

Post-Light Sequencing™ Using the Ion Torrent Personal Genome Machine (PGM)



Ion Torrent, 246 Goose Lane, Suite 100, Guilford, CT 06437 & 7000 Shoreline Court, Suite 201, South San Francisco, CA 94080, USA

ABSTRACT

Ion Torrent Systems has developed a DNA sequencing system that directly translates chemical signals into digital information on a semiconductor chip. This approach leverages a trillion dollars of investment from the semiconductor industry taking advantage of existing state-of-the-art chip fabrication technology, and the entire semiconductor design and supply chain. Unprecedented scalability and cost reduction result from decades of Moore's Law advances in semiconductor technologies that are brought to bear for DNA sequencing. Ion Torrent sequencing takes place in semiconductor microchips that contain detectors which have been fabricated as individual electronic sensors, allowing one sequence read per sensor. Current configurations have 1.5 million sensors in a 1 cm² chip, with proof of principle to enable densities of over 100 million sensors per chip.

The sequencing chemistry itself is remarkably simple. Native nucleotides are incorporated into the growing strand by DNA polymerase. As a base is incorporated, a direct electrical measurement of the incorporation event is made and the sequence is read out directly into the digital domain. Thus, sequencing is direct, efficient, and massively parallel, requiring no specialized reagents and no optical systems. Using native DNA chemistry with real time detection enables run times to be about 2 hours.

The Ion Torrent Personal Genome Machine provides a powerful tool for driving research, and is now commercially available. The simplicity of a semiconductor chip that reads itself means that small footprint instruments will be available at a fraction of the cost of other next-generation systems. The use of standard reagents, low reaction volume and high data density keep reagent costs low; Computational infrastructure and staff support requirements are modest; Finally, the short run time supports fast research cycle times and promotes the use of high throughput DNA sequencing in everyday research.

Sequencing Technology

The Ion Personal Genome Machine (PGM™) sequencer sequentially floods the chip with one nucleotide after another. If a nucleotide is incorporated into a strand of DNA, a hydrogen ion is released. The charge from that ion changes the pH of the solution, and is detected by the ion sensor, essentially going directly from chemical information to digital information. If the next nucleotide that floods the chip is not a match, no voltage change is recorded and no base is called. If there are two identical bases on the DNA strand, the voltage is double, and the chip records two identical bases called. Because this is direct detection—no scanning, no cameras, no light—each nucleotide incorporation is recorded in seconds.

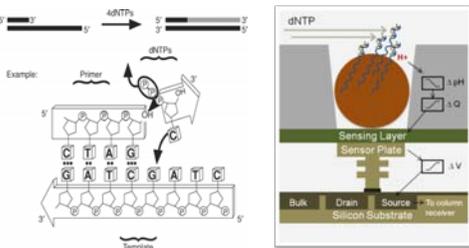


Figure 1: Sequencing Technology.

Scalability

Ion Chips consist of a high-density array of micro-machined wells. Beneath the wells is an ion-sensitive layer and beneath that a proprietary ion sensor. No cameras, no light—each nucleotide incorporation is recorded in seconds. Ion 314 Chip providing 1.3 million wells is currently available. Shortly the Ion 316 Chip will be available and provide a >5X increase in scaling to 6.3 million wells. Moore's law provides the precedent for the scalability of Ion's post-light sequencing technology, with unprecedented increases in throughput possible utilizing existing CMOS fabrication technologies. Sensors can be smaller and chips can be made a larger sizes to increase the number of features per chip.

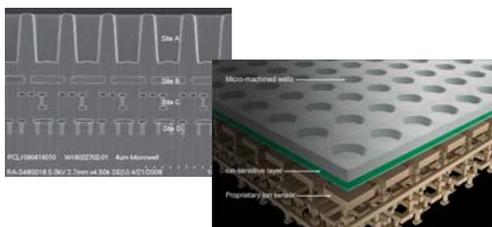


Figure 2: Scalability of Ion Chips

Workflow

Ion Sequencing workflow consists of a number of standard elements; creation of a fragment DNA library with a well defined size range or amplicon pool, a simple emulsion PCR process creating amplified templates on proprietary Ion Sphere particles, followed by an approximately 2 hour sequencing run on the PGM platform. Ion Torrent is working on improvements to the emulsion PCR process to make it simpler and with less manual intervention

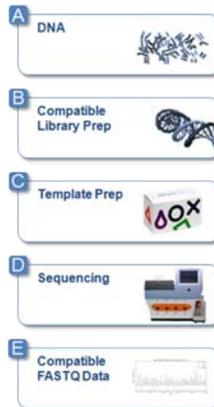


Figure 3: Workflow

Analysis Pipeline

With the completion of the PGM run, data automatically transfers to the dedicated Torrent Server that runs Torrent Suite software. Here raw ion signals are converted to base calls and stored in industry standard SFF or FASTQ files. These data can be processed by a variety of software packages. These packages help identify sequence variants of interest or differentially expressed mRNAs.

