

## Analisi Bioinformatica

### Scopo

La nostra azienda ha investito risorse significative nello stabilire una risorsa bioinformatica dedicata, sia come infrastruttura hardware e software che come personale qualificato con esperienza pluriennale specifica (accademica ed industriale) nell'analisi dati di sequenziamento massivo, sia sotto l'aspetto di annotazione che quantitativo/statistico e nella creazione di pipeline di analisi originali. Lo scopo dell'Unità di Analisi Bioinformatica in Genomnia è in primo luogo quello di consegnare in tempo utile ai partner scientifici o ai clienti dati di sequenza Ion Torrent ad alta qualità, mappati senza ambiguità al genoma di riferimento, annotati e classificati e quantificati con il massimo livello di precisione e contenuto informativo. Possiamo inoltre essere partner scientifici nelle attività di un progetto, anche in collaborazione con altre unità di bioinformatica e con altri investigatori.

### Risorse

Le risorse computazionali espressamente dedicate solo all'analisi bioinformatica consistono in:

- Workstation professionale Windows 7 64-bit con quad core, 32 Gb di RAM e ca. 800 Gb di disco;
- Workstation professionale Windows 7 64-bit con 12 core, 12 Gb di RAM e 1 Tb di disco;
- Workstation professionale Windows 7 64-bit con 6 core, 16 Gb di RAM e 1,5 Tb di disco;
- Server Linux 64 bit con 12 core (dual-Xeon esacore), 128 Gb di Ram e circa 4 Terabytes di spazio disco locale;
- Cluster (otto nodi computazionali) Linux 64 bit a risorse distribuite fra i nodi: 52 core Xeon, 290 Gb Ram e 8 terabyte di spazio disco; più di 15 Terabytes di spazio disco globale condiviso;
- Due server Linux multiprocessore dedicati alla produzione ed all'analisi primaria dei dati Ion Torrent;
- Switch di comunicazione ad alta velocità infiniband;
- Accesso diretto a livello UNIX-ssh ad una infrastruttura di High Performance Computing esterna presso il CINECA di Bologna per i lavori più paralleli ed intensivi.

Le competenze analitiche presenti in Genomnia sono focalizzate sulla quantificazione, classificazione, annotazione funzionale delle sequenze e sulla programmazione bioinformatica applicata:

- Analisi del trascrittoma, sia codificante che non codificante, nell'uomo o in organismi modello (Analisi del trascrittoma intero).
- Approcci basati sull'analisi del genoma (analisi delle varianti, delle fusioni/traslocazioni o delle inserzioni/delezioni; predizione e riconoscimento dei promotori; correlazione con la struttura genica; riconoscimento dei segnali nel genoma).
- Analisi epigenetiche (ChIP-Seq; metilazione).
- Analisi dei dati quantitativa (analisi esplorativa dei dati; approcci univariati e multivariati; disegno e calcolo di modelli lineari generalizzati).
- L'analisi statistica di dati di PCR quantitativa e dPCR è altresì disponibile.

### Descrizione dei servizi disponibili

Informazioni aggiuntive e dettagliate su analisi specifiche per applicazione sono disponibili su richiesta. I clienti e collaboratori possono scegliere tra le opzioni descritte di seguito quando finalizzano la loro proposta di progetto con Genomnia. Il prezzo di questi servizi varierà in funzione della complessità del progetto e del numero di campioni da analizzare.

DNA:

- Livello I (pannelli) / Livello II (esoma intero): allineamenti rispetto al genoma di riferimento in formato binario ".bam", consegnati su un HD portatile su richiesta del cliente. Tavole Excel con metriche di qualità e di mappaggio. Con riferimento alle regioni CCDS: identificazione di SNP e piccole inserzioni – delezioni (max 20 nt), includendo screening per varianti note già incluse in dbSNP. Generazione di files di varianti in formato tabellare e VCF. Correlazione di SNP ed indel con regioni geniche annotate dei database UCSC ed Ensembl. Annotazione funzionale, inclusa la valutazione degli effetti funzionali delle varianti con i programmi SIFT e Polyphen. Analisi di arricchimento e produzione di files con statistiche di coverage.

- Livello III (esoma intero in trio): allineamenti rispetto al genoma di riferimento in formato binario “.bam” su HD portatile, su richiesta del cliente. Tabelle in formato testo con metriche di copertura di sequenza per esone relativo al kit di arricchimento ampliseq. Con riferimento alle regioni relative al kit di cattura: identificazione di SNP e piccole inserzioni – delezioni (max 20 nt), includendo il confronto per varianti note già incluse in dbSNP ed annotazioni relative alle caratteristiche del sequenziamento (copertura di sequenza totale; dell’allele variante e di quello relativo alla referenza; valori di qualità degli allineamenti; caratteristiche delle SNP identificate come appartenenti a dbSNP). Generazione di files di varianti (SNP ed INDEL separatamente) in formato tabellare e standard “.vcf”. Identificazione tramite procedure proprietarie di mutazioni potenzialmente patogeniche ‘de novo’; eterozigoti composte; omozigoti recessive. Annotazione funzionale delle varianti identificate, inclusa la predizione dei possibili effetti funzionali con i programmi SIFT e Polyphen.

