

## 5.2 Directional RNA-Seq Protocol (H.sapiens)

### Input

#### Download

[Reference Annotation](#)

[Parameter File](#)

[Reference Genome](#)

#### Parameter

Expression		
NB_MOLECULES	5,000,000	Number of RNA molecules initially in the experiment
TSS_MEAN	50	Average deviation from the annotated transcription start site (TSS)
POLYA_SCALE	NaN	Scale of the Weibull distribution, shifts the average length of poly-A tail sizes
POLYA_SHAPE	NaN	Shape of the Weibull distribution describing poly-A tail sizes
Fragmentation		
FRAG_SUBSTRATE	RNA	Specifies RNA as the substrate of fragmentation
FRAG_METHOD	UR	Uniform random fragmentation
FRAG_UR_ETA	350	Average expected fragment size after fragmentations, i.e., number of breaks per unit length (exhaustiveness of fragmentation)
FRAG_UR_D0	1	Minimum length of fragments produced by UR fragmentation
FRAG_UR_DELTA	NaN	Geometry of molecules in the UR process depends logarithmically on molecule length
Reverse Transcription		
RTRANSCRIPTION	YES	Switch on the reverse transcription
RT_PRIMER	RH	Use random hexamer primers used for first strand synthesis
RT_MOTIF	default	A default PWM of the current Illumina protocol is used
RT_LOSSLESS	YES	Flag to force every molecule to be reversely transcribed
RT_MIN	500	Minimum length observed after reverse transcription of full-length transcripts
RT_MAX	5,500	Maximum length observed after reverse transcription of full-length transcripts
Amplification and Size Segregation		
PCR_DISTRIBUTION	default	Default PCR distribution with 15 rounds and 20 bins
GC_MEAN	NaN	Disable GC biases
PCR_PROBABILITY	0.05	PCR duplication probability
FILTERING	NO	Disable size selection
Sequencing		
READ_NUMBER	150,000,000	Produce 15 million reads
READ_LENGTH	75	Each read sequence is 75nt long
PAIRED_END	YES	Paired-end reads are simulated (two per fragment)

### Output

```
[INFO] I am collecting information on the run.
      initializing profiler *****
[INFO] Checking GTF file
***** OK (00:00:04)
[PROFILING] I am assigning the expression profile
***** OK (00:00:05)
      Reading reference annotation ***** OK (00:00:08)
      found 34102 transcripts
[PROFILING] Parameters
      NB_MOLECULES      5000000
      EXPRESSION_K      -0.6
      EXPRESSION_X0     5.0E7
      EXPRESSION_X1     9500.0
      PRO_FILE_NAME     /Users/micha/Desktop/hg19_stranded.pro
      profiling ***** OK (00:00:00)
      Updating .pro file ***** OK (00:00:00)
      molecules        4999517
[LIBRARY] creating the cDNA library
      Initializing Fragmentation File ***** OK (00:00:07)
      4999517 mol initialized
[LIBRARY] Fragmentation UR
[LIBRARY] Configuration
      DO: 1.0
      Delta: Not specified, depends on sequence length
      Eta: 350.0
      Processing Fragments ***** OK (00:01:32)
      47836171 mol: in 4999517, new 42836654, out 47836171
      avg Len 315.81723, maxLen 996
      preparing transcript sequences ***** OK (00:01:34)
[INFO] Initializing PWM cache
[INFO] Done
[LIBRARY] Reverse Transcription
[LIBRARY] Configuration
      Mode: RH
      PWM: default
      RT MIN: 500
      RT MAX: 5500
      Processing Fragments ***** OK (00:10:16)
      47890332 mol: in 47836171, new 54161, out 47890332
      avg Len 181.1249, maxLen 1148
      start amplification
[INFO] Loading default PCR distribution
[INFO] Initializing PWM cache
[INFO] Done
[LIBRARY] Amplification
[LIBRARY] Configuration
      Rounds: 15
      PCR Probability: 0.05
      Processing Fragments ***** OK (00:02:21)
      Amplification done.
      In: 47890332 Out: 1291507850
      47890332 mol: in 47890332, new 0, out 1291507850
      avg Len 181.17375, maxLen 1127
      Copied results to /Users/micha/Desktop/hg19_stranded.lib
      Updating .pro file ***** OK (00:00:00)
[SEQUENCING] getting the reads
      Initializing Fragment Index
      Indexing ***** OK (00:00:41)
      25701416 lines indexed (1291507850 fragments, 18682 entries)
      sequencing ***** OK (00:43:11)
      1291507850 fragments found (25701416 without PCR duplicates)
      150003190 reads sequenced
      8101165 reads fall in poly-A tail
      54585326 truncated reads
      Moving temporary BED file
      Updating .pro file ***** OK (00:00:00)
      Updating .pro file ***** OK (00:00:00)
      Updating .pro file ***** OK (00:00:00)
      Updating .pro file ***** OK (00:00:00)
[END] I finished, took me 3740 sec.
```

