

Ion Torrent Amplicon Sequencing

Introduction

The ability to sequence a genome or a portion of a genome has enabled researchers to begin to understand how the genetic code governs cellular function. Similarly, sequence data is becoming increasingly useful in understanding human health at the molecular level. In many instances sequencing only a portion of genome, or a “targeted” region of the genome, provides the data necessary to fully address a given hypothesis. A discrete region of the genome can be easily amplified from the entire genome using PCR and thus when

this method is used for targeted sequencing it is often referred to as “amplicon sequencing”.

Depending on the hypothesis, amplicon sequencing is typically used to investigate a few to several hundred genomic regions across multiple samples. Amplicon sequencing is especially useful for clinical applications where understanding human health is the goal. Traditional Sanger sequencing does not scale with the number of regions and samples. Next Generation Sequencing (NGS) technologies have dramatically increased throughput in the last five years. However, the presently available NGS technologies are too cumbersome, too slow or too expensive to sequence smaller regions of the genome. Ion Torrent sequencing technology is uniquely suited for Amplicon Sequencing because this revolutionary technology is simple, fast, scalable and cost effective.

This application note describes how to perform amplicon sequencing using the Ion Torrent technology. It will also offer guidance to design experiments to detect either germ-line or somatic mutations.

Ion Torrent Technology

Ion Torrent has invented the first commercially available device—a new semiconductor chip—capable of directly translating chemical signals into digital information. The first application of this technology is sequencing DNA. The device leverages decades of semiconductor technology advances, and in just a few years has brought the entire design, fabrication and supply chain infrastructure of that industry—a trillion dollar investment—to bear on the

challenge of sequencing. The result is Ion semiconductor sequencing, the first commercial sequencing technology that does not use light, and as a result delivers unprecedented speed, scalability and low cost. The technology will scale in just a few months from ~1 million sensors in the first-generation Ion 314™ chips to ~7 million sensors in the second-generation Ion 316™ chips—all while maintaining the same one to two hour runtime (Fig 1).

The sequencing chemistry itself is remarkably simple. Naturally, a proton is released when a nucleotide is incorporated by the polymerase in the DNA molecule, resulting in a detectable local change of pH. Each micro-well of

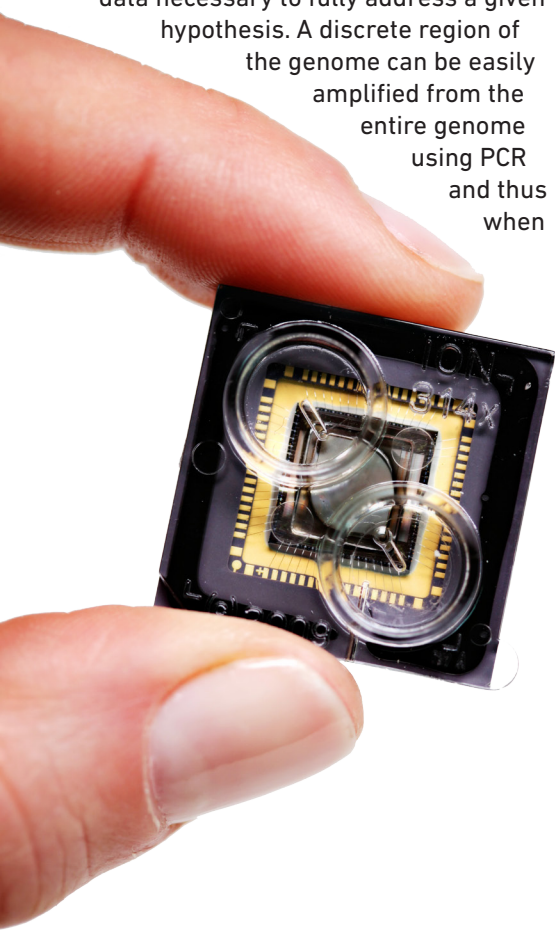


Figure 1. Ion 314™ chip. The grey elliptical shape at the center contains 1.4 million wells that sense and record the sequencing reaction.

