



# Metagenomica e diagnostica microbiologica: una sfida per il laboratorio del futuro



**MARTINA RUECA**

Laboratorio di Virologia, Istituto Nazionale per le Malattie Infettive "L. Spallanzani" Ircs, Roma



**MARIA ROSARIA CAOBIANCHI**

Laboratorio di Virologia, Istituto Nazionale per le Malattie Infettive "L. Spallanzani" Ircs, Roma

**La nascita della Biologia Molecolare** ha caratterizzato la seconda metà del Novecento ed ha permeato tutte le discipline biologiche portando con sé una vera e propria rivoluzione tecnologica. Essa ha segnato la scoperta dei meccanismi fondamentali che operano in tutte le cellule e ne controllano e regolano i processi. In quegli anni l'avanzamento tecnologico e l'incremento delle conoscenze infatti sono andati di pari passo, conducendo la comunità scientifica ad una conoscenza sempre più approfondita dei fenomeni biologici che interessano le macromolecole che compongono le cellule; questo ha permesso la costruzione delle solide fondamenta sulle quali oggi si basa la gran parte delle branche della biologia e della medicina.

Da quando Francis Crick nel 1958 ha formulato il dogma centrale della biologia molecolare che postula il flusso dell'informazione dal Dna alle proteine, attraverso la trascrizione dell'Rna, l'attenzione nel campo della ricerca si è concentrata sullo studio degli acidi nucleici: Dna, detentore dell'informazione genetica e responsabile del trasferimento dei geni alle cellule figlie; Rna, responsabile dell'espressione delle proteine de-

“

**Negli anni Ottanta si sono sviluppate molte delle tecnologie ad oggi in uso negli ambiti di ricerca e diagnostica molecolare che interessano varie discipline**

poste alle funzionalità cellulari. Da qui nasce l'interesse atto a decifrare l'informazione nascosta all'interno della struttura a doppia elica del Dna, fino ad arrivare alla risoluzione del codice genetico. Ad intricare il quadro, si aggiunge la scoperta di complessi meccanismi, operati dagli acidi nucleici in unione con particolari proteine, prefissi alla regolazione dell'espressione genica.

Negli anni Ottanta si sono sviluppate molte delle tecnologie ad oggi in uso negli ambiti di ricerca e diagnostica molecolare che interessano varie discipline. Inoltre sono nati diversi filoni di ricerca volti al sequenziamento del Dna, il cui punto di svolta venne segnato dal lavoro di Frederick Sanger, padre del sequenziamento di I generazione basato sul metodo dei terminatori di catena (metodo Sanger). Esso permette di conoscere la sequenza di un frammento di Dna, precedentemente amplificato con l'utilizzo di primers specifici; gli stessi vengono utilizzati nella reazione di sequenza insieme a nucleotidi modificati chimicamente che terminano la reazione di polimerizzazione del Dna, dando luogo alla formazione di frammenti di diverse lunghezze. I nucleotidi terminatori di catena, che nel metodo Sanger moderno sono marcati con diversi fluorofori, fanno sì che i frammenti, separati per elettroforesi capillare, siano riconosciuti da appositi macchinari che traducono l'informazione di sequenza in grafici visualizzabili al computer.

Un punto di svolta importante è da riconoscersi nel lancio del programma di sequenziamento del genoma umano avvenuto nel 1986 e terminato nel 2001, il quale ha portato all'ampliamento delle conoscenze sull'informazione genetica nonché ad un notevole incremento delle tecnologie che continua anco-

ra oggi. Esso ha inoltre contribuito notevolmente allo sviluppo della ricerca bioinformatica, portando alla nascita di software in grado di gestire immense quantità di dati di sequenza e favorendo la creazione di banche dati (come GenBank).

L'introduzione delle tecniche di sequenziamento di nuova generazione, Ngs (Next Generation Sequencing), nel primo decennio del 2000 ha rivoluzionato completamente l'approccio al sequenziamento del Dna. Queste tecnologie, a differenza del metodo tradizionale in cui si può sequenziare solo un frammento per volta, consentono l'estensione del processo a moltissimi frammenti contemporaneamente che vengono sequenziati in parallelo ed in modo rapido; per questa ragione vengono anche dette tecniche di sequenziamento ad alta resa (High-throughput). Per poter essere sequenziato, il Dna viene frammentato attraverso metodi fisici, o enzimatici; successivamente i frammenti ottenuti vengono legati a degli adattatori (frammenti di DNA che sono riconosciuti ed utilizzati dal macchinario preposto al sequenziamento) e subiscono un passaggio di amplificazione. Questi frammenti, che costituiscono la libreria di sequenziamento, servono da stampo per la sintesi di altri frammenti complementari, chiamati reads. A seconda delle esigenze è possibile impostare il sistema per ottenere il numero desiderato di reads necessarie allo scopo dell'analisi stessa. Il passaggio finale e molto complesso è quello di analisi dell'enorme numero di informazioni ottenute (milioni di sequenze) che richiede grande potenza computazionale per analizzare e allineare le reads con sequenze di riferimento presenti nei database. Sono molti gli ambiti che hanno beneficiato dei progressi del sequenziamento, primo tra tutti quello biomedico; le tecnologie NGS permettono di ottenere un panorama completo sulla variabilità genetica tra gli individui ed anche tra i diversi tessuti di uno stesso individuo; questo porta grandi implicazioni e sviluppi nello studio delle malattie genetiche e in ambito oncologico, campi dove si concentrano un gran numero di studi e nei quali i software e i tool bioinformatici si vanno via via affinando e semplificando, rendendo l'analisi sempre di più facile risoluzione.

