Sequencing and analysis with ION TORRENT System

Aims

Genomnia Ion Torrent platform comprises two sequencing systems, complementary in terms of throughput and capacity:

1. Ion Personal Genome Machine (PGM)
2. Ion S5

Based on the desired sequencing coverage, the required application and the sequencer model it is possible to choose among different chips that enable to sequence from 500,000 to 60-80M of reads in a single run.

The properties of this platform make it extremely versatile, allowing to perform several applications:

- TARGETED DNA SEQUENCING, using AmpliSeq technology, in order to study modifications in single genes or in gene panels linked to a specific disease;
- TRANSCRIPTOME sequencing (Total RNA-seq, Small RNA-seq; AmpliSeq transcriptome);
- TARGETED RNA SEQUENCING, using AmpliSeq technology, for the investigation of gene expression profile of specific genes or gene pathways;
- EPIGENOMICS, for genome-wide studies of transcription factors binding sites (Chip-seq) and DNA methyl-CpG enriched regions (MBD-seq);
- METAGENOMICS;
- *DE NOVO* SEQUENCING or RESEQUENCING of viral and bacterial WHOLE GENOMES.

TARGETED DNA SEQUENCING is performed when the sequence of the specific DNA target is known, either for a small subset of genes (Ion Ampliseq panel) or for thousands of genes, as in whole human exome sequencing studies. Targeted DNA sequencing is useful to detect snv/indel in specific genomic regions potentially related to cancer and Mendelian inherited diseases. The genomic targets are amplified via PCR and then used as input for the library preparation process.

This procedure is based on ***Ion AmpliSeq™*** technology: it enables the selective and simultaneous amplification of target genes in multiplex PCR reactions using hundreds of primer pairs. The PCR products are then sequenced on Ion Torrent platform. ***Ion AmpliSeq™*** technology requires fewer amounts of DNA (roughly from 1 to 100 ng, depending of type of application), thus allowing the sequencing of very poor samples (e.g. biopsy). Moreover, thanks to the small amplicon size this technology enables the analysis of partially degraded DNA samples, such as the one recovered by formaldehyde fixed-paraffin embedded (FFPE) specimens.

Exome sequencing, available only for *Homo sapiens* genomic DNA, is a targeted sequencing approach focused on protein-coding regions, estimated to be approximately 1% of the genome, yet containing roughly the 85% of disease-causing mutations. This technique enables the identification of single nucleotide variants (SNVs), copy number variations (CNVs), and small insertions or deletions (indels), as well as rare *de novo* mutations responsible for the heritability of Mendelian and complex disorders.

For the study of specific set genes, two different types of DNA panels are available: pre-designed panels (*ready to use*) by Life Technologies and custom panels tailor-made by Genomnia. The DNA panels are also classified as “Hot-Spot” panels, suitable to search for known SNPs in non-contiguous genomic regions, or as panels for the sequencing of long contiguous genomic regions, useful to find out both new and known genomic mutations. Custom panels are designed with Ion AmpliSeq™ Designer. The information required to create a panel are derived from databases as COSMIC (Catalogue of Somatic Mutations in Cancer) for cancer-related genes, from NCBI ClinVar for genes associated with inherited diseases and from dbSNP for a collection of all the known genetic alterations. These custom panels enable to perform targeted sequencing of specific genomic regions from various organism, such as human, mouse, pig, sheep, cow, chicken, maize and many others. You can also upload your own reference sequence for any other organism not already preloaded. Specific primer pairs are designed in one or two pools, depending on the genomic size and complexity.

For TRANSCRIPTOME sequencing details, i.e. *Total RNA-seq* and *Small RNA-seq*, please refer to the specific module.

Transcriptome analysis can be also performed using Ampliseq technology, now available for *Homo sapiens* and *Mus musculus* genes. *Ion Ampliseq transcriptome* enables the simultaneous measurement, in a single assay, of the expression levels for over 20,000 genes (representing more than 95% of *Homo sapiens* RefSeq genes and more than 20% of *Mus musculus* RefSeq genes). This panel is designed to produce small-sized amplicons, allowing to work directly with degraded RNA or RNA derived from challenging samples, such as formalin-fixed paraffin-embedded (FFPE) tissue. Total RNA is directly retrotranscribed using random primers, without performing polyA RNA isolation or rRNA depletion. Usually, 10 ng of total RNA are required for each reaction, but it is also possible to start from less RNA, if limiting.