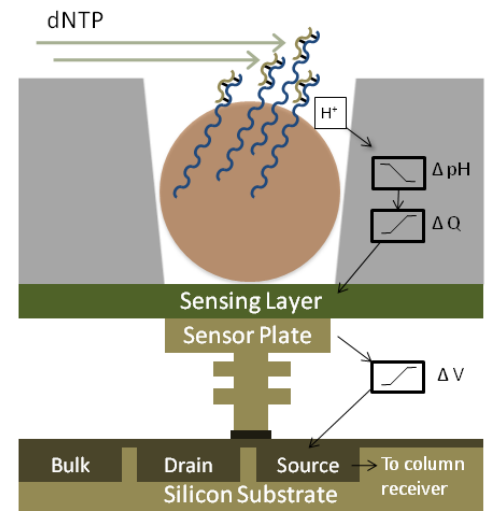


Figure 2. Schematic cross-section of a single well of an Ion Torrent sequencing chip. The well houses Ion Sphere™ particles containing DNA template. When a nucleotide incorporates, a proton releases and the pH of the well changes. A sensing layer detects the change in pH and translates the chemical signal to a digital signal.

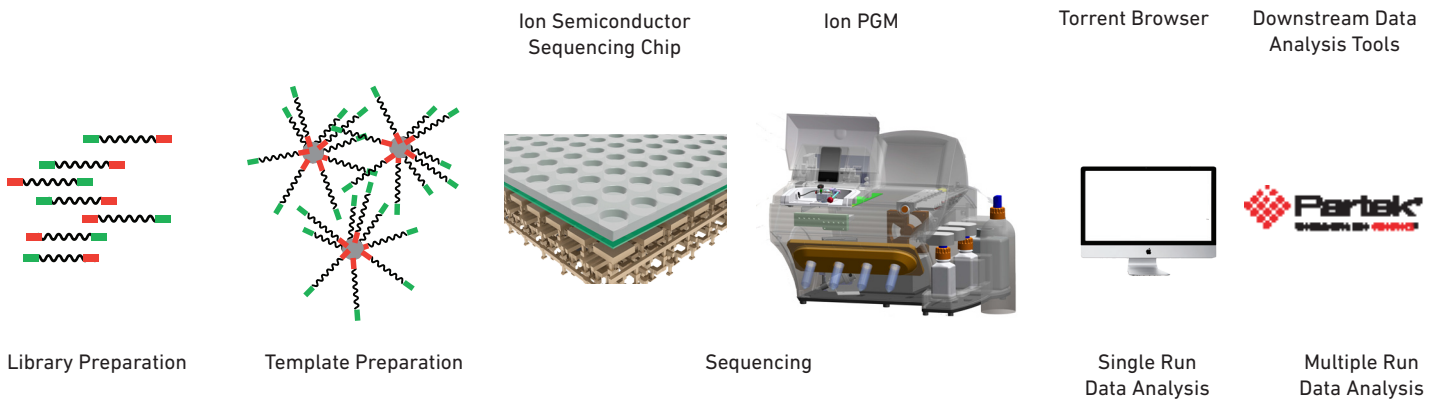


Figure 3. Schematic representation of the Ion Torrent sequencing workflow. A sequencing library is produced by generating DNA fragments flanked by the Ion Torrent sequencing adapters. These fragments are clonally amplified on the Ion Sphere™ particles by emulsion PCR. The Ion Sphere™ particles with the amplified template are then applied to the Ion Torrent chip and the chip is placed on the Ion PGM™. The sequencing run is set up on the Ion PGM™. Sequencing results are provided in standard file formats. Downstream data analysis can be performed using the DNA-Seq workflow of the Partek® Genomics Suite™.

the Ion Torrent semiconductor sequencing chip contains approximately one million copies of a DNA molecule. The Ion Personal Genome Machine (PGM™) sequencer sequentially floods the chip with one nucleotide after another. If a nucleotide complements the sequence of the DNA molecule in a particular micro-well, it will be incorporated and hydrogen ions are released. The pH of the solution changes in that well and is detected by the ion sensor, essentially going directly from chemical information to digital information (Fig 2). If there are two identical bases on the DNA strand, the voltage is double, and the chip records two identical bases. If the next nucleotide that floods the chip is not a match, no voltage change is recorded and no base is called. Because this is direct detection—no scanning, no cameras, no light—each nucleotide incorporation is measured in seconds enabling very short run times.

Workflow

The Ion Torrent Personal Genome Machine™ requires a very simple workflow (Fig 3). The first step in the workflow is to generate a library of DNA fragments flanked by the Ion Torrent adapters. This can be done by ligating the adapters to the PCR products or by adding the adapter sequences during PCR by designing PCR primers with the Ion adapter sequences at the 5' end (Fig 4).

