

In this document we count the number of reads and barcodes we use on different steps on Novel-X pipeline. With bam-files we used bxttools stats command.

CHM1

Initial:

Number of reads: 591130730

Number of barcodes: 2936374

Unmapped:

Number of reads: 11897752 (2%)

Number of barcodes: 1369871 (46.6%)

Mapped back to contigs:

Number of reads: 747944 (0.1%)

Number of barcodes: 172468 (5.9%)

Reassembly stage:

Total number of reads: 240758843 (40.5%) for 9448 assemblies

CHM13

Initial:

Number of reads: 591217596

Number of barcodes: 2870772

Unmapped:

Number of reads: 9142933 (1.5%)

Number of barcodes: 1254670 (43.7%)

Mapped back to contigs:

Number of reads: 770110 (0.1%)

Number of barcodes: 165324 (5.8%)

Reassembly stage:

Total number of reads: 252719350 (42.7%) for 9546 assemblies

NA12787

Initial:

Number of reads: 866774619

Number of barcodes: 3918208

Unmapped:

Number of reads: 27144858 (3.1%)

Number of barcodes: 2456850 (62.7%)

Mapped back to contigs:

Number of reads: 1019408 (0.1%)

Number of barcodes: 411315 (10.5%)

Reassembly stage:

Total number of reads: 377696832 (43.5%) for 11969 assemblies

NA19240

Initial:

Number of reads: 1002309662

Number of barcodes: 4155776

Unmapped:

Number of reads: 23749318 (2.4%)

Number of barcodes: 1826009 (43.9%)

Mapped back to contigs:

Number of reads: 1058405 (0.1%)

Number of barcodes: 212653 (5.1%)

Reassembly stage:

Total number of reads: 561705348 (56.0%) for 11832 assemblies