

RAMPAGE TSS expression matrix

The ENCODE RAMPAGE transcription start site (TSS) expression matrix includes expression of annotated GENCODE TSSs for multiple experiments. This matrix is obtained by collecting into a single file the annotated TSS quantification files produced by the ENCODE3 RAMPAGE pipeline. The input quantification files contain expression levels (raw read counts) for annotated TSS sites, extended by 50bp upstream. The output matrix file contains quantifications for all annotated TSSs with no additional filtering. Values from each biological replicate of each experiment are preserved.

Experiments in the RAMPAGE expression matrices are grouped by reference genome assembly and annotation version :

- hg19 and GENCODE v.19 (human)
- GRCh38 and GENCODE v.24 (human)

The output matrix is created using an in-house shell script and is provided in TSV format. It contains information about TSS as a rows and all replicates of the same experiment as columns. The TSS information is stored in three columns: the TSS ID as it is provided in the RAMPAGE pipeline output files, ENSEMBL gene ID, and the TSS region coordinates. The cells of the matrix contains the TSS quantification (unit: read counts).

Columns in the header contain the experiment identifier, followed by underscore (“_”) and the bio-replicate number. Values for each replicate are stored in the same cell, separated by colon (“:”).

Example for header column name two bio-replicates:

ExperimentId_1,2, where

“ExperimentId” - is the replicate.experiment.accession id, e.g. ENCSR000AAA

“1” - replicate.biological_replicate_number is 1

“2” - replicate.biological_replicate_number is 2

Expression values format for two bioreplicates:

count1:count2, where

“count1” is the TSS expression level in the ExperimentId, biological replicate 1

“count2” is the TSS expression level in the ExperimentId, biological replicate 2

Example

```
#TSS_id gene_id TSS_coordinates ENCSR000AEI_1,2,
```

```
TSS_ENSG00000134108.12_pk2 ENSG00000134108.12 chr3_5122219_5122269_+ 45.0:74.0
```

Contact

Questions, comments and requests should be addressed to Anna Vlasova (anna.vlasova@crg.eu) or Julien Lagarde (julien.lagarde@crg.eu)