

# 4.1.1 - Gene Expression Profile

## Parameters

Parameter Name	Variable	Default Value	Parameter Range	Description
NB_MOLECULES		5,000,000	>0	number of expressed RNA molecules simulated
EXPRESSION_K	Unknown macro: 'mathinline'	-0.6	Unknown macro: 'mathinline'	exponent of the expression power law ("Pareto coefficient")
EXPRESSION_X0	Unknown macro: 'mathinline'	9,500	Unknown macro: 'mathinline'	parameter of the exponential decay
EXPRESSION_X1	Unknown macro: 'mathinline'	9,500 <sup>2</sup>	Unknown macro: 'mathinline'	parameter of the exponential decay

## Algorithm

**Input:** reference annotation (REF\_FILE), transcript filtering parameter (LOAD\_CODING, LOAD\_NONCODING), expression parameters (NB\_MOLECULES, EXPRESSION\_K, EXPRESSION\_X0, EXPRESSION\_X1)

In the beginning, the Flux Simulator reads the transcripts of the reference annotation (REF\_FILE) and clusters genomic overlapping ones into loci. Transcripts that are annotated as non-/coding can be selectively disregarded (LOAD\_CODING, LOAD\_NONCODING). Then to assign a random

expression profile where not necessarily all transcripts of the reference are expressed. Expression levels

are connected with the relative expression rank form

Unknown macro: 'mathblock'

where

Unknown macro: 'mathinline'

denotes the rank number of a gene,

Unknown macro: 'mathinline'

respectively

Unknown macro: 'mathinline'

Unknown macro: 'mathinline'

which then are turned into relative expression levels by the modified Zipf's Law above, which determines the initial number of molecules by multiplication with the total numbers of molecules. Default values for parameters

Unknown macro: 'mathinline'

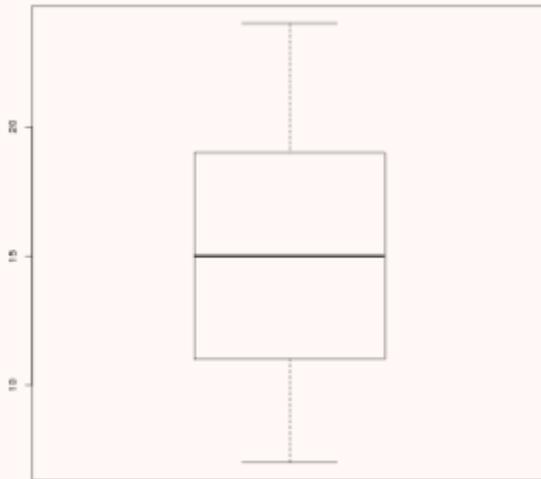
and

Unknown macro: 'mathinline'

have been estimated for mammalian cells by

non-linear fitting to expression levels observed in experimental results.

**Output:** Columnn 1-6 of the PRO\_FILE, i.e., (1) locus name, (2) transcript identifier, (3) coding flag, (4) length of the processed transcript, (5) relative fraction and (6) absolute number of the transcript species in the initial RNA extraction.



Although the Simulator program currently allows to set parameter `EXPRESSION_K` to a value of "0", please note that such settings remove the power-law distribution we usually observe in cellular transcriptomes. Below a boxplot representation of the distribution of the simulated number of molecules for `a//transcripts` in a reference annotation when `EXPRESSION_K` is set to "0", the distribution exhibits a mean of ~15 and a standard deviation of ~5.