

**Il DNA in 3D: con quali algoritmi studieremo l’architettura del genoma? Uno studio italiano pone le basi per individuarli.**

Le tecniche attualmente utilizzate per lo studio dei genomi consentono di analizzare non solo la sequenza, ma anche la struttura tridimensionale del DNA all’interno del nucleo. La ricostruzione dell’architettura tridimensionale del genoma, che richiede l’utilizzo di software scientifici e di algoritmi estremamente sofisticati e potenti, è fondamentale per la salute umana. Un gruppo coordinato da Unimore - Università di Modena e Reggio Emilia e dall’Istituto FIRC di Oncologia Molecolare (IFOM) di Milano ha analizzato e confrontato i diversi metodi matematici per ricostruire al computer la forma 3D del genoma. I risultati pubblicati oggi dalla prestigiosa rivista scientifica *Nature Methods* dimostrano la potenza di combinare strumenti computazionali e tecniche genomiche per svelare con sempre maggior dettaglio la struttura del DNA all’interno delle nostre cellule.

La disposizione spaziale del DNA all’interno delle cellule è molto complessa, e rende particolarmente difficile la sua osservazione. Organizzato nella struttura della cromatina, il DNA è infatti avvolto attorno a una serie di proteine e ripiegato all’interno del nucleo così da poter essere contenuto in uno spazio 200,000 volte più piccolo della sua lunghezza totale. Questo ripiegamento fa sì che regioni molto lontane tra loro nella sequenza lineare dei cromosomi siano in realtà prossime o addirittura a contatto nella struttura tridimensionale. Peraltro, l’organizzazione del DNA all’interno del nucleo non è statica, ma varia e le modificazioni strutturali contribuiscono, in maniera rilevante, a determinare il destino delle cellule, la loro funzione, il loro sviluppo ed eventuali alterazioni patologiche che possono per esempio dare origine a tumori. Determinare la struttura 3D del DNA permette quindi di comprendere, fin nel minimo dettaglio, l’utilizzo del genoma, di identificare le modificazioni conformazionali che trasformano le cellule e, in ultima analisi, potrebbe contribuire a sviluppare strategie più precise ed efficaci per correggere i loro comportamenti aberranti.

Le moderne tecniche di sequenziamento applicate allo studio dell’organizzazione del DNA all’interno del nucleo consentono di ottenere la mappatura completa di tutte quelle regioni, appartenenti allo stesso cromosoma o a cromosomi diversi, che nella struttura 3D sono prossimi e in contatto tra di loro. La tecnica più sofisticata attualmente disponibile per determinare una fotografia completa delle centinaia di milioni di contatti tra frammenti diversi di DNA all’interno del nucleo è chiamata Hi-C e richiede l’applicazione di algoritmi sofisticati e strumenti computazionali molto potenti per gestire, analizzare e interpretare l’enorme mole di dati genomici ottenuti. Lo sviluppo dell’Hi-C ha rappresentato pertanto un traguardo fondamentale per disegnare la mappa tridimensionale del genoma, ma ha anche creato un nuovo problema da risolvere: come sviluppare, ottimizzare e armonizzare i diversi metodi computazionali per l’analisi dei dati.

La sfida è stata raccolta da un gruppo tutto italiano di scienziati coordinati da Silvio Bicciato di Unimore e da Francesco Ferrari di IFOM. Dallo sforzo congiunto di bioinformatici, biologi, biotecnologi e ingegneri è nato quello che, a oggi, è la prima analisi esaustiva dei più complessi strumenti informatici per l’identificazione sistematica della struttura tridimensionale del DNA a partire dai dati di conformazione della cromatina.

