

Analisi dell'intero trascrittoma (total RNA-seq) o della frazione dei microRNA (small RNA-seq)

Obiettivi

Questa procedura ha l'obiettivo di esplorare tramite sequenziamento in modo quantitativo e senza ipotesi *a priori*, in un caso, l'intera popolazione dei lunghi RNA trascritti, codificanti e non, e nell'altro, l'intera popolazione dei piccoli RNA presenti in un dato campione. Questa tecnica è caratterizzata da un'alta sensibilità e specificità e da una risposta dinamica più ampia di quella che si ottiene con la tecnologia dei *microarrays*, ed è considerata ormai il metodo standard per ottenere profili di espressione genica a livello dell'intero genoma.

Il sequenziamento massivo offerto della piattaforma ION TORRENT permette non solo la rilevazione e la quantificazione dei trascritti noti ma anche la scoperta di nuovi esoni trascritti, di nuovi piccoli RNA, di giunzioni di splicing e di nuovi trascritti non processati o espressi a livello molto basso. Grazie ad una accurata analisi bioinformatica, si possono anche identificare e contare gli SNPs espressi, valutando quindi l'espressione genica allele-specifica, e scoprire trascritti di fusione. Il metodo usato per la preparazione della libreria (Ligase-Enhanced Genome Detection) preserva l'informazione relativa alla elica trascritta (*strandness*) delle molecole, facilitando la rilevazione di trascritti dall'elica opposta e parzialmente o totalmente sovrapposti (antisense).

Procedura

La metodica sviluppata per l'analisi dell'intero trascrittoma prevede un primo passaggio di purificazione dell'RNA totale e isolamento della frazione di RNA poliadenilato o depleto da RNA ribosomiale. Quantità piuttosto limitanti di RNA purificato sono frammentate per idrolisi chimica a dimensioni di 25-500 bp con picco intorno alle 100 paia di basi. 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 poi fosforilati usando la T4 chinasi. La frammentazione chimica ha il vantaggio di fornire un campionamento molto uniforme dei trascritti sul genoma.

La metodica sviluppata per l'analisi dei piccoli RNA prevede un primo step di arricchimento della frazione small RNA con cui costruire la libreria da sequenziare.

I frammenti di RNA, siano questi piccoli RNA o derivanti da trascritti più lunghi, sono convertiti in una libreria finale di cDNA a doppia elica attraverso diversi passaggi. Le molecole di RNA sono prima ibridate e quindi ligate, utilizzando una RNA ligasi, ad una miscela di adattatori costruiti in modo tale da permettere il sequenziamento a partire sempre e solo dall'estremità 5' del filamento senso. L'RNA legato agli adattatori viene convertito in cDNA a singola elica tramite trascrittasi inversa e purificato utilizzando un sistema di cattura con biglie magnetiche. Per ottenere la quantità di campione necessaria per il sequenziamento con la piattaforma ION TORRENT e per attaccare a ciascuna molecola le sequenze terminali necessarie, la libreria di cDNA è amplificata usando un numero variabile di cicli di PCR. Durante questo passaggio è possibile aggiungere all'estremità 3' della molecola corte sequenze di DNA che agiscono come codici a barre per identificare i differenti campioni: ad oggi sono disponibili 96 codici a barre. Le molecole della libreria ottenuta come prodotto finale sono quindi amplificate in modo clonale tramite PCR in emulsione su biglie che poi costituiranno il supporto per il sequenziamento.

Tipo e quantità dei campioni di RNA

Come indicato in precedenza, per l'analisi dell'intero trascrittoma due diverse specie di arricchimenti di RNA possono essere usati come materiale di partenza, RNA poly(A) o RNA totale depleto di rRNA. Per l'analisi dei piccoli RNA, invece, si deve partire dalla frazione degli small RNA arricchita dall'RNA totale. In entrambi i casi è necessario partire con un campione di RNA totale che abbia un RIN ≥ 7 , preferibilmente ≥ 8 , misurato con lo strumento 2100 Bioanalyzer™ della Agilent® (RNA 6000 Nano o Pico Kit). Ciò è particolarmente rilevante per gli studi dei piccoli RNA, in quanto i prodotti di degradazione degli RNA più grandi competono con gli small RNA nella produzione delle librerie riducendo quindi il numero di reads mappate correttamente.

L'RNA poliadenilato è selezionato con uno o più cicli di cattura con oligo(dT) mentre l'RNA totale depleto di rRNA viene preparato usando un kit di cattura della parte rRNA, differente a seconda della specie del campione in esame. La frazione degli small RNA viene infine selezionata per dimensione utilizzando un metodo basato sulla tecnologia SPRI e

biglie magnetiche. Dopo la purificazione, l'assenza di rRNA 18S e 28S o l'arricchimento della frazione di small RNA può essere controllata usando il Bioanalyzer e il kit Agilent RNA 6000 Pico.

La quantità standard di RNA, da tessuti o cellule, richiesta per questa procedura deve essere completamente priva di DNA contaminante ed è indicata nella tabella sottostante. Le quantità minime sono indicazione del limite più basso raggiunto dalla tecnica e possono essere base di discussione per ottimizzare condizioni sperimentali particolarmente critiche.

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

Analisi bioinformatica

L'analisi bioinformatica del trascrittoma intero (disponibile in due livelli) ha lo scopo di investigare sia sotto l'aspetto 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.) l'intero set di trascritti di cellule o tessuti sotto esame. Il protocollo di analisi bioinformatica comprende un passaggio di correlazione dei trascritti con il genoma di riferimento (mappatura); 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 dalle banche dati e sistemi di annotazione internazionali (NCBI, UCSC Genome Browser, EBI, Ensembl).

L'output di questa procedura consiste in più worksheet Excel contenenti:

- i files relativi al mappaggio sul genoma, al coverage, all'analisi di qualità degli allineamenti, alle statistiche di mappaggio (ambedue i livelli);
- i valori di espressione differenziale per i confronti richiesti (livello I);
- i valori di splicing differenziale (analisi delle isoforme) e di espressione differenziale elaborati con un secondo approccio (livello II);
- l'intersezione tra i valori di LogFC delle valutazioni di espressione differenziale ottenute con i due algoritmi (livello II);
- l'annotazione con Cytoscape / Functional Interactions module delle liste di geni sovra e sotto espresso ottenuti dalla convergenza dei due approcci (livello II);
- l'analisi di espressione differenziale dei trascritti non codificanti (RefSeq e/o Ensembl) ottenuta con una comparazione diretta dei files ottenuti dall'estrazione delle read dai files di allineamento con un database specializzato di trascritti ncRNA.

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. Il workflow concettuale di questa analisi è riportato in extenso nell'articolo scientifico "Deep-sequencing of endothelial cells exposed to hypoxia reveals the complexity of known and novel microRNAs" (ref.)

Referenze Bibliografiche (in neretto i coautori Genomnia)

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.

Informazioni per gli ordini

Prodotto	N. Catalogo
QC: controllo di qualità di preparazioni di RNA totale	RNA03
Purificazione di RNA Poly(A) ⁺ 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 (analisi avanzata del trascrittoma)	RNA-BF02
Analisi Bioinformatica III: RNA (analisi custom)	RNA-BF03
Analisi Bioinformatica I: smallRNA	Small-BF01
Analisi Bioinformatica II: smallRNA (avanzata)	Small-BF02