Shared Resource Updates

The Illumina HiSeq 2000 sequencer has undergone major changes since its launch in February 2010. This includes larger flow cells and new image analysis software to bring sequencing output to 1.6 billion (200 million reads per lane, 8 lanes per flow cell) passed filter reads. In our first HiSeq 2000 run, some of our samples achieved 240 million reads per lane or 1.9 billion reads per flow cell. The new sequencing chemistry (version 3) significantly improves coverage uniformity by reducing density-dependent GC bias and by increasing in cluster density, thereby resulting in the lowest number of gaps and minimal risk of missing variants in sequencing data. The Illumina primary alignment software algorithm ELAND_v2e significantly improved SNP, indel and structural variant calling and increase aligned reads by 5 percent to 7 percent. The amount of mismatch-free reads per run has also increased threefold since 2010. Illumina achieved major breakthroughs in data processing software (RTA 1.12), resulting in a significant reduction in the disk space required to store the sequencing output. For example, the amount of space needed to store HiSeq 2000 data from a human genome sequenced to a depth of 30-fold has dropped by an order of magnitude. Despite the huge amount of data output, network bandwidth needed for data transfer from the HiSeq 2000 instrument to our computation server is minimal (5.56 out of typical 100 mbs sustained capacity over a 1 gbs network).

The increase in sequencing output and the reduction in sequencing time enabled by the Illumina HiSeq 2000 Sequencer allows researchers to sequence more samples, increasing the statistical power and quickening the time to data while staying within their research budgets. For example, with the current data output, 125 indexed samples can be sequenced per run (two concurrent HiSeq flow cells at about 25 million reads per sample) for high throughput gene expression profiling study. If the intent is to perform high resolution transcriptome analysis, more than 50 samples can be analyzed per run at about 100 million reads per sample. The run time for both types of runs is around two days.

Illumina HiSeq 2000 Sequencer

Illumina cBot (creates clonal clusters from sample fragments for sequencing)