4.5.1 - The Sequencing Process

***Requires**:* LIB_FILE_NAME, READ_LENGTH, READ_NUMBER, PAIRED_END, FASTA, GEN_DIR, ERR_FILE, UNIQ_IDS
***Outputs**:* SEQ_FILE_NAME

This step produces about READ_NUMBER sequencing reads from the library in LIB_FILE_NAME. The simulator iterates the input annotation and maps the READ_LENGTH long stretches from the ends of cDNA molecules in the library LIB_FILE_NAME to genomic coordinates. In the case of READ_LENGTH exceeds the length of a cDNA molecule, the read is truncated to the length of the molecule. For each fragment a Bernoulli trial is carried out, by Unknown macro: 'mathinline' with Unknown macro: 'mathinline' being a uniformly sampled random variable in the boundaries [0;1[compared to the sequencing probability Unknown macro: 'mathinline'

Unknown macro: 'mathblock'

and Unknown macro: 'mathinline' denoting the number of molecules in the library. By this, never more reads (respectively, read pairs) are generated than there are LIBRARY_NUMBER cDNA molecules. In the case of single end sequencing, randomly one end of the cDNA molecule that succeeded the Bernoulli trial is sequenced, and if PAIRED_END is set, correspondingly both ends are sequenced. The FLUX SIMULATOR shows you the number of reads and their fraction (relative to the planned number), the number of splicing loci represented in these reads (and the ratio they constitute of the total number of expressed loci), and the number of transcripts (ratio of total expressed spliceforms, respectively).

Please note that the final number of molecules you obtain provides an upper limit on your sequencing capacity, as over sampling a small amount of molecules will not enlarge the diversity in the produced reads — it means, if you would produce a 1000 reads from 10 molecules left after RT /fragmentation, you will find groups of about 100 that map to identical locations. Upon termination the step copies the .LIB file from the temporary directory to the project directory and updates column 7 and 8 of the .PRO file.