«Siamo partiti raccogliendo tutti gli algoritmi e i software disponibili per lo studio dei dati prodotti da Hi-C, e li abbiamo applicati a decine di campioni rappresentanti tipi cellulari diversi per verificare l’efficacia di ciascun metodo nell’identificare interazioni e strutture tridimensionali del DNA», spiega Mattia Forcato, ricercatore presso il Dipartimento di Scienze della Vita dell’Università di Modena e Reggio Emilia e primo autore della ricerca. «Questo studio ha richiesto moltissime ore di lavoro al computer e risorse di calcolo molto importanti, ma gli sforzi sono stati ampiamente premiati dalla scelta di *Nature Methods* di pubblicare la nostra ricerca e di dedicarne la copertina, disegnata da Chiara Nicoletti, assegnista del Dipartimento di Scienze della Vita e co-autore del lavoro» aggiunge Forcato.

«L’enorme mole di dati analizzati e il numero di algoritmi confrontati ci hanno permesso di fornire un’immagine dettagliata di quali siano i punti di forza e i limiti degli strumenti bioinformatici attualmente disponibili per lo studio dei contatti del DNA all’interno del nucleo» commenta Francesco Ferrari, ricercatore che, dopo il dottorato ottenuto a Unimore nel 2008, ha avuto una lunga esperienza di ricerca all’Harvard Medical School. Rientrato in Italia nel 2015, ha aperto il laboratorio di Genomica Computazionale presso IFOM, grazie a un finanziamento Start up dell’AIRC. «Questo studio è nato dall’esigenza di fare chiarezza in un campo in rapida evoluzione come lo studio dell’architettura 3D del DNA. L’articolo appena pubblicato aiuterà i ricercatori a orientarsi nell’analisi e interpretazione dei dati sperimentali. Noi stessi stiamo già usando questi risultati per migliorare la caratterizzazione dell’organizzazione spaziale del genoma» aggiunge Ferrari.

«Fornire le basi computazionali per studiare il ruolo della struttura tridimensionale del DNA e delle sue modificazioni è un passaggio fondamentale per identificare quei processi molecolari che **portano alla rigenerazione dei tessuti o alla crescita tumorale**» spiega Bicciato, a capo dell’Unità di Bioinformatica del Centro Interdipartimentale di Ricerche Genomiche. Lo scienziato, professore di Bioingegneria del Dipartimento di Scienze della Vita dell’Università di Modena e Reggio Emilia, conduce, con il suo gruppo e diversi collaboratori, ricerche che, attraverso l’analisi bioinformatica dei dati genomici, mirano a svelare i meccanismi alla base dell’utilizzo del genoma da parte delle cellule, sia nei loro stati normali sia nelle trasformazioni patologiche. «Se arriviamo a comprendere come le cellule sfruttano la struttura del genoma nelle loro trasformazioni, possiamo concretamente aumentare le nostre possibilità di intervenire per correggere quei meccanismi che, ad esempio, sono fattori chiave della trasformazione tumorale».

I computer rappresentano, quindi, uno strumento fondamentale per addentrarsi nell’organizzazione tridimensionale del genoma e comprendere, attraverso l’uso sinergico di tecniche di biologia molecolare, dati genomici e algoritmi matematici, i meccanismi molecolari attraverso i quali le cellule determinano il loro destino e le loro funzionalità.

Questa ricerca è stata condotta grazie al programma di oncologia clinica molecolare AIRC 5 per mille Molecular basis for triple negative breast cancer metastasis: new tools for diagnosis and therapy, a uno startup grant di AIRC e al progetto europeo ERC DENOVOSTEM. Il lavoro è inoltre parte di Epigen, il progetto italiano di epigenomica, sostenuto dalle risorse del Ministero della Ricerca e diretto dal prof. Giuseppe Macino, per lo studio delle modifiche epigenetiche che si trovano sull’intero genoma e che contribuiscono a determinare il profilo di espressione genica delle cellule in condizioni fisiologiche e patologiche.

Dettagli editoriali

29 Giugno 2017, *Nature Methods*

Titolo originale dell’articolo: “Comparison of computational methods for Hi-C data analysis”

Autori: Mattia Forcato, Chiara Nicoletti, Koustav Pal, Carmen Maria Livi, Francesco Ferrari, Silvio Bicciato