

## Analisi del trascrittoma

### Obiettivi

Nell'ambito dell'analisi del trascrittoma Genomnia è in grado di offrire diverse tipologie di servizio:

- **Total RNA-seq:** questa procedura permette di eseguire tramite sequenziamento un'analisi quali-quantitativa e senza ipotesi *a priori* dell'intera popolazione dei trascritti, codificanti e non, presenti in un campione;
- **Small RNA-seq:** tale approccio consente di analizzare in modo specifico, secondo lo stesso rationale descritto per il Total RNA-Seq, la popolazione dei soli piccoli RNA presenti in un campione;
- **Trascrittoma Ampliseq:** questa metodica fornisce la possibilità di investigare in modo rapido e con approccio massivo l'espressione genica di campioni di RNA umani o murini, utilizzando pool di primers specifici per l'amplificazione degli RNA target.

L'analisi del trascrittoma mediante sequenziamento è caratterizzata da un'elevata sensibilità e specificità e da una risposta dinamica più ampia di quella che può essere ottenuta con la tecnologia dei *microarrays*. Per questo, è ormai da tempo considerata il metodo d'elezione per lo studio dei profili di espressione genica su scala genomica.

Grazie alle differenti strategie adottabili, il sequenziamento massivo eseguito con piattaforma Ion Torrent permette non solo la rilevazione e la quantificazione degli RNA noti, ma anche l'individuazione di nuovi esoni trascritti, di nuovi piccoli RNA, di giunzioni di splicing alternativo e di trascritti non processati o espressi a livelli molto bassi. Grazie ad una accurata analisi bioinformatica, è inoltre possibile identificare e contare gli SNPs espressi, valutando quindi l'espressione genica allele-specifica, ed individuare trascritti di fusione. Inoltre, la metodica utilizzata per la preparazione delle librerie per Total RNA-seq e Small RNA-seq (*Ligase-Enhanced Genome Detection*) preserva l'informazione relativa alla direzionalità del filamento trascritto (*strandness*), facilitando così la rilevazione di eventuali trascritti antisenso parzialmente o totalmente sovrapposti.

### Procedura

La metodica per il **Total RNA-Seq** prevede un primo passaggio di purificazione dell'RNA totale e, successivamente, l'isolamento della frazione di RNA poliadenilato o la deplezione dell'RNA ribosomiale. Le ridotte quantità di RNA d'interesse così ottenute vengono frammentate per idrolisi chimica in molecole di 25-500 nt (con la maggior parte dei frammenti aventi dimensioni di circa 100 nt): questo tipo di frammentazione è omogeneo, privo di selettività e garantisce pertanto un campionamento molto uniforme delle reads rispetto ai trascritti. Sfruttando l'azione combinata di cationi ( $Mg^{++}$  e  $Zn^{++}$ ) e calore (tra 90-100°C), vengono prodotti frammenti di lunghezza variabile a seconda della durata del trattamento, successivamente fosforilati usando la T4 chinasi.

La procedura che permette di eseguire l'analisi selettiva dei piccoli RNA (**Small RNA-Seq**) si caratterizza per la presenza di un iniziale passaggio di arricchimento della frazione di small RNA, successivamente utilizzata come template per la costruzione della libreria.

Per entrambi gli approcci sopra citati, le molecole di RNA vengono convertite in una libreria finale di cDNA attraverso una serie di passaggi. Gli RNA vengono dapprima ibridati e quindi ligati, utilizzando una RNA ligasi, ad una miscela di adattatori costruiti in modo da permettere il sequenziamento a partire sempre e solo dall'estremità 5' del filamento senso. Successivamente le molecole di RNA legate agli adattatori vengono retroscritte in cDNA a singolo filamento, purificato utilizzando un sistema di cattura con biglie magnetiche. La libreria di cDNA viene amplificata mediante PCR per ottenere la quantità di campione richiesta per i successivi passaggi di lavorazione e, al contempo, per permettere l'aggiunta delle sequenze terminali necessarie per il sequenziamento. In questa fase, all'estremità 3' delle molecole possono essere aggiunti specifici *barcodes*, corte sequenze di DNA che permettono l'identificazione univoca dei campioni in esperimenti di sequenziamento in multiplexing: ad oggi sono disponibili 96 differenti *barcodes*. La libreria così ottenuta è amplificata in modo clonale tramite PCR in emulsione su biglie.

L'analisi del **trascrittoma con tecnologia Ampliseq** permette di valutare i livelli di espressione genica umana e murina. Per ciascuno degli organismi è stato sviluppato un pannello costituito da un singolo pool di primers, che consente l'interrogazione in multiplexing di più di 20.000 geni (oltre il 95% dei geni RefSeq per *Homo sapiens* e oltre il 90% per *Mus musculus*). I primers sono stati progettati per produrre ampliconi (uno solo per ogni gene) di dimensioni ridotte, in modo da poter analizzare materiale estratto da campioni anche non ottimali, quali tessuti fissati in formalina e inclusi in paraffina (FFPE) e disponibili in quantità limitate.

