Bioinformatic Analysis

Aims

Our company has invested significant resources in establishing a dedicated bioinformatic analysis infrastructure, in terms both of a powerful hardware and software infrastructure and of dedicated skilled personnel with consolidated academic and industrial specific experience in deep sequencing data analysis (functional annotation and statistical analysis). The aim of the Bioinformatics Analysis Unit at Genomnia is in a first instance to deliver timely to the scientific partners or customers high quality Ion Torrent sequence data, unambiguously mapped to the reference genome, annotated, classified and quantified with the maximum level of precision and information content. In addition, we can also be partners in the discovery part of a project, in collaboration with other bioinformatics units and with the Investigators.

Resources

Current hardware resources dedicated to bioinformatics analysis consist in:

- Professional Workstation Windows 7 64-bit, 8 cores, 16 Gb of RAM, 1,5 Tb of disc space;
- Professional Workstation Windows 7 64-bit, 12 cores, 12 Gb of RAM, 1 Tb of disc space;
- Professional Workstation Windows 7 64-bit, 6 cores, 16 Gb of RAM, 1,5 Tb of disc space;
- Server Linux 64 bit con 12 cores Xeon, 128 Gb of Ram, 4 Terabytes of local disc space;
- Cluster (8 computational nodes) Linux 64 bit shared resources: 52 core Xeon, 290 Gb Ram and 8 terabytes of local disc space; 15 Terabytes of shared disk storage;
- A dedicated Linux 64 bit shared memory server with 4 Intel Xeon 5130 cores, 32 Gb RAM and ca. 2 Terabytes of local disk space dedicated to external users;
- High-speed infiniband switch device.

Current bioinformatic analysis skills at Genomnia are focused on sequence count, classification, functional annotation and applied bioinformatics programming. These activities include:

- Transcriptome analysis, both coding and non-coding, of human and model organisms (Whole Transcriptome Analysis).
- Genome-based approaches (analysis of variants, fusions/translocations or indels and SNPs; identification and annotation of promoters; correlation with gene structure; genome signal recognition).
- Epigenetic analyses (ChIP_Seq; methylation).
- Quantitative data analysis (exploratory data analysis; univariate and multivariate approaches; generalized linear models design and evaluation).
- Statistical analysis for qPCR high-throughput datasets and dPCR is also available.

Description of the available services

Details on application-specific analyses are available separately on request. Customers and collaborator scientists can choose in the following range of bioinformatic services when they finalize their project proposal with Genomnia. Pricing of services will vary with the complexity of the tasks and the number of samples to analyze.

DNA:

- Level I (panels) / Level II (whole exome): alignments on the reference genome in binary format ".bam", delivered on a portable HD on request from the customer. Excel tables with mapping and quality metrics. With reference to the CCDS regions: identification of SNPs and small indels (max 20 nt), including the screening for already known variants included in dbSNP. Generation of files of variants in tabular and VCF format. Correlation of SNPs and indels with genome regions annotated in the UCSC and Ensembl databases. Functional annotation, including the evaluation of functional effects of the variants with the SIFT and Polyphen programs. Enrichment analyses and production of files with sequence coverage statistics.
- Level III (trio whole exome): alignments on the reference genome in binary format ".bam", delivered on a portable HD on request from the customer. Tables in text format with sequence coverage metrics for exon with reference to the Ampliseq enrichment kit. With reference to the regions identified in the capture kit: identification of SNPs and small insertion-deletions (max 20 nt), including the

comparison for already known variants and annotations relative to the sequencing features (global sequence coverage; coverage of the variant and reference allele; quality values of the alignment; features of SNPs identified as belonging to dbSNP). Generation of files of variants (SNPs and INDELs) in tabular and standard ".vcf" format. Identification with proprietary procedures of potentially pathogenic 'de novo' mutations; compound heterozygotes; recessive homozygotes. Functional annotation of identified variants, including the prediction of possible functional effects with the SIFT and Polyphen softwares.

RNA:

- Level I (transcriptome analysis): mapping on the reference genome. Generation of quality metrics in text and graphical format. Identification of known transcripts (RefSeq or Ensembl), coding or non-coding, from the whole transcriptome sequencing. Differential expression and functional annotation of the differentially expressed transcripts.
- Level II (advanced transcriptome analysis): identification and differential expression of isoforms of the identified transcripts. Evaluation of differential expression using two different algorithms. Network analysis of Functional Interactions for differentially expressed coding genes.

Small RNA:

- Level I: alignments of the reads on the target genome (delivered on request), filtered for non-microRNA small RNAs (snoRNAs, piRNAs, processed UTRs, tRNAs etc.), in binary format ".bam". Mapping statistics. Identification, annotation and quantification of known microRNAs (identifiable in Mirbase) in the sample, including the differential expression statistical analysis (expression profiling).
- Level II: prediction of putative novel miRNAs. Where possible with the reference organism: identification, classification and differential expression of isoMIRs (variants of known miRNAs); differential expression of novel miRNAs.

Metagenomics:

- Level I: quality control and alignment of the sequences; classification of sequences at different taxonomical levels; differential analysis.

De novo:

- Level I: quality control; assembly of sequences in contigs and generation of metrics such as the number of contigs and the N50 value.

Methylation:

- Level I: alignments against the reference genome in ".bam" format (delivered on request) and mapping statistics. Generation of quality metrics and enrichment diagnostics in text and graphical format. Identification of the genomic methylated regions.
- Level II: in addition to level I, elaboration of differentially methylated genes and regions according to the comparisons planned in the project. Annotation at gene level of the differentially methylated regions. Functional Genic Annotation.
- Level III: in addition to level II, estimated quantification of the methylation level of families of trasponible sequences at elevated redundancy level (LINE, SINE); differential methylation analysis in repeat families according to the comparisons planned in the project.

ChIP-seq:

- Level I: alignments against the reference genome in ".bam" format and mapping statistics. Generation of quality metrics and of enrichment diagnostics in text and graphical format. Identification of peaks (algorithm MACS2). Annotation of peaks with reference to RefSeq or EnsEMBL gene structures.

Results are delivered in Excel tables and summary sheets, accompanied by an explicatory report. Bioinformatics support for data analysis discussions by email, phone and remote connection for four weeks after data delivery is included in the Bioinformatic Analysis activities. Extended support, including face-to-face meetings, is available as a separate service for all the applications with final invoicing

Ordering information

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 (differential analysis of methylated regions)	MBD-BF02
Bioinformatic Analysis III: Methylation (SINE/LINE differential analysis)	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