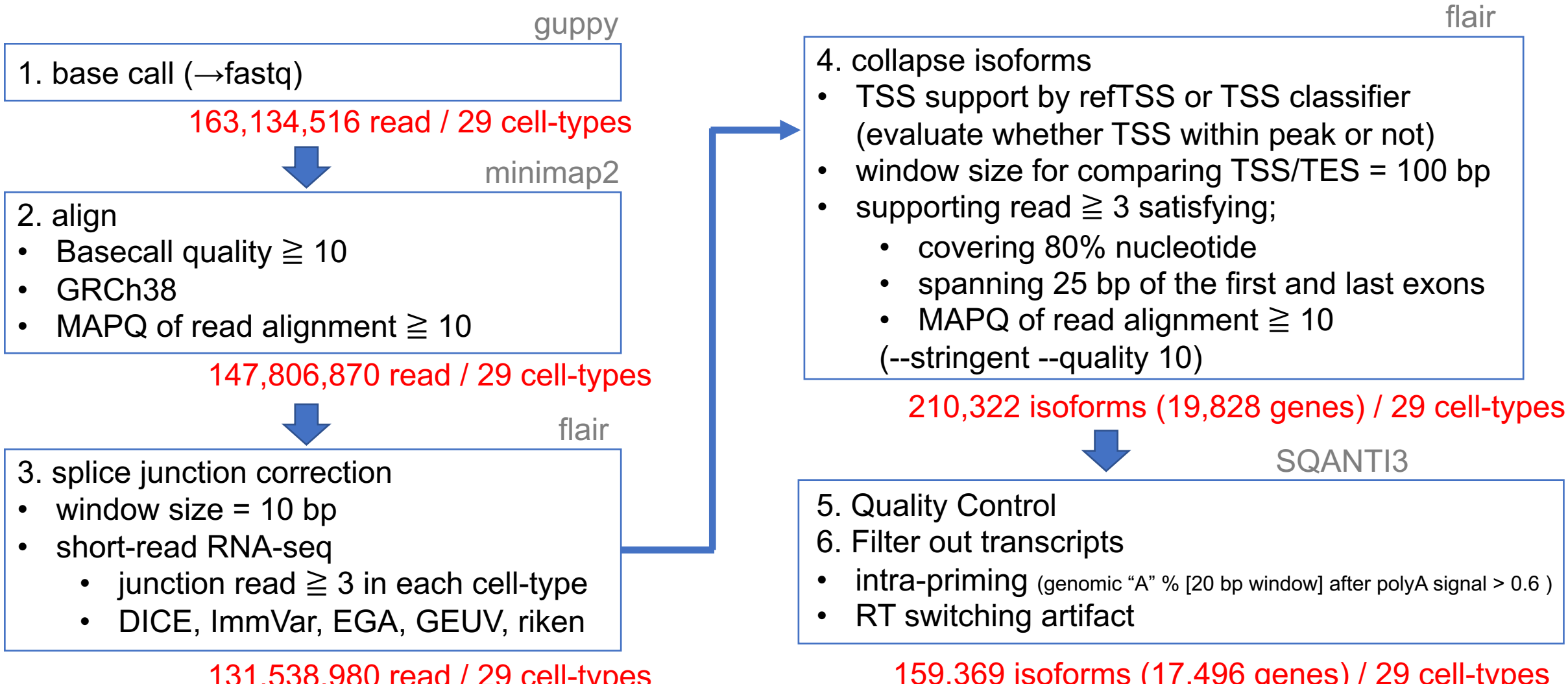
guppy
1. base call (→fastq)
163,134,516 read / 29 cell-types
minimap2
2. align
• Basecall quality ≧ 10
• GRCh38
• MAPQ of read alignment ≧ 10
147,806,870 read / 29 cell-types
flair
3. splice junction correction
• window size = 10 bp
• short-read RNA-seq
• junction read ≧ 3 in each cell-type
• DICE, ImmVar, EGA, GEUV, riken
131,538,980 read / 29 cell-types
flair
4. collapse isoforms
• TSS support by refTSS or TSS classifier (evaluate whether TSS within peak or not)
• window size for comparing TSS/TES = 100 bp
• supporting read ≧ 3 satisfying;
• covering 80% nucleotide
• spanning 25 bp of the first and last exons
• MAPQ of read alignment ≧ 10
(--stringent --quality 10)
210,322 isoforms (19,828 genes) / 29 cell-types
SQANTI3
5. Quality Control
6. Filter out transcripts
• intra-priming (genomic "A" % [20 bp window] after polyA signal > 0.6 )
• RT switching artifact
159,369 isoforms (17,496 genes) / 29 cell-types