The library fragments are then clonally amplified onto the proprietary Ion Sphere™ particles. Clonal amplification is accomplished by emulsion PCR (emPCR). The Ion Sphere™ particles coated with template are applied to the Ion chip. The Ion Sphere™ particles are then deposited in the chip wells by a short centrifugation step. The chip is placed on the PGM and the PGM touch-screen guides the user to set up the sequencing run.

Ion Fragment Library Kit User Guide, Ion

Template 314 Kit User Guide, and the Ion Sequencing 314 Kit User Guide contain detailed descriptions of the protocols.

Once data is generated on the Ion PGM™ sequencer, it is automatically transferred to the required Torrent Server. Here data are run through signal processing and base calling algorithms that produce the DNA sequences associated with individual reads. Torrent Server hosts web pages where summarized data results can be viewed and the data themselves can be downloaded using industry-standard data formats like SFF, FASTQ, or SAM/BAM. The Ion Torrent data can then be imported into any number of NGS data analysis solutions. The last section of this document describes one such solution – Partek® Genomics Suite™.

Experimental Design Considerations

There are several points to consider in order to obtain optimal amplicon

Forward primer (Primer A-key):

5' -CCATCTCATCCCTGCGTGTCTCCGACTCAG-template-specific-sequence-3'
 <----- 30 bases ----->

Reverse primer (Primer P1-key):

5' -CCTCTCTATGGGCAGTCGGTGAT-template-specific-sequence-3'
 <----- 23 bases ----->

Figure 4. PCR primer sequence structure required to generate the amplicon sequencing library. Standard software tools to aid PCR primer design such as Primer3 (<http://frodo.wi.mit.edu/primer3/>) should be used when choosing template specific sequences.

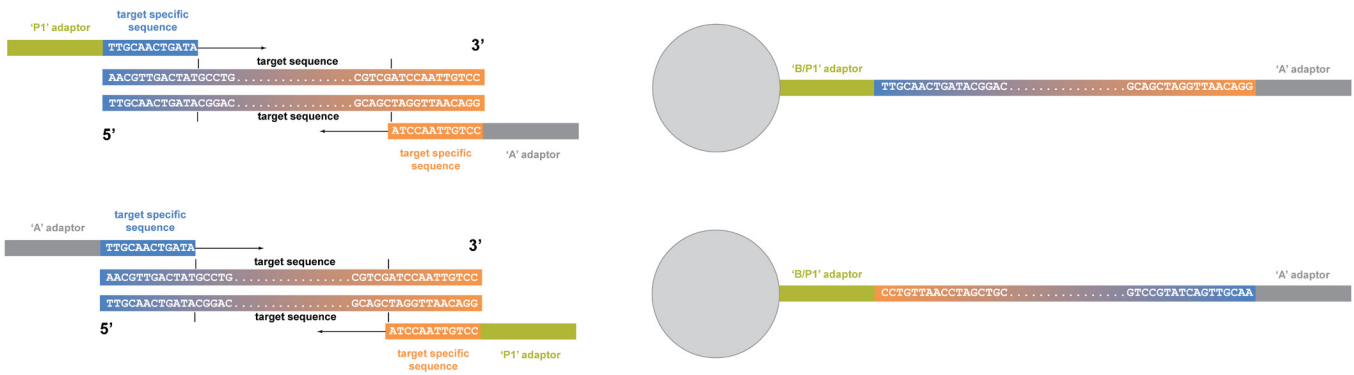


Figure 5. Schematic representation of how the amplicon libraries and the resulting templates are generated for bidirectional and unidirectional sequencing via appropriate selection of primers and target-specific amplification of genomic DNA prior to emulsion PCR. For bidirectional sequencing the two templates need to be generated. Only one is required for unidirectional sequencing.

sequencing results.

- Bidirectional/unidirectional sequencing
- Length of the target region
- Depth of coverage

Bidirectional sequencing is recommended for optimal results. Bidirectional sequencing will produce high quality reads from both ends and across the full length of the amplicons. The bidirectional sequencing approach requires four fusion primers per target region—each of the two adapter sequences must be fused to each of the two target-specific sequences. One of the two fusion primer pairs will have the A adaptor region followed by the proximal end of the target sequence and the P1 adaptor region followed by the distal end of the target region. The other fusion primer pair will have the A and P1 adaptor sequences swapped (Fig 5). Unidirectional sequencing requires only one fusion primer pair and will produce reads from only one end of the amplicon.