Il protocollo sviluppato per questa applicazione permette di evitare la selezione dell'RNA poly(A) o l'eliminazione dell'RNA ribosomiale prevedendo l'utilizzo come input di campioni di RNA totale, inizialmente retroscritto con primers random. Le regioni bersaglio del cDNA sono quindi amplificate in multiplex utilizzando pools di coppie di



primers specifici. Gli ampliconi ottenuti vengono trattati enzimaticamente in modo da digerire parzialmente i primers di amplificazione e da fosforilare le estremità degli ampliconi stessi. Quindi, alle due estremità di ciascun amplicone vengono ligati covalentemente adattatori specifici (compresi i *barcodes*, se necessario), richiesti per il sequenziamento con piattaforma Ion Torrent. Per maggiori dettagli in merito alla tecnologia Ampliseq si rimanda alla scheda specifica.

### Caratteristiche quali-quantitativi dei campioni di RNA

Come indicato in precedenza, per l'analisi dell'intero trascrittoma possono essere usati come materiale di partenza due diverse tipologie di campione di RNA selettivamente arricchito: l'RNA poly(A) o l'RNA totale privato della frazione ribosomiale. Per l'analisi dei piccoli RNA è invece necessario partire dalla frazione degli small RNA isolata dall'RNA totale.

In entrambi i casi è necessario utilizzare come materiale di partenza un campione di RNA totale caratterizzato da un valore di RIN  $\geq 7$  (preferibilmente  $\geq 8$ ), misurato con lo strumento Agilent® 2100 Bioanalyzer™ (RNA 6000 Nano o Pico Kit). Questo aspetto risulta particolarmente rilevante per l'analisi dei piccoli RNA, in quanto i prodotti di degradazione degli RNA più grandi hanno dimensioni tale da determinare una competizione con le molecole di small RNA nella produzione delle librerie, riducendo quindi il numero di reads relative alle specie di interesse.

L' RNA poliadenilato è selezionato con uno o più cicli di cattura con oligo(dT) mentre l'RNA totale depleto di rRNA viene preparato utilizzando appositi kit per la rimozione della frazione di RNA ribosomiale, differenti a seconda della specie a cui appartiene il campione in esame. La frazione degli small RNA viene invece selezionata in base alle dimensioni, utilizzando un metodo basato sulla tecnologia Solid Phase Reversible Immobilization (SPRI) e sull'impiego di biglie magnetiche. Dopo la purificazione, l'effettiva rimozione dell'rRNA 18S e 28S o l'arricchimento della frazione di small RNA può essere verificata eseguendo analisi elettroforetica con Bioanalyzer, utilizzando il kit Agilent RNA 6000 Pico.

I campioni di RNA da tessuti o cellule richiesti per queste procedure devono essere completamente privi di DNA contaminante e devono essere disponibili nelle quantità indicate nella tabella sottostante. Il valore più basso si riferisce alla quantità minima richiesta da protocollo per l'esecuzione della procedura. Tuttavia, Genomnia si rende disponibile a valutare insieme al cliente la fattibilità dell'esperimento anche a partire da quantità inferiori, cercando di ottimizzare le condizioni sperimentali in casi particolarmente critici.

Si precisa che, ove possibile, è preferibile disporre di quantità di partenza maggiori, in modo da poter eseguire gli opportuni controlli qualitativi iniziali.

Materiale di partenza	Quantità (µg)
RNA totale, per purificazione Poly(A)	1 - 5
RNA totale, per deplezione rRNA	0.5 - 5
RNA totale, per arricchimento small RNA	1 - 20
	Quantità (ng)
RNA Poly(A)	10 - 500
RNA totale depleto di rRNA	10 - 500
RNA arricchito in small RNA	1 - 100
RNA totale per tecnologia Ampliseq	10

### Analisi bioinformatica

L'analisi bioinformatica del trascrittoma intero (disponibile in due livelli) ha lo scopo di investigare l'intero set di trascritti di cellule o tessuti sotto esame, sia dal punto di vista quantitativo (espressione differenziale di trascritti tra due o più condizioni sperimentali, fasi dello sviluppo, tessuti, etc.) che qualitativo (identificazione e classificazione di isoforme, RNA non codificanti, etc.). Il protocollo di analisi bioinformatica comprende un passaggio di correlazione dei trascritti con il genoma di riferimento (mappaggio); un passaggio di calcolo statistico (analisi differenziale); un

passaggio di annotazione funzionale. Questo protocollo si applica ad organismi che dispongano di un genoma di riferimento, anche incompleto, annotato e disponibile nelle banche dati e nei sistemi di annotazione internazionali (NCBI, UCSC Genome Browser, EBI, Ensembl).

