4.2 - Reverse Transcription

Parameters

Parameter Name	Default Value	Parameter Range	Description
RTRANSCRIPTION	true	{true,false}	switch reverse transcription on/off
RT_PRIMER	RH	{PDT,RH}	chooses random (RH) or poly-dT primers for first strand synthesis
RT_MOTIF			sequence motif that
RT_MIN	500	>0	the minimum stretch that is polymerized by the reverse transcriptase enzyme
RT_MAX	5,500	>0	the maximum stretch that is polymerized by the reverse transcriptase enzyme (template-fidelity)

Algorithm

Input: RNA polymers annotated by their start end end coordinates on the transcript sequence they originate from (LIB_FILE) and parameters for reverse transcription.
Current sequencing technologies have to transform RNA into double-stranded DNA molecules before sequencing, either before or after fragmentation. For the first strand synthesis the Flux Simulator provides random priming or poly-dT primers (RT_PRIMER). According to the nature of primers and the template fidelity of the reverse transcriptase enzyme (RT_MIN and RT_MAX), the algorithm determines start point and extension separately for first and for second strand synthesis.

1. During first strand synthesis, poly-*d*T primers induce priming events in the poly-A tail, whereas random primers provoke successful initiation events along the entire molecule, and anchored primers trigger exactly one priming event at the 3-end of the respective fragment. In sequencing protocols without sequence-specific biases, each priming event is assigned a random location uniformly sampled along the corresponding stretch. Optionally, start points of first strand synthesis are determined by importance sampling according to weights of an optional PWM capturing sequence-biases.
2. The point where second strand synthesis initiates is simulated by the length of the first DNA strand, which can be between RT_MIN and RT_MAX nucleotides, but maximally the distance of the first strand synthesis priming event from the 5'-end of the RNA template. The point of priming second strand synthesis in the presence of sequence biases is drawn from a distribution according to the PWM capturing the bias, or from a uniform distribution otherwise.

In the case of multiple priming events with random primers, several enzymes concurrently transcribe parts of the RNA molecule, and collisions with downstream DNA-RNA hybrids are resolved by displacement according to a Bernoulli trial.
Output: Reversely transcribed DNA molecules annotated by their start end end coordinates on the transcript sequence they originate from (LIB_FILE).