is present in some parts of a gene even if overall expression is not significantly different.

**The analysis proceeds as such**: gapped alignments are performed and corrected for length, PCR and fragment bias so that a Fragment Per (effective) Kilobase per Million of reads (FPKM) is estimated as well as the simpler Reads Per Kb per Million of reads (RPKM). When provided with multiple isoforms and in the ‘contextual’ mode, corrections include a expectation maximization algorithm estimates effective expression profiles and a corrected alignment is produced. For each gene the user provides, DEW computes coverage descriptive statistics (RPKM, FKMP and total, mean and median corrected counts), expression profiles (normalized as R/FPKM and trimmed mean of fold change). DEW allows users to explore the data by providing interactive graphs that show the (corrected) depth of read coverage across each gene (as a line graph), fold change versus absolute abundance (as 2D and 3D scatter plots with data points coloured by (2D) or with the third axis being (3D) the False Discovery Score for each test) and relative expression.

**The results are presented** as an interactive report that allows domain-experts to explore and interface with the results. Except the interactive graphs, the report includes: a list of genes that are housekeeping (i.e. very little fold change) or significantly differentially expressed between i) each pair of readsets or ii) across all readsets.

## To do

* A demo is available on <http://insectacentral.org/bench/dew>
* Support replicates
* Allow filtering by FDR and FC