Oltre ai file di allineamento sul genoma, l'output di questa procedura consiste in più worksheet Excel contenenti:

- Informazioni su coverage, qualità degli allineamenti, statistiche di mappaggio;
- i risultati dell'analisi differenziale per i confronti richiesti;
- i risultati dell'analisi delle isoforme (solo con il livello II) e di espressione differenziale elaborati con un secondo approccio (livello II);
- i risultati dell'analisi differenziale elaborati con un secondo approccio e confronto dei risultati ottenuti con i due algoritmi (solo con il livello II);
- l'annotazione con Cytoscape / Functional Interactions module delle liste di geni sovra e sotto espressi ottenuti dalla convergenza dei due approcci (solo con il livello II);
- l'analisi di espressione differenziale dei trascritti non codificanti (solo con il livello II).

L'analisi di profiling dei microRNA serve per analizzare in maniera qualitativa e quantitativa il risultato del deep sequencing di piccoli RNA con tecnologia ION, cioè ad eseguire un profiling di espressione di smallRNA in chiave prima assoluta e successivamente differenziale. I microRNA, in particolare, sono una parte minoritaria ma funzionalmente importante degli smallRNA. Questa procedura costituisce l'analisi di livello I e prosegue poi con la procedura di identificazione degli isomiR, la scoperta di nuovi putativi microRNA (livello II) e la loro analisi differenziale.

## **Referenze Bibliografiche (in neretto i coautori Genomnia)**

Megiorni F, **Colaiacono M**, Cialfi S, McDowell HP, **Guffanti A**, Camero S, **Felsani A**, Losty PD, Pizer B, Shukla R, Cappelli C, Ferrara E, Pizzuti A, **Moles A**, Dominici C. A sketch of known MYCN-associated miRNA networks in neuroblastoma. *Oncology Reports*. 2017;38(1):3-20.

Callari M, **Guffanti A**, Solda G, Merlino G, Fina E, **Brini E**, et al. In-depth characterization of breast cancer tumor-promoting cell transcriptome by RNA sequencing and microarrays. *Oncotarget*. 2016;7(1):976-94.

Soreq L, Salomonis N, **Guffanti A**, Bergman H, Israel Z, Soreq H. Whole transcriptome RNA sequencing data from blood leukocytes derived from Parkinson's disease patients prior to and following deep brain stimulation treatment. *Genom Data*. 2015; 3:57-60.

Fazi B, **Felsani A**, Grassi L, **Moles A**, D'Andrea D, Toschi N, et al. The transcriptome and miRNome profiling of glioblastoma tissues and peritumoral regions highlights molecular pathways shared by tumors and surrounding areas and reveals differences between short-term and long-term survivors. *Oncotarget*. 2015;6(26):22526-52

Soreq L, **Guffanti A**, Salomonis N, Simchovitz A, Israel Z, Bergman H, et al. Long non-coding RNA and alternative splicing modulations in Parkinson's leukocytes identified by RNA sequencing. *PLoS Comput Biol*. 2014;10(3): e1003517.

Megiorni F, Cialfi S, McDowell HP, **Felsani A**, Camero S, **Guffanti A**, et al. Deep Sequencing the microRNA profile in rhabdomyosarcoma reveals down-regulation of miR-378 family members. *BMC cancer*. 2014; 14:880.

**Guffanti A**, Simchovitz A, Soreq H. Emerging bioinformatics approaches for analysis of NGS-derived coding and non-coding RNAs in neurodegenerative diseases. *Front Cell Neurosci*. 2014; 8:89.

Papait R, Cattaneo P, Kunderfranco P, Greco C, Carullo P, **Guffanti A**, **Viganò V**, Stirparo GG, Latronico MV, Hasenfuss G, Chen J, Condorelli G. Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc Natl Acad Sci U S A*. 2013, 110:20164-9

Voellenkle C, Van Rooij J, **Guffanti A**, **Brini E**, Fasanaro P, Isaia E, et al. Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs. *RNA*. 2012;18(3):472-84.

**Guffanti A**, Iacono M., Pelucchi P \*, Kim N., Soldà G., Croft L.J., Taft R.J., Rizzi E., Askarian-Amiri M., Bonnal R.J., Callari M., Mignone F., Pesole G., Bertalot G., Rossi Bernardi L., Albertini A., Lee C., Mattick J.S., Zucchi I., de Bellis G. A transcriptional sketch of a primary human breast cancer by 454 deep sequencing. *BMC Genomics* 2009, 10:163.



Prodotto	N. Catalogo
QC: controllo di qualità di preparazioni di RNA totale	RNA03
Purificazione di RNA Poly(A) <sup>+</sup> da RNA totale	RNA06
Arricchimento nella frazione degli Small RNA	RNA07
Deplezione di rRNA da RNA totale	RNA10
Preparazione di libreria RNA con codice a barre	LRb
Sequenziamento forward 200 bp tags con codice a barre	SEQI200B
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