TARGETED RNA SEQUENCING is based on ion Ampliseq technologies and is supported by the same rationale described for targeted DNA sequencing. Custom panels allow the quantification of hundred and thousand genes in a single amplification reaction using specific primer pairs in multiplexing to amplify one single amplicon for each gene. As far as for DNA, Ion AmpliSeq™ RNA panels are available both as *ready-to-use* and custom synthetized by Ion Ampliseq Designer. Whenever possible, amplicons are designed to bridge an exon-exon boundary, minimizing the impact of residual genomic DNA in the isolated total RNA. Moreover, the target region is selected between the transcripts more represented for a single RefSeq gene.

This type of RNA panel also allows the analysis and detection of specific gene fusion.

For a complete and up-to-date list of RNA and DNA panels, please see and log in at www.ampliseq.com

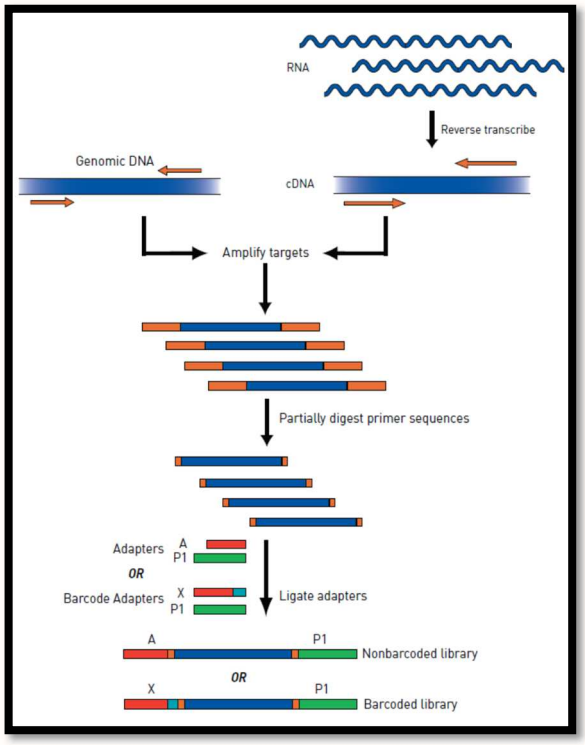
Genomnia offers EPIGENOMICS analyses comprising ChIP-seq, genome-wide study of protein–DNA interactions, and MBD-seq, i.e. the analysis of DNA methyl CpG enriched regions (see specific module). ChIP-seq analysis is used to identify the distribution of DNA transcription factor binding sites or histone epigenetic modifications within the whole genome. The chromatin, native or fixed with formaldehyde, is first enzymatically digested or sonicated and then immunoprecipitated with specific antibody against the protein of interest, in one case, or against the specific histone modification, in the other. DNA is then extracted from the immunoprecipitated complex and ligated to Ion adaptors for performing standard DNA library construction. At the end of sequencing process, the mapped reads identify the genomic coordinates of the transcriptional binding sites or of the histone modification sites.

MBD-seq analysis allows to investigate the genome-wide distribution of methyl-CpG enriched regions. DNA is mechanically fragmented (by sonication) and then selected through a *pull down* step performed with magnetic beads functionalized with the *methyl binding domain* of human MDB protein. These beads specifically bind to methyl-CpG regions thus allowing the selection of genomic fragments with the highest level of methylation. The selected DNA fragments are then used as template to prepare the library. Its sequencing enables to map the genomic distribution of methyl-CpG enriched regions. Both ChIP-seq and MBD-seq provide quantitative information: it is thus possible to perform differential analysis, making a comparison between different samples.

METAGENOMIC studies can be performed analyzing polybacterial samples within a mixed population by sequencing specific regions of 16S rRNA genes. The two primer sets designed for this purpose selectively amplify, respectively, V2, V4 and V8, and V3, V6-7 and V9 hypervariable regions of 16S rDNA. Thanks to this approach, it is possible to obtain the sequence-based identification of a broad range of bacteria at family, genus and, in some cases, even species level.