The length of the target regions is important for obtaining optimal results and must be considered carefully. The current typical read length is 100 nucleotides. However, the first 20-25 nucleotides of sequence correspond to the target specific sequence of the PCR primers and they will not produce informative data. Hence, we currently recommend target regions (the region of the amplicon that will produce informative sequence) of around 75

nucleotides long. The technology is rapidly evolving and will open the door to study larger amplicons as read-length increases. Longer amplicons (up to 150 nucleotides) can also be considered. However, reads from both directions (see above: bidirectional sequencing) will not overlap.

The depth of coverage needed to ensure enough statistical confidence to make accurate mutation calls will depend on the expected frequency of the mutation within the sample. The depth of coverage necessary for the experiment will

dictate the number of amplicons that can be included given a fixed amount of sequence throughput per chip. Two different use cases should be considered to guide on the depth of coverage required and the number of amplicons that can be sequenced on a single sequencing run.

1. For analysis of germ-line mutations that follow standard Mendelian inheritance patterns, it is expected that either 100% or 50% of the reads contain a given sequence variant. This would be the case for inherited diseases such as

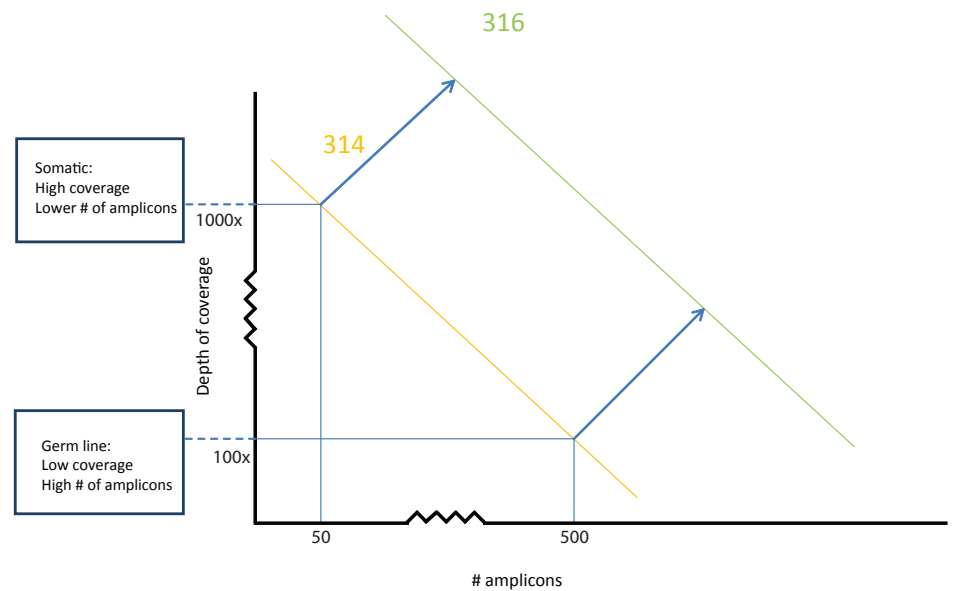


Figure 6. Representation of the correlation between depth of coverage and number of amplicons for two use cases of amplicon sequencing. For germ line mutation detection, moderate depth of coverage is required. For somatic mutations, higher depth of coverage is required and as a consequence, fewer amplicons can be sequenced in the same run. The Ion 316™ chip allows multiple samples per chip or an increased number of amplicons per experiment.

Cystic Fibrosis (CF), familial cancers, mental retardation, developmental diseases, etc. Homogeneous cell lines also produce all or half of the reads containing a given sequence variant. In these cases, an average depth of coverage of 100-200X provides a sufficient number of reads to detect variants with statistical confidence. At that depth of coverage the Ion 314™ chip produces enough reads to sequence up to 500 amplicons bidirectionally (Fig 6). The 316™ chip will offer 10 times higher throughput than the Ion 314™ chip and thus enables sequencing more amplicons or multiple samples per chip. Sample multiplexing is enabled by employing molecular barcodes, which are a set of unique sequences (typically 6-10 nucleotides in length) that will be part of each read and can be used to identify all reads associated with a given sample. Ion Torrent is currently developing molecular barcodes that are

compatible with the Ion Torrent sequencing platform.

