.PAR Simulation Parameters

The PAR format in the Flux Simulator is used to administrate all parameters of a run. It is a simple format containing key value pairs (one per line) with the following parameter names (i.e., keys):

File Locations

Key	Type	Default Value	Description
REF_FILE_NAME	String		Path to the GTF reference annotation, either absolute or relative to the location of the parameter file
PRO_FILE_NAME	String	{REF_FILE_NAME}.PRO	Path to the profile of the run, either absolute or relative to the location of the parameter file; the default profile uses the name of the parameter file with the extension .pro.
LIB_FILE_NAME	String	{REF_FILE_NAME}.LIB	Path to the library file of the run, either absolute or relative to the location of the parameter file; the default profile uses the name of the parameter file with the extension .lib.
SEQ_FILE_NAME	String	{REF_FILE_NAME}.BED	Path to the sequencing file of the run, either absolute or relative to the location of the parameter file; the default profile uses the name of the parameter file with the extension .bed.
GEN_DIR	String		Path to the directory with the genomic sequences, i.e., one fasta file per chromosome/scaffold/contig with a file name corresponding to the identifiers of the first column in the GTF annotation.
TMP_DIR	String	$TMP_DIR	Temporary directory, can also be specified by the environment variable $TMP_DIR.

Expression

Key	Type	Default Value	Description
LOAD_CODING	Boolean	YES	Coding messengers, i.e., transcripts that have an annotated CDS, are extracted from the cell.
LOAD_NONCODING	Boolean	YES	Non-coding RNAs, i.e., transcripts without an annotated ORF are extracted from the cell.
NB_MOLECULES	Long	5,000,000	Number of RNA molecules initially in the experiment.
EXPRESSION_K	Double	(-0.6)	Exponent of power-law underlying the expression profile [-1;0]
EXPRESSION_X0	Double	9,500	Linear parameter of the exponential decay.
EXPRESSION_X1	Double	90,250,000	Quadratic parameter of the exponential decay.

Transcript Modifications

Key	Type	Default Value	Description
TSS_MEAN	Double	25	rate of the exponential for deviation of simulated transcription starts from annotated transcription start point, set to NaN (i.e., "not a number") to deactivate simulated transcription start variability
POLYA_SCALE	Double	300	scale parameter of the Weibull distribution describing poly-A tail lengths, set to NaN (i.e., "not a number") to deactivate simulated poly-A tails
POLYA_SHAPE	Double	2	shape paramter of the Weibull distribution describing poly-A tail lengths, set to NaN (i.e., "not a number") to deactivate simulated poly-A tails

Library prepeparation

Fragmentation

Key	Type	Default Value	Description
FRAGMENTATION	Boolean	YES	Turn fragmentation on/off.
FRAG_SUBSTRATE	{DNA, RNA}	RNA* **DNA in Simulator* *v1.2 and earlier*	Substrate of fragmentation, determines the order of fragmentation and reverse transcription (RT): for substrate DNA, fragmentation is carried out after RT, substrate RNA triggers fragmentation before RT.

FRAG_METHOD	{EZ,NB, UR}	UR	Fragmentation method employed: * [EZ] Fragmentation by enzymatic digestion * [NB] Fragmentation by nebulization * [UR] Uniformal random fragmentation
Enzymatic Digestion			
FRAG_EZ_MOTIF	String		Sequence motif caused by selective restriction with an enzyme, choose pre-defined NlaIII, DpnII, or a file with a custom position weight matrix.
Nebulization			
FRAG_NB_LAMBDA	Double	900.0	Threshold on molecule length that cannot be broken by the shearfield of nebulization.
FRAG_NB_THOLD	Double	0.1	Threshold on the fraction of the molecule population; if less molecules break per time unit, convergence to steady state is assumed.
FRAG_NB_M	Double	1.0	Strength of the nebulization shearfield (i.e., rotor speed).
Uniformal Random (UR) Fragmentation			
FRAG_UR_ETA	Double	NaN	Average expected framgent size after fragmentations, i.e., number of breaks per unit length (exhautiveness of fragmentation); NaN optimizes the fragmentation process w.r.t. the size filtering
FRAG_UR_DELTA	Double	NaN	Geometry of molecules in the UR process: * NaN= depends logarithmically on molecule length, * 1= always linear, * 2= always surface-diameter, * 3= volume-diameter, ...
FRAG_UR_D0	Double	1.0	Minimum length of fragments produced by UR fragmentation.

Reverse Transcription (RT)

Key	Type	Default Value	Description
RTRANSCRIPTION	Boolean	YES	Switch on/off Reverse Transcription.
RT_PRIMER	{RH,PDT}	RH	Primers used for first strand synthesis: * [RH] for random hexamers or * [PDT] for poly-dT primers
RT_MIN	Integer	500	Minimum fragment length observed after reverse transcription of full-length transcripts.
RT_MAX	Integer	5,500	Maximum fragment length observed after reverse transcription of full-length transcripts.

Filtering

Key	Type	Default Value	Description
FILTERING	Boolean	NO	Switches size selection on/off.
SIZE_DISTRIBUTION	String	default	Size distribution of fragments after filtering, either specified by the fully qualified path of a file with an empirical distribution where each line represents the length of a read (no ordering required), or attributes of a gaussian distribution (mean and standard deviation) in the form Unknown macro: 'latex-inline', for example Unknown macro: 'latex-inline'. If no size distribution is provided, an empirical Illumina fragment size distribution is employed.

Amplification

Key	Type	Default Value	Description
PCR_DISTRIBUTION	String	default	PCR distribution file, 'default' to use a distribution with 15 rounds and 20 bins, 'none' to disable amplification.
PCR_PROBABILITY	Float	0.1	PCR duplication probability when GC filtering is disabled by setting GC_MEAN to NaN.
GC_MEAN	Float	0.5	Mean value of a gaussian distribution that reflects GC bias amplification probability, set this to 'NaN' to disable GC biases.

GC_SD	Float	0.1	Standard deviation of a gaussian distribution that reflects GC bias amplification probability, inactive if GC_MEAN is set to NaN.

Sequencing

Key	Type	Default Value	Description
READ_NUMBER	Integer	5,000,000	Number of reads.
READ_LENGTH	Integer	36	Length of the reads.
PAIRED_END	Boolean	NO	Switch on/off paired-end reads.
FASTA	Boolean	NO	Creates .fasta/.fastq output. Requires the genome sequences in a folder specified by GEN_DIR. If a quality model is provided by parameter ERR_FILE, a .fastq file is produced. Otherwise read sequences are given as .fasta.
ERR_FILE	String		Path to the file with the error model. With the values '35' or '76', default error models are provided for the corresponding read lengths, otherwise the path to a custom error model file is expected.
UNIQUE_IDS	Boolean	NO	Create unique read identifiers for paired reads. Information about the relative orientation is left out of the read id and encoded in the pairing information. All /1 reads are sense reads, all /2 reads are anti-sense reads. This option is useful if you want to identify paired reads based on the read ids.