5.4 Poly-dT Priming and Nebulization (A.thaliana)

In this example, we investigate a protocol that uses poly-*d*T primers to reversely transcribe mRNA molecules, that later-on are fragmented by a mechanical shearing known as *nebulization*. Subsequently, reads are sequenced without PCR or size filtering.

Input

Download

[Reference Annotation](#)

[Parameter File](#)

Parameters

Expression		
NB_MOLECULES	5,000,000	Number of RNA molecules initially in the experiment
TSS_MEAN	100	Average deviation from the annotated transcription start site (TSS)
POLYA_SCALE	200	Scale of the Weibull distribution, shifts the average length of poly-A tail sizes
POLYA_SHAPE	1.5	Shape of the Weibull distribution describing poly-A tail sizes
Reverse Transcription		
RTRANSCRIPTION	YES	Switch on the reverse transcription
RT_PRIMER	PDT	Use poly-dT primers used for first strand synthesis
RT_LOSSLESS	YES	Flag to force every molecule to be reversely transcribed
RT_MIN	400	Minimum length observed after reverse transcription of full-length transcripts
RT_MAX	2,600	Maximum length observed after reverse transcription of full-length transcripts
Fragmentation		
FRAG_SUBSTRATE	DNA	Specifies DNA as the substrate of fragmentation
FRAG_METHOD	NB	Nebulization as fragmentation method
FRAG_NB_LAMBDA	600	Threshold on molecule length that cannot be broken by the shearfield of nebulization
FRAG_NB_M	5	Strength of the nebulization shearfield (i.e., rotor speed)
Amplification and Size Segregation		
PCR_DISTRIBUTION	none	Disable PCR amplification
GC_MEAN	NaN	Disable GC bias
FILTERING	NO	Disable size filtering
Sequencing		
READ_NUMBER	2,000,000	Produce 2 million reads
READ_LENGTH	100	Each read sequence is 100nt long
PAIRED_END	NO	Single reads are simulated, one per fragment

Output

```
[INFO] I am collecting information on the run.
[INFO] Checking GTF file
*[WARN] Unsorted in line 27 - chr/strand Chr1 + already read.
********* OK (371580:28662:09)
[GTF FILE] The GTF reference file given is not sorted, but we found a sorted version.
[GTF FILE] The Simulator will use /Users/micha/Desktop/TAIR9_GFF3_genes_sorted.gtf
[GTF FILE] You might want to update your parameters file
[PROFILING] I am assigning the expression profile
********** OK (371580:28662:09)
    Reading reference annotation **[WARN] merging exon (-21073927,-21073974) with exon (-21073898,-21073924) in
transcript AT1G56280.1 because intervening intron has 4 or less nt.
********[WARN] skipped chromosome ChrM
 OK (00:00:03)
    found 38564 transcripts
[PROFILING] Parameters
    NB_MOLECULES    5000000
    EXPRESSION_K    -0.6
    EXPRESSION_X0    5.0E7
    EXPRESSION_X1    9500.0
    PRO_FILE_NAME    /Users/micha/Desktop/t9_nebulization.pro
    profiling ********** OK (00:00:00)
    Updating .pro file  ********** OK (00:00:00)
    molecules    4999395
[LIBRARY] creating the cDNA libary
    Initializing Fragmentation File ********** OK (00:00:04)
    4999395 mol initialized
[LIBRARY] Reverse Transcription
[LIBRARY] Configuration
        Mode: PDT
        PWM: No
        RT MIN: 400
        RT MAX: 2600
    Processing Fragments ********** OK (00:00:18)
        4999395 mol: in 4999395, new 0, out 4999395
        avg Len 1039.7405, maxLen 2600
[LIBRARY] Nebulization
[LIBRARY] Configuration
        Lambda: 600.0
        M: 5.0
        Max Length: 2600.0
        Recursions: 3
    Processing Fragments ********** OK (00:00:23)
        7498699 mol: in 4999395, new 2499304, out 7498699
        avg Len 693.1967, maxLen 2590
        start amplification
[LIBRARY] PCR disabled, skipping amplification
    Copied results to /Users/micha/Desktop/t9_nebulization.lib
    Updating .pro file  ********** OK (00:00:00)
[SEQUENCING] getting the reads
    Initializing Fragment Index
    Indexing ********** OK (00:00:09)
    7498699 lines indexed (7498699 fragments, 18849 entries)
    sequencing ***[WARN] merging exon (-21073927,-21073974) with exon (-21073898,-21073924) in transcript
AT1G56280.1 because intervening intron has 4 or less nt.
*******[WARN] skipped chromosome ChrM
 OK (00:08:37)
    7498699 fragments found (7498699 without PCR duplicates)
    2001148 reads sequenced
    0 reads fall in poly-A tail
    42854 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 579 sec.
```