ion torrent
by life technologies™

Exome sequencing using the Ion Proton™ System

Integrated, easy-to-implement workflow solution for exome sequencing

- The Ion Proton™ System removes the high cost and complexity of today's genome-scale sequencing, bringing it to labs and scientists on their budgets and schedules
- Simple bioinformatics through point-and-click run setup and data analysis in Torrent Suite Software v3.0 combined with exome-specific workflows in Ion Reporter™ Software to identify, prioritize, and report the most biologically interesting variants
- Comprehensive and simple target capture using the Ion TargetSeq™ Exome Kit, which is optimized for use with the Ion Proton™ System, to ensure that exome target enrichment with the highest capture-probe density results in high on-target sequencing reads

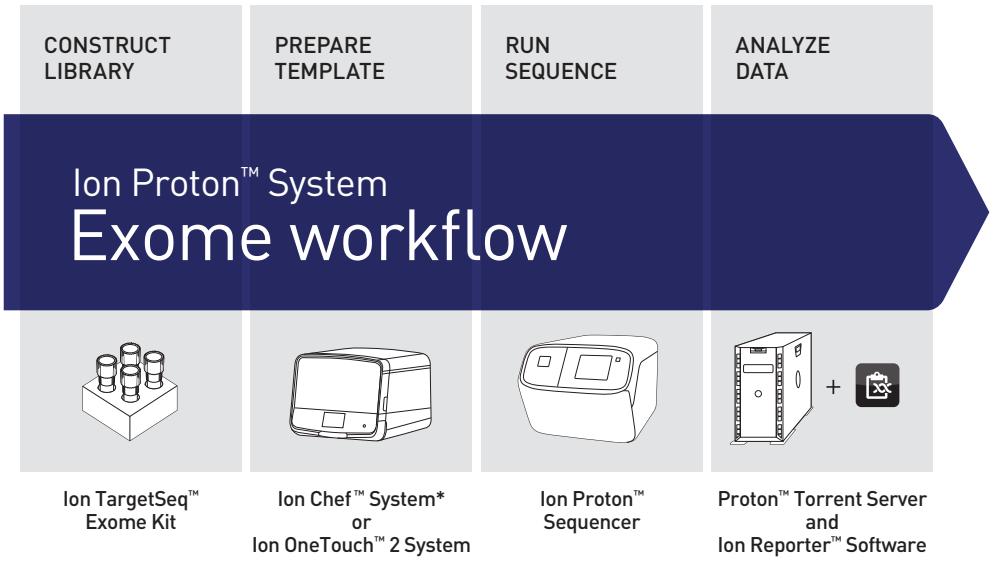


Figure 1. Integrated exome sequencing workflow with ease of use to support labs of all abilities.

Exome sequencing reveals important genetic variation

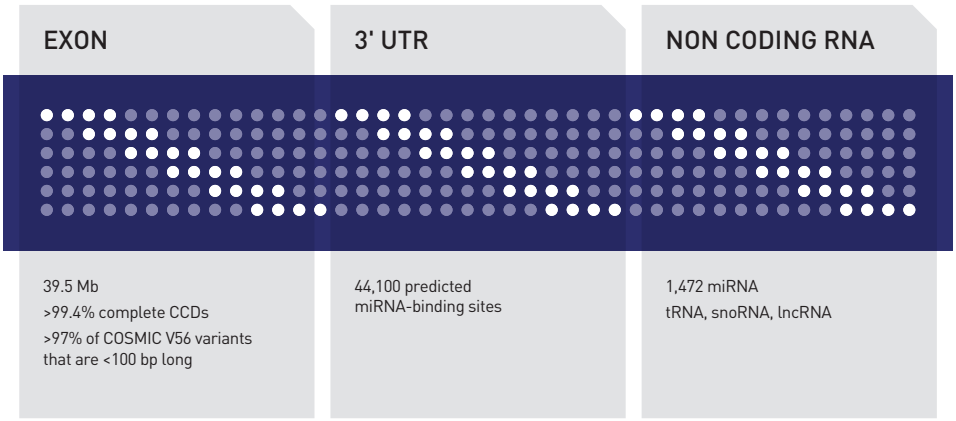
While whole genome sequencing (WGS) provides an optimal solution to variant identification, the combination of data analysis hurdles and expense of WGS has resulted in the development and successful adoption of exome sequencing. Exome sequencing interrogates the interpretable part of a genome through enrichment strategies that target all the coding regions of a genome, regulatory and 3′ untranslated regions, and other functionally annotated regions of interest such as miRNA genes and various noncoding RNAs. For genetic researchers, exome sequencing enables the identification of the single-nucleotide variants (SNVs) and small insertions or deletions (indels) responsible for Mendelian diseases, as well as rare *de novo* mutations that explain the heritability of complex diseases [1]. Exome sequencing has already proven to be a powerful discovery tool, uncovering the causative mutations in a number of developmental disorders, neurological diseases, metabolic disorders, inherited forms of vision loss, and deficiencies of the blood and lymphatic system. Additionally, exome sequencing has provided fundamental insights into the genetic basis of cancer.

