MACE Results: Short Description of the Excel-File

The results of the digital gene expression analyses are provided in tab-separated format.
The table should be opened with Excel 2007 or higher, otherwise the amount of rows is limited to 65,536.

Decimal separators:
The numbers are provided with points as separators for decimals. Excel uses decimal separators according to the regional settings. If in your region commas are used as separators for decimals, please follow the instructions described under the following link to change this temporarily:
http://office.microsoft.com/en-us/excel-help/change-the-separator-for-thousands-and-decimals-HP003089676.aspx?CTT=1

Filter:
The data can be conveniently analyzed by using the "filter" option of Excel, to filter for example for keywords e.g. "Transcription" or for tags that are strongly differentially expressed (pvalue, fold change etc.). The filters can be combined. For more information please visit
http://office.microsoft.com/en-us/excel-help/filter-data-in-a-range-or-table-HP010073941.aspx

Group:
Replicates are referred to as "group". Calculation for p-value and FDR consider the variation within the replicates.

The following information is provided in separated columns:

Id:	The **identifier** of the best matching database entry(s) (=hits) identified. When a genome was used for annotation, this corresponds to the feature of the genomic region were the tag matches to. The ID can be hyperlinked to the database entry in the WWW or to a separate file.
gene_symbol:	If available, the gene symbol is displayed here (http://ghr.nlm.nih.gov/glossary=genesymbol).
description:	Description of the transcript, originating from the database, if available.

Mapping types:

cluster_bed: Reads that map successfully to the genomic regions with no annotation (feature) are clustered. The clusters (consensus sequences) are annotated via BLASTX to the Uniprot database. The result is displayed in the “description” column, if no hit was found it will be displayed as "uncharacterized RNA".

genome: Amount of reads mapping to a genomic feature

genome/transcriptome: Amount of reads mapping to the same gene either on the genome or on the transcriptome level. The counts were summed up.

blast_assembly: Reads that could neither be mapped on the genome nor transcriptome are assembled and annotated via BLASTX to the Swiss-Prot and TrEMBL databases, respectively. The outcome is displayed in the “description” column, if no hit was found it will be coined as "uncharacterized RNA".

Normalized_(group ID): The average raw count of each gene within a library are divided by sum of all reads in the sample multiplied with one million.

log2FoldChange (group ID1_vs_ID2: The logarithm to the basis 2 of the Fold Change of the average of the normalized values

p-value (group ID1_vs_ID2): The p-value describing the likelihood for DE based on normalized read counts. It is only provided for pairwise comparisons, it describes the probability of a transcript to be differentially expressed. For samples without biological replicates the p-value for differential expression is calculated using the DEGseq R package http://bioinfo.au.tsinghua.edu.cn/software/degseq).
If biological replicates are available, the p-values are calculated using the DEseq R package (https://bioconductor.org/packages/release/bioc/html/DESeq2.html)

FDR (group ID1_vs_ID2): False Discovery Rate for differential expression (DE) according to Benjamini and Hochberg (1995). The lower the FDR, the more likely is the DE.

Raw_(group ID): raw-counts of reads annotated to the database ID