**Biotecnologie**

***La trascrittomica***

Sono trascorsi circa 80 anni da quando **Beadle e Tatum** con gli esperimenti sulla Neurospora e **Garrod** con lo studio sugli errori congeniti del metabolismo operarono un diretto collegamento tra la funzione del gene e un’ attività enzimatica. Al tempo la scoperta fu entusiasmante ma oggi sappiamo che si trattava di una visione estremamente riduttiva di quella che è realmente la funzione dei geni.

Nell’era post-genomica, la **trascrittomica**è uno dei campi di ricerca più sviluppati. Benchè questo termine sembri definire qualcosa di decisamente futuristico, risalgono agli anni ’90 i primissimi tentativi di **studio del trascrittoma**, ovvero: l’insieme completo dei trascritti in uno specifico tipo di cellula o tessuto in un certo stadio di sviluppo e/o in una specifica condizione fisiologica.

Secondo il dogma centrale della biologia, la **trascrizione** è il primo e fondamentale passaggio normativo dell’espressione genica. Il trascritto può dare origine ad una proteina che di fatto sarà l’operatore materiale di una certa funzione nella cellula. Oggi sappiamo però che solo una minima percentuale dei geni 1,5-2% sarà tradotto in proteina e ciò contrasta con l’evidenza che la maggior parte del genoma è continuamente trascritto ma non è tradotto in proteina. Ma quali sono le funzioni attribuibili a questi trascritti che non vengono tradotti in proteina?

La **trascrittomica**, che è diventata un campo ispiratore della ricerca sulle scienze della vita, mira principalmente a catalogare tutte le specie di trascritti (compresi mRNA, ncRNA e piccoli RNA), a determinare la struttura trascrizionale dei geni (in termini di siti di partenza, estremità 5’ e 3’, schemi di splicing e altre modifiche post-trascrizionali) e a quantificare i mutevoli livelli di espressione di ogni trascrizione.

L’analisi del trascrittoma non solo aiuta a spiegare l’incoerenza tra numero di geni codificanti e numero di proteine prodotte, ma è anche il punto di partenza per lo studio della regolazione traduzionale. Recentemente, sono state scoperte un gran numero di specie di RNA trascritte da regioni genomiche *non* *codificanti*, i cui ruoli sono risultati importanti nella regolazione genica. Ad esempio, il 93% della parte di genoma umano tradotto è trascritto in RNA, ma di questo solo il 2% proviene dalle regioni codificanti proteine.

**Cenni storici e tecnologie antecedenti**

Nel 1991 fu pubblicato su Science il primo tentativo di catturare – parzialmente – il trascrittoma umano, dove vennero riportate 609 sequenze di mRNA del cervello umano. Nel 2008 furono pubblicati altri due trascrittomi umani composti da milioni di sequenze derivate da trascritti che coprono circa 16.000 geni. Dopo il completamento del Progetto Genoma Umano e dopo lo studio di molti altri genomi di organismi modello e non, le sequenze del genoma sono state rese disponibili come vettori di informazioni genetiche.

**Partendo dagli EST alla RT-qPCR.**

Gli studi delle singole trascrizioni sono stati eseguiti diversi decenni prima che fossero disponibili gli approcci di trascrittomica. Alla fine degli anni ’70 alcuni mRNA vennero raccolti in “*librerie geniche*“, convertiti in DNA complementare (cDNA) mediante trascrittasi inversa per scopi conservativi. Negli anni ’80, il sequenziamento Sanger detto sequenziamento di prima generazione cominciò ad essere utilizzato per sequenziare singoli trascritti casuali da queste librerie, chiamate tag di sequenza espressa (**ESTs**). Questi TAG possono essere usati come sonde, esche che permettono di catturare singoli messaggeri mediante esperimenti di ibridazione.

