FAST5 (raw) = targeted amplicon reads 2-4 kb

a. Quality-based filtering: 2D reads (Metrichor)

FASTQ

b. Remove reads < 1000bp (or 2000bp) and convert to FASTA format

FASTA
c. Reads alignment to human genome GRCh37 (LAST)

MAF
d. Parse LAST output and select reads with an insert (Java)
   - Insert with length >= 50bp
   - Two flanking regions mapping to the Switch (length >= 100bp)
   - Gap <= 100bp between the insert and the Switch flanking region(s)

FASTA BED
e. Merge inserts' coordinates (for overlapping inserts) and annotate them with GENCODE v19
   (bedtools, BEDOPS)