# 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 sequence data, unambiguously mapped to the reference genome, annotated and classified with the maximum level of precision and information content and quantified with precision. In addition, we can also be partners in the discovery part of a project, in collaboration with other bioinformatics units and with the Investigators. Genomnia also offers to customers with experienced local bioinformatics staff, after formal agreement by contract, direct temporary access to its HPC bioinformatic resources and extended methodological and scientific support.

## *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
- ssh-UNIX direct access to an external HPC infrastructure, localized in CINECA Bologna, for the most intensive and parallel applications.

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; ChIP-Seq).
- 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 (Whole Exome Resequencing in trio and mutation identification; Whole Transcriptome analysis and quantification; annotation, quantification and discovery of microRNAs; methylation 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 analyse.

### *Bioinformatic Analysis I*

- miRNA: Alignments of the reads on the reference genome in binary ".bam" format (delivered on request), filtered for non-microRNA small RNAs (snoRNAs, piRNAs, processed UTRs etc.). Mapping statistics. Identification, annotation and statistics of known Mirbase miRNAs detectable in the sequencing sample, including statistical analysis of differential expression (expression profiling).
- Whole Transcriptome: Mapping on the reference genome. Generation of exploratory quality metrics in text and graphics format. Identification of known genes and transcripts (RefSeq or Ensembl), coding and non-

coding, from whole transcriptome (WTA) sequencing. Functional annotation of the differentially expressed transcripts.

- ChIP-Seq: Reference genome mapping in ".bam" format and mapping statistics. Generation of quality metrics and enrichment diagnostics in text and graphical format. Peak calling (MACS2 algorithm). Peak correlation with RefSeq or Ensembl gene structures.
- Targeted resequencing and All Exome resequencing: Alignments versus the reference genome in binary .bam format, delivered on a portable HD according to the customer request. Excel tables with quality and mapping metrics. With reference to CCDS regions: SNP and small indel (up to 20 nt) detection, including allele calls, coverage calculation and screening for known variants already included in dbSNP. Generation of sequence variant files in text and VCF format. Correlation of SNPs and indels with UCSC or Ensembl annotated gene features. Functional annotation, including the evaluation of the functional effects of variants with SIFT and Polyphen. Enrichment analysis and coverage statistics files.
- Methylation Analysis: Genome alignments with respect to reference genome in ".bam" format (delivered on request) and mapping statistics. Generation of quality metrics and enrichment diagnostics in textual and graphical format. Identification of methylated genome regions. Elaboration of differentially methylated regions and genes according to the project-associated comparisons. Gene-level annotation of the differentially methylated regions.

*Bioinformatic Analysis II: as in Bioinformatica Analysis I, plus*

- miRNA: target identification and prediction of putative novel miRNAs. Where possible with the organism of choice and from the prediction results: identification, classification and differential expression of isoMIRs; differential expression of isoMIRs and of novel small RNAs.
- Whole Transcriptome: Identification and differential expression of transcript isoforms. Evaluation of differential expression with the convergence of two different algorithms. Network analysis of Functional Interactions for the differentially expressed coding genes.
- ChIP-Seq: Differential peak calling and annotation of the differential analysis results obtained with the MACS2 algorithm.
- All Exome 'in trio' resequencing: Alignments to the reference genome in binary ".bam" format and related indexes ("bai" files). Tables in text format with per-exon coverage metrics. With reference to the genome regions included in the capture kit design: identification of SNPs and small insertions – deletions (max 20 nt), including the comparison with known variants already included in dbSNP (classification in NEW and KNOWN variants) and annotation related to the sequencing features (total sequence coverage; sequence coverage of the variant and reference alleles; quality values of alignments; features of SNPs identified as belonging to dbSNP). Generation of files of variants (SNPs and INDELs separately) in tabular and standard ".vcf" format. Identification, with proprietary procedures, of potentially pathogenic 'de novo' mutations; compound heterozygotes; recessive homozygotes. Functional annotation of the identified variants, including the prediction of the possible functional effects with the SIFT and Polyphen programs.
- Methylation Analysis: As in Methylation Bioinformatic Analysis I. In addition: estimated quantification of methylation of transposable sequences at elevated redundancy level (LINE, SINE); differential analysis of methylation in repeat families according to the project-associated comparisons.

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.

***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 (transcriptome advanced analysis) | RNA-BF02 |
| Bioinformatic Analysis III: RNA (custom analysis) | RNA-BF03 |
| Bioinformatic Analysis I: smallRNA | Small-BF01 |
| Bioinformatic Analysis II: smallRNA (advanced) | Small-BF02 |
| Bioinformatic Analysis I: Metagenomics | METAGEN-BF01 |
| Bioinformatic Analysis I: De Novo | ASSEMBLY-BF01 |
| Methylation Bioinformatic Analysis | MBD-BF01 |
| Methylation Bioinformatic Analysis (advanced) | MBD-BF02 |
| 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