Il metodo del sequenziamento Sanger è stato predominante fino all’avvento di metodi ad alto rendimento, come il sequenziamento per sintesi (**Solexa / Illumina**, San Diego, California) o di seconda generazione detto anche **Next Generation Sequencing**. Gli EST o tracce di cDNA sono diventati importanti negli anni ’90 come metodo efficiente per determinare il contenuto genico di un organismo senza l’obbligo di dover sequenziare l’intero genoma. Furono popolari anche la quantificazione delle singole trascrizioni mediante Northern blotting, matrici di membrana di nylon e la **Retro-trascrizione**  **PCR** quantificata mediante real-time (RT-qPCR), ma questi metodi sono risultati troppo laboriosi e capaci di catturare solamente una minuscola sottosezione del trascrittoma.

**Le tecnologie contemporanee o gli Arrays.**

Le tecniche dominanti contemporanee, i microarrays e l’RNA-Seq, furono sviluppati a metà degli anni ’90 e negli anni 2000. Il primo articolo in cui si descrisse un esperimento con i microarray, che misurano l’abbondanza di un definito insieme di trascrizioni attraverso la loro ibridazione con una serie di sonde complementari, è stato pubblicato nel 1995*.*La tecnologia dei microarray ha permesso il dosaggio simultaneo di migliaia di trascritti per gene ad un costo notevolmente ridotto. Dagli anni 2000, sono stati prodotti una serie di microarray per coprire geni noti di organismi modello o economicamente rilevanti. I progressi nella progettazione e produzione di array hanno migliorato la specificità delle sonde e consentito la verifica di più geni su un singolo array. I progressi nel rilevamento della fluorescenza hanno inoltre aumentato la sensibilità e l’accuratezza della misurazione per basse quantità di trascrizioni.



Un microarray è un insieme di sonde di DNA adese ad una superficie solida. Le molecole di mRNA da identificare vengono estratte dalla cellula e convertite in cDNA (target) mediante l’enzima trascrittasi inversa e poi marcate con sostanze fluorescenti. L’ibridazione tra il cDNA target e la sonda adesa alla matrice è facilmente rilevata dal segnale di fluorescenza rilasciato dalla molecola target in seguito alla ibridazione.

La presenza di artefatti di cross-ibridazione e la necessità di conoscere il genoma di riferimento a priori per la costruzione delle sonde complementari alle molecole di cDNA target hanno permesso l’avanzamento dell’RNA-Seq, rendendo la trascrittomica una scienza basata sui metodi di sequenziamento.

Le tecnologie di next-generation sequencing hanno permesso lo sviluppo dell’RNA-sequencing (RNA-Seq) da cui si possono ottenere corte sequenze chiamate **read**. Tali read sono utilizzate per ricostruire il trascrittoma mediante due modalità principali: le read, tra loro assemblate, possono essere allineate ad un genoma di riferimento (transcriptome assembly genome guided) oppure, se non è disponibile un genoma di riferimento per la ricostruzione del trascrittoma, si allineano le sequenze read tra loro per determinare regioni di sovrapposizione che potrebbero determinare read continue (**contig**) e quindi rappresentare una regione più consistente del trascrittoma (de novo assembly). Qualunque sia la strategia utilizzata per l’assemblaggio delle read in un esperimento di RNA-Seq, è necessario l’utilizzo di programmi bioinformatici per valutare la qualità delle read derivanti dal sequenziamento ed il corretto assemblaggio. La ricostruzione del trascrittoma mediante la tecnologia di RNA-Seq permette di identificare trascritti derivanti da splicing alternativo, di caratterizzare modificazioni post-trascrizionali e di rivelare cambiamenti nell’espressione genica.

Sequenziamento dell’RNA. Le corte sequenze lette possono essere allineate ad un genoma di riferimento. In assenza del genoma di riferimento sono allineate fra di loro per ricavarne la continuità della sequenza genica di riferimento.



**L’RNA-Seq** oggi utilizza tecnologie deep-sequencing di recente sviluppo. In generale, una popolazione di RNA (totale o frazionato) viene convertita in una libreria di frammenti di cDNA con adattatori collegati a una o entrambe le estremità. Ciascuna molecola, con o senza amplificazione, viene quindi sequenziata con metodi high-throughput per ottenere sequenze brevi da un’estremità (sequenziamento single-end) o da entrambe le estremità (sequenziamento dell’estremità della coppia). Le letture sono in genere 30-400 bp, a seconda sulla tecnologia di sequenziamento del DNA utilizzata.