Exome sequencing results

Exome sequencing effectively finds the middle ground between the breadth of WGS, with its constraints of data analysis, and the need to be cost-effective. Further, exome sequencing eliminates the time-consuming and costly need to sequence multiple genes and gene panels when attempting to discover rare variants of interest for diseases of

Target capture with the Ion TargetSeq™ Exome Kit

The Ion TargetSeq™ Exome Kit offers a simple and flexible workflow that provides single-tube enrichment using solution-phase DNA probe capture technology for highly specific enrichment of exons and other targeted regions within the human genome (see figure 2). The Ion TargetSeq™ Exome Kit utilizes the highest capture-probe density (>2 million biotinylated capture probes) along with a probe tiling strategy that provides exceptional capture efficiency, resulting in high on-target sequencing read percentages. The kit combines the advantages of low DNA input, as little as 125 ng of each individual library, with affordability through sample multiplexing of up to four barcoded samples in a single target-enrichment reaction. Further, the exome sequencing workflow leverages the automatable enzymatic shearing protocol of the Ion Xpress™ Plus Fragment Library Kit and an integrated workflow that permits ease of use with either the Ion OneTouch™ 2 System or the Ion Chef™ System*† for automated template preparation and chip loading.



Lengths: 60–115 bp, mean 75 **Quantity:** 2.1 million DNA baits
Total: 52.7 Mb (tiled by probes); 46.2 Mb (bases targeted)

Figure 2. Ion TargetSeq™ Exome Kit design.

unknown etiology. We have made available an easy-to-implement, cost-effective, scalable exome sequencing workflow using the Ion Proton™ System with the Ion PI™ Chip, the Ion TargetSeq™ Exome Kit, and data analysis using Ion Reporter™ Software (Figure 1).

Genomic DNA of HapMap CEU sample NA12878 (a female genome of European ancestry that has been deeply sequenced by the 1000 Genomes Project) was used to assess the performance of the Ion Proton™ System with the Ion TargetSeq™ Exome Kit. On an Ion PI™ Chip, out of a total of 168 million wells, 140 million wells (83%) were loaded with Ion Sphere™ particles (ISPs). Of these, a total of 97% (135 million) were productive library ISPs resulting in 9.6 Gb of raw sequence. After quality filtering, 89.7 million reads were generated, of which, 97% (8.8 Gb) were mapped to the human genome using the TMAP alignment tool (Table 1). The mean read length was 107 bp, with a most common read length of 125 bp. Of the aligned reads, 79% (68.9 million) of the reads were on target with a consensus accuracy of 99.996%.

Table 1. Exome sequencing results on the Ion Proton™ System using the Ion PI™ Chip and the Ion TargetSeq™ Exome Kit. The highly characterized NA12878 sample was used to create a fragment library, enriched using the Ion TargetSeq™ Exome Kit, sequenced with the Ion Proton™ System using the Ion PI™ Chip, and analyzed with Torrent Suite Software v3.0.

Sequencing statistics

Raw reads	Reads mapped	Percent reads mapped	Reads on target	Percent reads on target
89,782,719	87,156,364	97.1%	68,899,957	79.1%

Coverage

Mean depth of coverage	Target bases at 1x	Target bases at 10x	Target bases at 20x
119x	98.5%	95.3%	92.5%

Variants

Type	Number of variants	Concordance with dbSNP135
SNVs	30,095	98.0%
Heterozygous SNVs	18,031	97.1%
Homozygous SNVs	12,046	99.4%

Intuitive and simplified exome sequencing data analysis

Torrent Suite Software v3.0 coordinates all the experiment planning and data processing steps necessary to complete your exome sequencing workflow. New product templates enable users to plan exome runs simply and quickly. Raw data are processed on the Ion Proton™ Sequencer and transferred to the Proton™ Torrent Server for base calling and variant calling. This entire process takes several hours and can be followed up by paired or trio exome analysis on the cloud-based Ion Reporter™ Software, or further analyzed using third-party software solutions such as the Ingenuity® Variant Analysis™ Knowledge Module for Ion Reporter™ Software.