#### RNA:

- Livello I (analisi del trascrittoma): mappaggio sul genoma di riferimento. Generazione di metriche di qualità in formato testuale e grafico. Identificazione di trascritti noti (RefSeq o Ensembl), codificanti e non-codificanti, dal sequenziamento di trascrittoma intero. Espressione differenziale ed annotazione funzionale dei trascritti differenzialmente espressi.
- Livello II (analisi avanzata del trascrittoma): identificazione ed espressione differenziale di isoforme dei trascritti identificati. Calcolo dell’espressione differenziale tramite l’utilizzo combinato di due algoritmi differenti. Analisi di Network di Interazioni Funzionali per i geni codificanti differenzialmente espressi.

#### Small RNA:

- Livello I: allineamenti delle read sul genoma (consegnati su richiesta), filtrati per piccoli RNA non-microRNA, (snoRNA, piRNA, UTR processati, tRNA etc.) in formato binario “.bam”. Statistiche di mappaggio. Identificazione, annotazione e conta dei micro RNA noti (identificabili in Mirbase) nel campione, inclusa l’analisi statistica dell’espressione differenziale (expression profiling).
- Livello II: predizione di putativi nuovi miRNA. Dove possibile con l’organismo di riferimento: identificazione, classificazione ed espressione differenziale degli isoMIR (varianti dei miRNA noti); espressione differenziale dei nuovi miRNA.

#### Metagenomica:

- Livello I: controllo di qualità e allineamento delle sequenze, classificazione delle sequenze a vari livelli tassonomici; analisi differenziale.

#### De novo:

- Livello I: controllo di qualità, assemblaggio delle sequenze in contigui e generazione di metriche come il numero di contigui ed il valore N50.

#### Metilazione:

- Livello I: allineamenti rispetto al genoma di riferimento in formato “.bam” (consegnati su richiesta) e statistiche di mappaggio. Generazione di metriche di qualità e diagnostiche di arricchimento in forma testuale e grafica. Identificazione delle regioni genomiche metilate.
- Livello II: in aggiunta al livello I, elaborazione delle regioni e dei geni differenzialmente metilati secondo i confronti previsti dal progetto. Annotazione a livello genico delle regioni differenzialmente metilate. Annotazione Genica Funzionale.
- Livello III: in aggiunta al livello II, quantificazione stimata del livello di metilazione di famiglie di sequenze trasponibili a elevato livello di ridondanza (LINE, SINE); analisi differenziale della metilazione nelle famiglie di repeat secondo i confronti previsti dal progetto.

#### ChIP-seq:

- Livello I: allineamenti rispetto al genoma di riferimento in formato “.bam” e statistiche di mappaggio. Generazione di metriche di qualità e diagnostiche di arricchimento in forma testuale e grafica. Identificazione dei picchi (algoritmo MACS2). Annotazione dei picchi rispetto a strutture geniche RefSeq od Ensembl.



I risultati sono consegnati nella forma di tavole Excel e sommari, accompagnati da un rapporto esplicativo. Supporto bioinformatico a distanza (email, telefono, connessione remota) per discussioni sull'analisi dati è incluso per tutte le attività analitiche sopra descritte, per un periodo di quattro settimane dalla consegna dei dati. Supporto bioinformatico esteso, compreso contatto diretto, è disponibile per tutte le analisi come servizio aggiuntivo a richiesta e fatturazione a consuntivo.

## **Informazioni per gli ordini**

Prodotto	N. Catalogo
Analisi Bioinformatica I: DNA (pannelli)	DNA-BF01
Analisi Bioinformatica II: DNA (esoma intero)	DNA-BF02
Analisi Bioinformatica III: DNA (esoma intero in trio)	DNA-BF03
Analisi Bioinformatica I: RNA (analisi del trascrittoma)	RNA-BF01
Analisi Bioinformatica II: RNA (isoforme e network)	RNA-BF02
Analisi Bioinformatica I: smallRNA (microRNA noti)	Small-BF01
Analisi Bioinformatica II: smallRNA (nuovi microRNA, isomiR e targets)	Small-BF02
Analisi Bioinformatica I: Metagenomica	METAGEN-BF01
Analisi Bioinformatica I: De Novo	ASSEMBLY-BF01
Analisi Bioinformatica I: Metilazione (identificazione delle regioni metilate)	MBD-BF01
Analisi Bioinformatica II: Metilazione (analisi di metilazione differenziale)	MBD-BF02
Analisi Bioinformatica III: Metilazione (analisi differenziale delle LINE/SINE)	MBD-BF03
Analisi Bioinformatica CHIP-Seq I	CHIP-BF01