2. For high confidence detection of somatic mutations present at variable and typically low frequencies in heterogeneous cancer samples, deeper coverage of up to 1000-2000X is required. At this depth of coverage the Ion 314™ chip will produce enough reads to sequence up to 50 amplicons bidirectionally (Fig 6). The higher throughput offered by the Ion 316™ chip enables sequencing an even greater number of amplicons or detecting variants present at even lower frequencies by increasing the depth of coverage.

Identification of DNA Variants

Once base call data have been generated and aligned against a reference DNA sequence on the Torrent Server, then

SAM (sequence alignment format) files can be directly imported into Partek® Genomics Suite™ (GS). Pre-defined workflows are available within Partek® GS to perform a variety of analyses. Within the DNA-seq workflow, a step by step process is available to perform a multi-sample analysis of Ion Torrent amplicon sequencing data. The process starts in a logical manner by stepping the user through import of the SAM files (one for each PGM run) into Partek® GS. Data is grouped based on replicate structure and whether biological or technical replicates were used. Because differences in coverage can occur across even a short amplicon, it is recommended that normal, wild type samples are run and analyzed in parallel to experimental samples. Partek® GS implements a statistical test at each base comparing the base pair composition in the experimental sample against the base composition in the wild type/

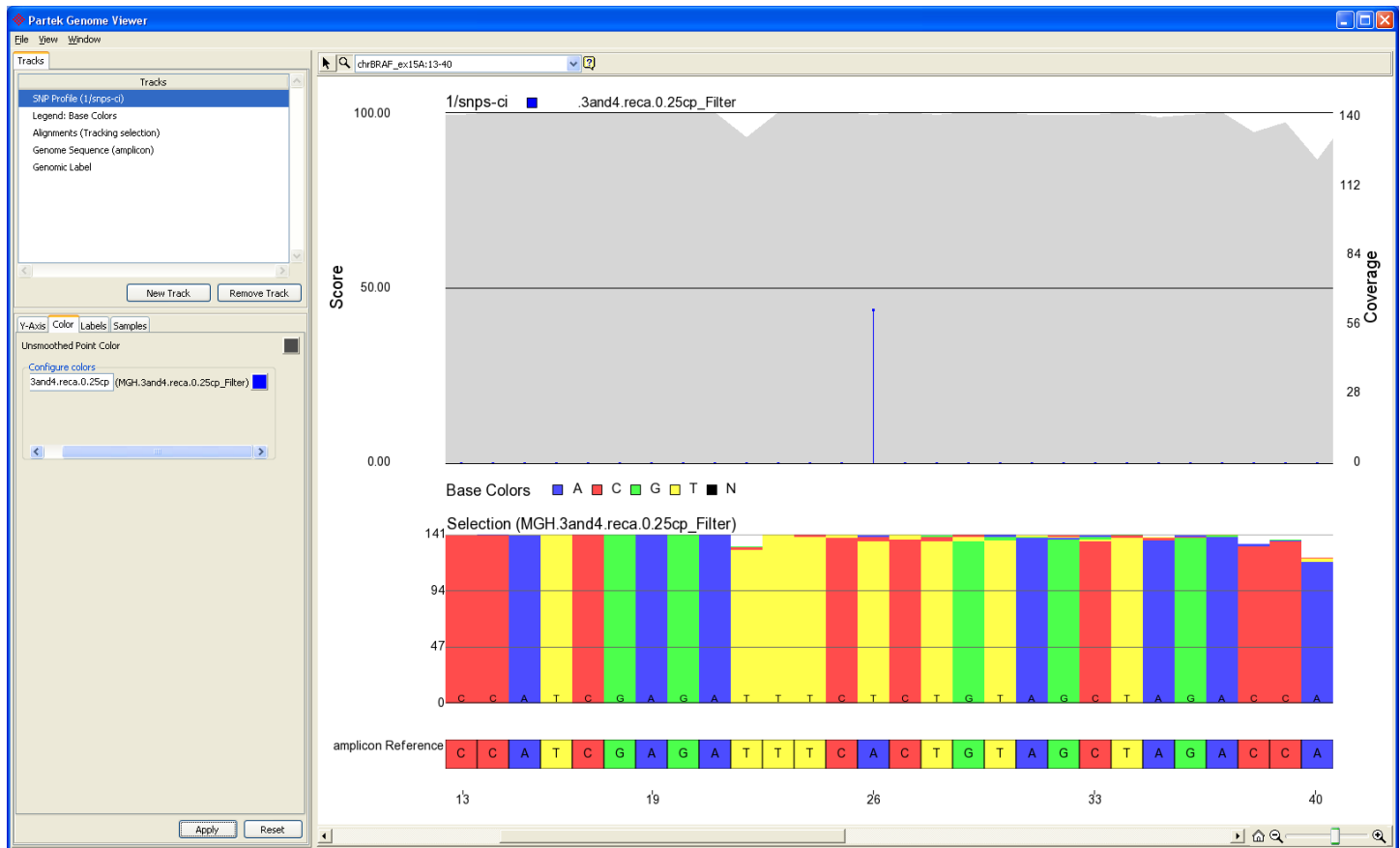


Figure 7. Identifying DNA Variants in a Homogeneous Cell Line Across BRAF. TOP - The graph displays a 22 base pair window within the BRAF gene. Position 26 shows a high mutation score (-log of the p-value) in the form of a purple vertical bar. The gray area displays the depth of Ion Torrent sequencing coverage (approximately 140X) across the region. This variation corresponds to a T to A substitution mutation at position 1799 in BRAF. Approximately 100% of the reads within the sample showed this mutation suggesting a homozygous mutation in a homogenous sample. BOTTOM - The colored bars display the distribution of base calls across the full depth of the Ion Torrent reads at each location. Each of the four nucleotides is represented by a different color according to the bottom legend.

normal sample. For each base along the amplicon a p-value is produced that measures if a variation exists and the probability that the identified variant is due to chance. These results are produced in an exportable data table and also displayed graphically (Fig 7). This graphical image allows scientists to view the genomic location of where mutations are found (based on the p-value) and the associated sequence coverage/depth at each position. Running a statistical test at each genomic location helps remove some of the ambiguity associated with finding significant mutations.