Ion Reporter™ Software comprises a suite of bioinformatics tools that streamline and simplify analysis of semiconductor sequencing data. Data generated on the Ion Proton™ System are automatically uploaded from the Torrent Browser to the hosted Ion Reporter™ Software environment for read mapping, annotation and reporting of common and rare variants, and multi-sample analysis—all with version control and audit traceability.

The Ingenuity® Variant Analysis™ Knowledge Module for Ion Reporter™ Software* helps researchers identify causal variants from exome data. Ingenuity® Variant Analysis™ software can be accessed from within Ion Reporter™ Software to help identify, prioritize, and report the most biologically interesting variants for follow-up. Variant Analysis™ software uses a series of filters that users can apply to quickly exclude common variants (based on the 1000 Genomes Project) and nondeleterious variants, and then relate the remaining variants to relevant biology. The biological filtering in Variant Analysis™ software is made possible by the rich biological content in the Ingenuity® Knowledge Base. Published findings about each variant of interest can be reviewed to assess the likely strength of its effect. Also, regulatory diagrams depicting how each variant may impact disease progression, with supporting literature, are readily accessible.

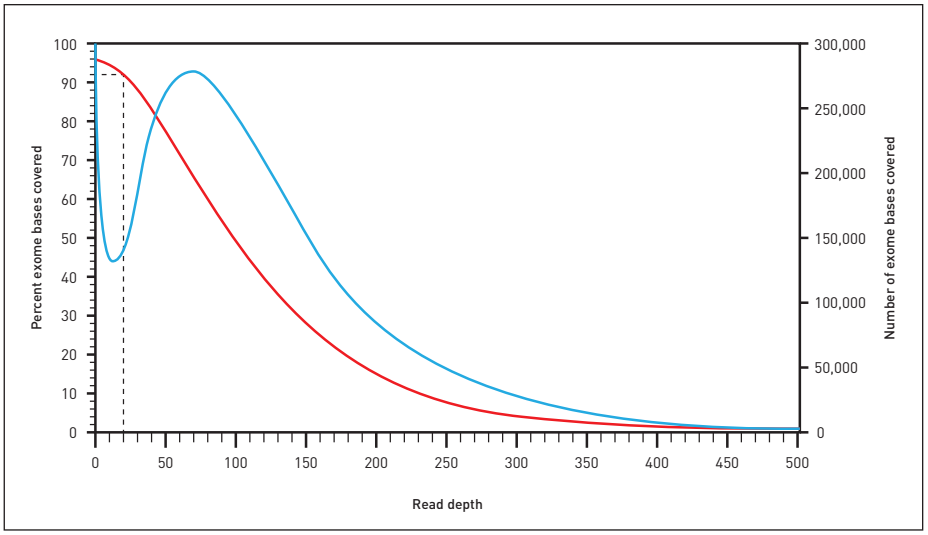


Figure 3. Exome capture data reveal excellent uniformity of target coverage and targeting efficiency with the Ion TargetSeq™ Exome Kit. The red plot indicates targeting efficiency, with the percentage of exome bases covered (left scale) at a particular read depth. The blue plot illustrates uniformity, with the number of exome bases covered (right scale) at a particular read depth.

Exome coverage and variant identification

An important consideration for variant discovery is coverage, i.e., the average number of reads that align to a reference base. Coverage typically translates into confidence in variant calling, with greater coverage increasing the confidence in putative variants.

The targeting efficiency was assessed by determining the coverage for all targeted bases (Figure 3). The mean coverage depth was 119x, with 98.5% of bases covered at least once and >92.5% covered at ≥20x (Table 1). High uniformity of coverage was observed as indicated by the distinct peak with the highest number of on-target bases near the mean depth of coverage of 119x (Figure 3). High uniformity in coverage is important for exome sequencing since this minimizes the amount of sequencing needed to achieve a desired coverage threshold (e.g., 20x) with a significant proportion of exome bases targeted (e.g., >90%).

To evaluate the ability to detect SNVs in exome sequencing data, a total of 30,095 SNVs were identified using variantCaller v3.0.41461, with 98.0% (out of 30,095) of the SNVs found to be concordant with dbSNP135 (Table 1). 97.1% (out of 18,031) of the heterozygous SNVs and 99.4% (out of 12,046) of the homozygous SNVs were found to be concordant with dbSNP135. The transition/transversion (Ti/Tv) ratio was 2.74, which closely corresponds to the Ti/Tv ratio of ~2.8–3.1 estimated for human exomes. The heterozygous to homozygous ratio for SNVs was 1.50, within the range (1.25–1.7) typically observed for genomes of European ancestry.