La tecnica è stata quindi fortemente influenzata dallo sviluppo di tecnologie di sequenziamento ad alto rendimento (high throughput). Il primo lavoro di RNAseq è stato pubblicato nel 2006 con 105 trascritti sequenziati dalle cellule tumorali della prostata con la tecnica ROCHE 454 che usa come sistema di rivelazione il pirosequenziamento. Con questa tecnica sono sequenziate 400-600 megabases of DNA ogni 10 ore.

Il sistema consiste nel legare dei linker al DNA genomico frammentato per nebulizzazione. Dopo denaturazione i frammenti si separano e ibridano con degli oligo fissati su palline di vetro. Gli oligo fissati sulla pallina oltre ad ibridare con il singolo filamento di cDNA forniscono l’innesco 3’OH per copiare lo stampo. Il doppio strand così generato viene denaturato ed uno rimane fissato alla pallina di vetro e l’altro va a complementare con un altro oligo. Reiterando il processo la pallina di vetro sarà ricoperta di copie dello stesso frammento. Ogni pallina di vetro sarà poi sottoposta al processo di sequenziamento parallelo di Sanger e da ogni pallina verrà fuori un segnale di sequenziamento fortemente amplificato.

L’ RNA-Seq ha iniziato ad aumentare la propria popolarità dal 2008, quando le nuove tecnologie Solexa/Illumina hanno permesso di registrare 109 sequenze di trascrizione nel lievito (*Schizosaccharomyces pombe),*nell’uomo e nel topo. La resa di tali tecnologie è ora sufficiente alla quantificazione accurata di interi genomi umani.

[**https://pt.slideshare.net/cursoNGS/ana-conesa-introduccion-a-ngs/20?smtNoRedir=1**](https://pt.slideshare.net/cursoNGS/ana-conesa-introduccion-a-ngs/20?smtNoRedir=1)

**Applicazioni: diagnostica e profilazione di malattie**

Le strategie trascrittomiche sono state ampliamente applicate in diverse aree della ricerca biomedica, tra cui la diagnosi e la profilazione di malattie. Gli approcci RNA-Seq hanno permesso l’identificazione su larga scala dei siti di inizio della trascrizione, promotori alternativi e nuove alterazioni dello splicing. Questi elementi regolatori svolgono un ruolo fondamentale nelle malattie umane e, pertanto, la definizione di tali varianti è cruciale per l’interpretazione degli studi sull’associazione della malattia. L’RNA-Seq è capace, inoltre, di identificare *polimorfismi a singolo nucleotide* (**SNP**) associati alla malattia, l’espressione allele-specifica e la fusione di geni, contribuendo alla completa comprensione delle varianti che causano la malattia.

L’ RNA-Seq può inoltre fornire informazioni sulla trascrizione di retrotrasposoni endogeni che possono influenzare, tramite vari meccanismi epigenetici, la trascrizione di geni vicini. I **retrotrasposoni** sono elementi trasponibili che proliferano all’interno dei genomi eucariotici attraverso un processo che coinvolge la trascrizione inversa.  Allo stesso modo, è in rapida espansione la possibilità di utilizzare RNA-Seq per comprendere le malattie immunitarie grazie alla capacità di dissezionare le popolazioni di cellule immunitarie e di sequenziare i gruppi di recettori delle cellule T e delle cellule B dai pazienti.

**Trascrittomi umani e patogeni**

L’RNA-Seq dei patogeni umani è diventato un metodo consolidato per quantificare i cambiamenti di espressione genica, identificare nuovi fattori di virulenza, prevedere la resistenza agli antibiotici e svelare le interazioni immunitarie del patogeno ospite. Uno degli obiettivi principali di questa tecnologia è lo sviluppo di misure di controllo delle infezioni e di un trattamento mirato e individualizzato.

Oggi è possibile analizzare il trascrittoma tenendo in conto l’eterogeneità di un campione o di un tessuto mediante il sequenziamento a singola cellula. Sequenziando il trascrittoma di ciascuna cellula è possibile generare cluster di cellule che presentano un trascrittoma simile e quindi identificare differenti popolazioni cellulari producendo dei cluster basati su similitudini .