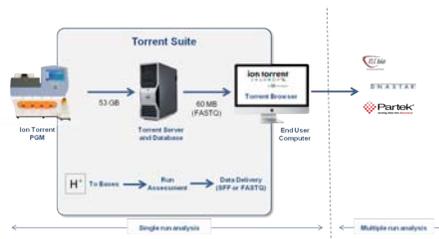


Figure 4: Workflow

Quality Assessment

With the completion of base calls, data on Torrent Server automatically flows through the open source TMAP algorithm. TMAP aligns the reads to a supplied reference genome. This alignment gives the researcher an assessment of quality of both the sequencing run and the input library. Scientists can then view details about the number and length of reads within a PGM run, as well as percentage coverage across the reference genome.

Web-based Data Delivery

Torrent Browser (hosted on Torrent Server) provides an embedded web portal to allow users remote, web-enabled data access. View analysis reports, export data graphics, download summarized data, or share a link with a colleague. File sizes are small enough to download remotely over the web. Torrent Browser also provides a searchable, sortable, centralized data repository for all the PGM runs and analyses. Users can even archive raw data on the server right from within the Torrent Browser web interface.



Figure 5: Torrent Browser - the web interface to download PGM data.

Accuracy

The Ion Torrent Personal Genome Machine rapidly generates sequencing data while maintaining highly accurate base calls. Each run on an Ion 314 chip currently generates greater than 100K high quality, filtered reads, producing a minimum of 10Mb of high quality sequence. Figure 6 shows rapidly improving per base accuracy versus total megabases of sequence. The data was generated from a single run of 314 using *E. coli* DH10B and base calls were generated with default parameters in v1.2.0 of the Torrent Suite software. High quality base calls allow for lower levels of coverage to detect variants, with fewer false positive calls. The level of accuracy of Ion Torrent semiconductor sequencing provides the basis for reproducible and consistently high data quality and minimizes the overall cost of sequencing projects.

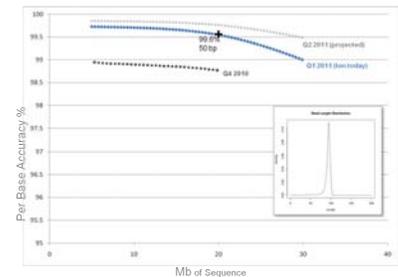


Figure 6: Rapidly improving per base accuracy

Read Length

Figure 7 shows the improvement in read length over time at Q17, Q20, and perfect reads. The long read lengths generated by the PGM and the 314 Chip enable a variety of applications including microbial genomes, library assessment and quality control as well as targeted resequencing.

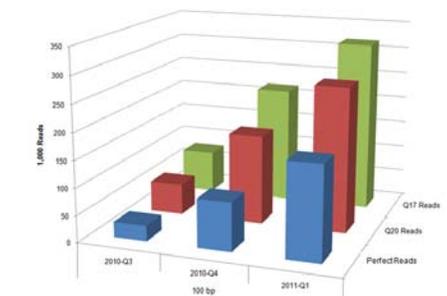


Figure 7: Rapid performance improvements

Lack of Systematic Error

Ion Torrent sequencing offers truly enabling sequencing through high consensus accuracy and lack of systematic error. The data for Figure 8 is generated from a single run of 314 using *E. coli* DH10B and the base calls are generated with default parameters in v1.2.0 of the Torrent Suite software. The circular plot in the left panel of figure 8 shows percentage of G+C content in the inner red circle and regional coverage level in the outer blue circle. Despite differences in regional levels of G+C content, Ion Torrent produces exceptionally even coverage across the genome. The right panel of figure 8 shows a histogram of expected versus actual coverage. The green line (Poisson) shows what coverage would be if sequencing occurs in a truly unbiased manner. The red line represents the actual sampling of the Ion Torrent reads. It tightly matches the green line illustrating exceptionally low bias from the Ion Torrent platform.

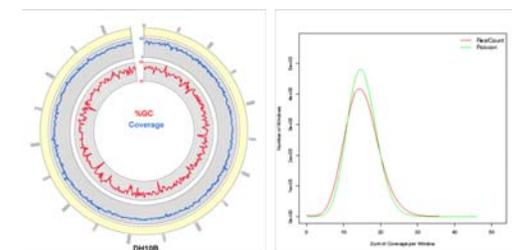


Figure 9: Exceptionally Uniform Genome Coverage

Trademarks/Licensing
© 2011 Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners. For Research Use Only. Not intended for any animal or human therapeutic or diagnostic use.