Oltre alla genetica umana queste analisi vengono applicate allo studio del genoma di qualsiasi specie vivente; questo rende il sequenziamento di nuova generazione uno strumento chiave per

“  
**Per poter essere sequenziato, il Dna viene frammentato attraverso metodi fisici, o enzimatici; successivamente i frammenti ottenuti vengono legati a degli adattatori (frammenti di DNA che sono riconosciuti ed utilizzati dal macchinario preposto al sequenziamento) e subiscono un passaggio di amplificazione**

molti campi di studio che includono la microbiologia e la virologia. In particolare la tecnologia Ngs si sposa all'analisi metagenomica, una scienza che studia l'insieme dei diversi materiali genetici (metagenoma) presente in un dato campione. Questo approccio apre la strada a numerosi studi che riguardano la componente microbiologica (microbioma) e virologica (viroma) di diversi distretti corporei. Il metagenoma della comunità microbica che è ospitata nel corpo umano possiede nel complesso una quantità di geni che è considerata approssimativamente 100 volte più grande di quella contenuta nel genoma umano. Questa grande e diversificata collezione di geni potrebbe nascondere un gran numero di funzioni biologiche e biochimiche la cui conoscenza, in condizioni di salute e patologia, potrebbe rivelare grandi potenzialità dal punto di vista biomedico e della diagnosi delle malattie ad eziologia infettiva. Nel caso della metagenomica batterica, l'analisi in questione può essere effettuata tramite l'utilizzo di sequenze target conservate ed universali, come l'Rna ribosomale 16s. Esso è tradotto da una regione del genoma batterico che risulta abbastanza conservata da poter essere utilizzata per amplificare il DNA, grazie all'utilizzo di primers specifici, ma che allo stesso tempo contiene alcune regioni con un grado di variabilità abbastanza alto da permettere l'identificazione di genere e di specie dei batteri presenti nel campione.

D'altro canto per lo studio del viroma l'analisi è maggiormente complessa, in quanto non esistono nel genoma dei virus (che sia ad Rna o a Dna) regioni conservate universalmente che possano essere sfruttate per l'identificazione; dunque l'unico approccio possibile è quello che prevede il sequenziamento della totalità del materiale genetico presente all'interno del campione (approccio Shot-gun). Questo pone grandi sfide per quanto riguarda l'analisi bioinformatica delle informazioni ottenute, in quanto tra le sequenze risultanti più del 90% è occupato da sequenze di Dna appartenenti all'ospite; dunque aumenta la difficoltà di rilevazione di sequenze virali presenti nel campione che rappresentano una piccolissima parte di quest'ultimo. Ciò implica che l'analisi dei risultati richieda lo sviluppo di strumenti per l'analisi bioinformatica sempre più raffinati.

Gli sviluppi di questo approccio sono promettenti dal punto di vista della diagnostica virale e possono essere utiliz-

zati per analizzare campioni di pazienti affetti da patologie la cui eziologia rimane sconosciuta. L'analisi metagenomica ha un grosso potenziale nell'identificazione di sequenze virali, anche appartenenti a quei virus, non ancora conosciuti, che potrebbero essere troppo divergenti da quelli noti e che quindi non possono essere identificati tramite l'utilizzo di tecniche di amplificazione di sequenze note. Il campo dell'analisi metagenomica tramite sequenziamento di nuova generazione per la caratterizzazione del viroma di diversi distretti corporei pone continuamente nuove sfide nell'ambito della ricerca ed è in continuo divenire e grande espansione. Dunque le potenzialità di questo metodo avanguardistico potrebbero portare alla scoperta di nuovi agenti patogeni e/o all'identificazione di virus che non ci si aspetta di trovare in un determinato distretto corporeo o area geografica e che quindi non sono ricercati con le tecniche diagnostiche utilizzate di routine. Si conclude quindi che gli sforzi che la comunità di biologi e bioinformatici sta affrontando oggi nel cercare di risolvere la difficoltosa analisi della componente virale del corpo umano, potrebbe contribuire significativamente ad implementare i protocolli e le procedure della diagnostica di domani.

Prima che tali approcci siano veramente fruibili nella realtà diagnostica sarà necessario sviluppare protocolli tecnici e strumenti di analisi standardizzati e facilmente trasferibili dall'ambito della ricerca all'ambito applicativo. Questo rappresenta uno sforzo enorme nel quale si stanno cimentando moltissimi gruppi di ricerca, in sintonia con le ditte che intendono investire nella diagnostica del futuro raccogliendo la sfida dell'innovazione tecnologica.

Come già avvenuto per le applicazioni di Ngs alla diagnostica genetica, accanto alla sfida tecnologica si accompagna la sfida dell'utilizzo delle informazioni, che riveste ambiti etici di non facile gestione, come l'emergere di informazioni inattese, le cosiddette "incidental findings", che per la microbiologia possono consistere nell'evidenziazione di infezioni non sospettate, potenzialmente foriere di delicato contenuto medico e sociale.

In conclusione, il futuro della diagnostica microbiologica potrebbe essere rappresentato come un vaso di Pandora, da gestire con delicatezza e rispetto, affinché il suo contenuto si riveli come una risorsa anziché risolversi in una tempesta ingestibile.