VIRAL OR BACTERIAL WHOLE GENOME SEQUENCING and *DE NOVO SEQUENCING OF MICROORGANISM GENOMES,* can be also performed to identify new species or to monitor known bacteria and virus as an indication of contamination, disease or healthy conditions.

Library preparation procedure

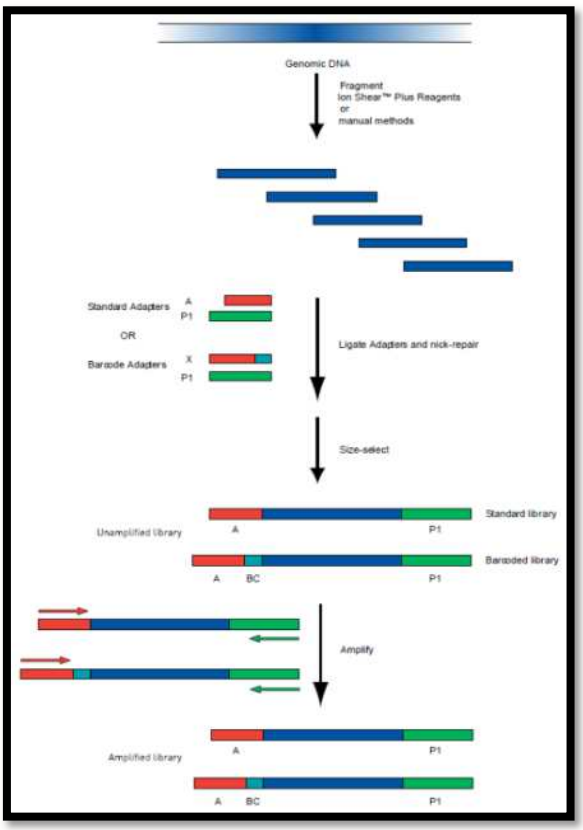


AmpliSeq library preparation procedure is the same for both DNA and RNA samples. As shown in the picture, if the starting material is RNA, the first step is represented by reverse transcription.

Targeted regions of genomic DNA or cDNA are amplified in multiplex using pools of specific primer pairs. The resulting amplicons are enzymatically treated to partially digest the amplification primers and phosphorylate the amplicons. Specific adapters (A/P1) for Ion Torrent sequencing platform are then ligated to the amplicon ends. During this step, the A adapter can be replaced by barcoded adapter for multiplexing analysis.

The adapter ligated amplicons are purified by Agentcourt AMPure XP beads (SPRI technology) and amplified by PCR reaction to enrich for target molecules with adaptors at both end. Library quality and quantity are then checked by fluorometric (Qubit) or capillary electrophoresis (Agilent Bioanalyzer) assays.

Finally, libraries are ready to be sequenced individually or pooled for multiplexed sequencing.



The procedure for fragment library preparation starts with an initial gDNA fragmentation to produce a population of appropriately sized fragments via enzymatic digestion or by sonication. Alternatively, DNA fragments can be obtained by previous steps of specific protocols (e.g. ChIP-seq). The fragments are end-repaired and covalent ligated to Ion-compatible adapters. The adapter-ligated library is then size-selected and PCR-amplified to obtain the required DNA amount for template preparation on Ion Sphere Particles and sequencing.

The procedure described above is followed to produce ChIP-seq library, prepared from the DNA fragments of chromatin complexes immunoprecipitated with specific antibody, and MBD-seq library, with CpG enriched methylated DNA fragments.

Metagenomics libraries also belong to this category: in particular, metagenomic DNA is amplified with primer sets that selectively amplify the hypervariable regions of the 16S rDNA before undergoing library preparation procedure. PCR amplicons are then pooled and processed according to all the steps already described for fragment library preparation.

Sample type and amount

Required sample type and amount depend on the application to be perfomed and on the Quality Control on input samples. Usually, AmpliSeq libraries required from 1 to 100 ng of input material. If possible, an extra amount should be provided to perform initial QC control.

For all the protocols of fragment library preparation 10 ng to 1 µg of DNA per sample are needed. More detailed information will be given while discussing about project design.

Bioinformatic Analysis

Bioinformatic analyses are differentiated by application and are described in detail in every application sheet. In general, for the projects with reference to man or model organisms, all the analyses start from the mapping of sequence reads to the reference genome in a way which is non-ambiguous, highly precise and quantifiable without uncertainties.