Summary

Amplicon sequencing is one of the fastest growing applications of NGS. Its relevance in the clinical research field is especially noticeable. Amplicon sequencing is also broadly used in basic research projects such as following up on regions of interest identified in genome wide association studies, pathway focused research, etc. This application note has shown how to use the Ion Personal Genome Machine™ to detect germ-line mutations or somatic mutations in biologically relevant regions of the genome.

An end to end workflow and experimental design guidelines to sequence amplicons have been provided. Partek® Genomics Suite™ has extended its DNA-Seq data analysis workflow with a statistical test specifically developed for amplicon sequencing projects. The software not only provides all the statistical calculations needed to make confident variant calls, it also produces intuitive graphics that couple variant scores, depth of coverage and genomic positioning. These visualizations are customizable and exportable in a variety of high quality formats for publication.

The simplicity of the Ion Torrent technology allows the completion of a sequencing run in less than 2 hours, in comparison with several days or even weeks needed for other NGS technologies. The simplicity of the technology is the result of combining basic chemistry and semiconductor-based detection. No imaging, fluorescence detection, cameras or scanners are involved.

The great scalability of this technology resides in the fact that the chip is the sequencing machine. The Ion 316™ chip will have 10 times higher throughput than the Ion 314™ chip. The additional throughput will allow the study of multiple barcoded samples

per chip, measure variants present at very low frequencies in heterogeneous samples or analyze more amplicons per sequencing run. Only one capital equipment purchase gives users the access to several levels of throughput depending on their experimental needs. Moreover, longer read-length will allow sequencing bigger amplicons in the near future. Ion Torrent has successfully produced internally good quality reads well over 200 nucleotides long.

Ion Torrent has introduced a revolutionary technology ideally suited for amplicon sequencing. This affordable, fast, simple, scalable post-light sequencing technology is a perfect fit for laboratories of all sizes that want to have access to high throughput sequencing.

For research use only. Not intended for any animal or human therapeutic or diagnostic use.

© 2011, Life Technologies Corporation. All rights reserved. The trademarks mentioned herein are the property of Life Technologies Corporation or their respective owners.

Ion Torrent by Life Technologies | 7000 Shoreline Court | Suite 201 | South San Francisco, CA 94080 USA
Phone +1-203-458-8552 | Toll Free in North America 1-87-SEQUENCE (1-877-378-3623)

www.lifetechnologies.com
www.iontorrent.com
<http://ioncommunity.iontorrent.com>

C022802

