DRAGEN is the world’s first next-generation sequencing (NGS) Bio-IT processor and radically reduces the computational cost and increases the speed without compromising the accuracy of the NGS data analysis pipeline.

The reference-based mapping, aligning, sorting, deduplication and variant calling that is currently performed by software on extensive cluster or cloud-based platforms is performed by the DRAGEN processor, which is loaded with highly optimized algorithms for this analysis. DRAGEN can accurately analyze over 50 whole human genomes (from FASTQ to VCF) in less than a day, and reduces the need for clusters of large servers to process the data, lowering costs related to storage space and IT infrastructure.

The DRAGEN Bio-IT Processor is integrated on a PCIe form-factor card and is provided with accompanying software as a Platform-as-a-Service (PaaS) that can be integrated directly into sequencing instruments and next-generation sequencing bioinformatics servers.

DRAGEN takes raw read data produced by a sequencing instrument, such as Illumina HiSeq X Ten, in standard FASTQ format (paired-end or single-end reads) as an input. DRAGEN’s output, after variant calling, is a standard VCF file ready for tertiary analysis.

Benefits

- Low cost
- Very fast
- High accuracy
- Very low power
- Embeddable form factor
- Local analysis
- Massively scalable
- Reduced storage
- Push button workflow
- No IT staff needed

Availability

- Early Access: Jun 2014
- Commercial: Oct 2014

Pricing

- Available on request
Earlier this year Illumina announced their HiSeq X Ten, which is a cluster of 10 HiSeq X instruments capable of sequencing up to 18,000 whole human genomes each year with continuous operation.

The HiSeq X Ten has a run cycle of ~3 days and produces ~150 genomes each run cycle. This means that the data generated must be analyzed within 3 days or a backlog will occur. Simple math thus provides a target of ~28 minutes for the completion of the mapping, aligning, sorting, deduplication and variant calling of each genome.

Running the industry standard BWA+GATK analysis pipeline to perform this analysis on a reasonably high-end (Dual Intel Xeon E5-2697v2 CPU – 12 core, 2.7 GHz with 96 GB DRAM) compute server takes ~24 hours per genome. To achieve the required throughput of 150 genomes every three days, at least 50 of these servers are required.

With DRAGEN’s run time for the full pipeline being well under 28 minutes, only a single card is required to handle the data produced by a full HiSeq X Ten system running at maximum capacity!

When calculating the costs associated with the legacy compute server infrastructure required to analyze and store the data produced by the HiSeq X Ten, we use the same method that Illumina themselves use to calculate their “$1,000 Genome.” Equipment costs are amortized over 4 years and staff and overheads are included.

By taking into account standard pricing from Dell for the 50 compute servers, industry standard costs for associated IT professionals, rack space and power, as well as AWS pricing for upload and storage of the genomic data over a 4 year period, the $1,000 Genome becomes a $1,130 Genome.

By comparison, DRAGEN significantly reduces the compute cost over the same period while also removing the need for IT professionals and lowering power and rack space requirements. In addition, since DRAGEN implements CRAM-based compression within its pipeline to transparently compress the large amounts of data to be uploaded and stored for later re-analysis, these costs are also significantly lower. When combining all of these savings over the 72,000 genomes sequenced during the 4 year period, DRAGEN provides cost savings of ~$6 million!

Case Study: The $1,000 Genome on Illumina’s HiSeq X Ten