The sequence 'signal' (whether it is mutational, epigenetic or transcriptomic) is then quantified and filtered. We use statistical approaches for the evaluation of differential aspects (transcriptome analysis or analysis of epigenetic transcriptional control) or for the description of the genomic variation panorama (SNPs, insertions/deletions).

Results are annotated with reference to the annotation of gene structure in its basic aspects (exons, introns, transcription start site, promoters, etc) and in relation with advanced annotation elements specific for the application.

The results of mutation analysis (genomes or exomes) are instead analyzed using integrated functional prediction algorithms and compared with a list of putative disease genes derived from literature, databases or given by the customer. Always for mutation analysis, we have developed an 'ad hoc' procedure for the 'in trio' mutation analysis, very powerful for the identification of causative disease genes (further details in the correlated service sheet).

For what pertains epigenetic analyses (ChIP-seq, MBD-seq), the regions of interest are annotated with reference to the CpG islands, to the repeated sequences and to UCSC tracks correlated with epigenetic modifications, such as histone methylation and acethylation. A specific service sheet is dedicated to the MBD-seq.

The transcriptome analysis with Ampliseq technology allows to quantify the level of gene expression and to identify the differentially expressed genes between two biological groups. A specific service sheet is dedicated to RNA-seq.

For the smallRNA analysis, in addition to the procedure applied for the identification, quantification and differential analysis of the transcripts, we also apply an advanced analysis which let us to identify and quantify the sequence and length variants (Isomirs) or the putative novel miRNAs. The analytical procedure ends with the differential analysis of these two molecular species.

For the microbiological approaches, we propose analytical strategies which allow not only to identify the composition and quantification of the diversity in single samples, but also to compare statistically the temporal dynamics or the effects of pharmacological and/or biological treatments in different samples.

Ordering information

Please ask to admin@genomnia.it the desired project form specific for each application of interest.

DNA and RNA treatments

Item	Catalog N.
QC: Quality and size control of MBD/CHIP enriched DNA	DNA05
QC: Quality control of DNA preparations	DNA06
QC: Quality control of total RNA preparations	RNA03
Poly(A)$^+$ RNA purification from total RNA	RNA06
Enrichment in Small RNA fraction	RNA07
rRNA depletion of total RNA	RNA10
Enrichment of hypervariable regions of the 16S	16S
MBD-enrichment from genomic DNA	MBD
CHIP-enrichment from genomic DNA	CHIP

Library preparation

Item	Catalog N.
Design and synthesis of a custom Ampliseq panel	PAN
Barcoded DNA library preparation	LDb
Barcoded RNA library preparation	LRb

Sequencing

Item	Catalog N.
Forward sequencing 200 bp tags with barcode	SEQI200B
Forward sequencing 400 bp tags with barcode	SEQI400B
Forward sequencing 600 bp tags with barcode	SEQI600B

Bioinformatic Analysis

Item	Catalog N.
Bioinformatic Analysis I: DNA (panels)	DNA-BF01
Bioinformatic Analysis II: DNA (full exome)	DNA-BF02
Bioinformatic Analysis III: DNA (full exome in trio)	DNA-BF03
Bioinformatic Analysis I: RNA (transcriptome analysis)	RNA-BF01
Bioinformatic Analysis II: RNA (isoforms and networks)	RNA-BF02
Bioinformatic Analysis I: smallRNA (known microRNAs)	Small-BF01
Bioinformatic Analysis II: smallRNA (novel microRNAs, isoMIRs and targets)	Small-BF02
Bioinformatic Analysis I: Metagenomics	METAGEN-BF01
Bioinformatic Analysis I: De Novo	ASSEMBLY-BF01
Bioinformatic Analysis I: Methylation (identification of methylated regions)	MBD-BF01
Bioinformatic Analysis II: Methylation (analysis of differential methylation)	MBD-BF02
Bioinformatic Analysis III: Methylation (differential analysis of LINE/SINE)	MBD-BF03
ChIP-Seq Bioinformatic Analysis I	CHIP-BF01

Genomnia srl
Via L. Ariosto, 21 - 20091 Bresso (MI) - tel +39 02 93305700 - fax +39 02 93305777
www.genomnia.it - info@genomnia.it