Seamless variant identification and biological interpretation with Ion Reporter™ Software

It is critical that variants of interest can be quickly identified and prioritized while common variants and nondeleterious variants (those predicted to have no effect on protein function or expression) can be reliably filtered from exome sequencing experiments. Using the Ingenuity® Variant Analysis™ Knowledge Module for Ion Reporter™ Software, a number of loss-of-function variants in NA12878 were identified from a total of 12,046 homozygous SNVs. The loss-of-function variants identified correspond to genuine high-confidence loss-of-function variants in NA12878 that were qualified in a recent study after the application of stringent filters and orthogonal validation [2]. An example of a high-confidence loss-of-function variant identified in a single day using the Ion Proton™ System was the homozygous recessive variant in *FUT2* (W154X) that results in a "non-secretor" phenotype that confers protection to norovirus, a major cause of stomach flu (Figure 4).

Exome sequencing using the Ion Proton™ System with the Ion PI™ Chip

The Ion Proton™ System delivers high-quality exome sequencing that scales with research needs. For instance, up to two exomes per run are possible using the Ion PI™ Chip, and up to 8 exomes per run will be possible using the Ion PII™ Chip.* The Ion PII™ Chip* will support genome sequencing in a few hours with no additional hardware upgrades. The combination of affordable instrument pricing with scalable chips allows the easy implementation of exome sequencing for studies with sample size variability. Further, the sequencing platform and exome target enrichment are optimized with the 100–200 bp single reads of the Ion Proton™ System, and when combined with the Ion TargetSeq™ Exome Kit (which utilizes 60–115 bp probes), results in the optimal sequence capability for targeting human exons that average ~150 bp in size.

Conclusions

The Ion Proton™ System removes the high cost and complexity of genome-scale sequencing and enables exome sequencing using the Ion PI™ Chip and higher exome multiplexing or human genome–scale sequencing with the Ion PII™ Chip.* The Ion Proton™ System combined with Ion TargetSeq™ technology delivers fast, flexible, and high-quality exome sequencing that scales with your research needs. The freedom to plan and execute genomic research on your timeline: confidently go from exome library to variant identification in a single day.

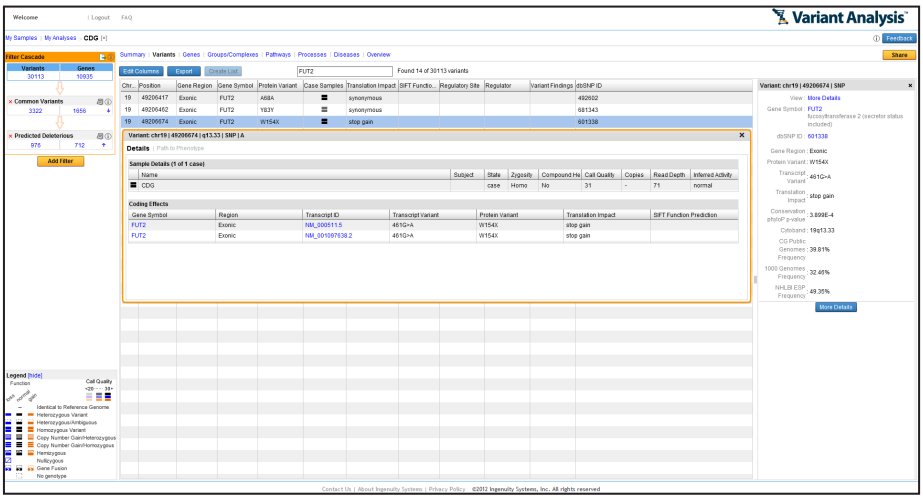


Figure 4. Identification of homozygous loss-of-function variant in *FUT2* (W154X) using the Ingenuity® Variant Analysis™ Knowledge Module for Ion Reporter™ Software.

References

Marth GT et al. (2011) The functional spectrum of low-frequency coding variation. *Genome Biol* 12(9):R84.

MacArthur DG et al. (2012) A systematic survey of loss-of-function variants in human protein-coding genes. *Science* 335(6070):823–828.

Find out more about the Ion Proton™ System at lifetechnologies.com/proton

*The content provided herein may relate to products that have not been officially released and is subject to change without notice.

† Ion Chef™ System is expected to be available for quotation and purchase in Q1, 2013 and is expected to begin shipping in the first half of 2013.

life technologies™