



**POLITECNICO**  
MILANO 1863

## **Genomica computazionale: parte un progetto per la Medicina del futuro**

Stefano Ceri del Politecnico di Milano vince un Advanced Grant ERC, il secondo consecutivo

*Milano, 12 maggio 2016* - Il prof. Stefano Ceri del Politecnico di Milano ha raggiunto un eccezionale risultato: è uno dei due ricercatori europei di Informatica ad aver vinto due Advanced Grant ERC.

Il suo precedente progetto, Search Computing, si è svolto dal 2008 al 2013, mentre **GeCo (Data-Driven Genomic Computing) sta per partire** e si concluderà nel 2021. **GeCo ha un obiettivo rivoluzionario:** sviluppare e consolidare **un nuovo approccio alla medicina integrando l'analisi di "big data" derivanti dal sequenziamento del genoma** per trovare una risposta più precisa ed efficace a tante domande della biologia e della medicina, incluse la modalità di sviluppo dei **tumori** e la loro dipendenza da cause ambientali. GeCo vuole rivisitare la genomica computazionale tramite l'uso estensivo di banche dati pubbliche, ideando nuovi modelli, linguaggi e strumenti per la loro analisi e gestione, solidi dal punto di vista dei concetti utilizzati e capaci di operare in modo super-efficiente su sistemi "cloud".

Cos'è la Genomica Computazionale?

E' la scienza che, partendo dal sequenziamento del genoma e grazie all'uso di analisi statistiche e computazionali, **decifra la funzione delle regioni del genoma** e costituisce pertanto il presupposto per le future scoperte nel campo della biologia e della medicina. Le tecniche di sequenziamento del genoma di nuova generazione (NGS) consentono oggi la produzione dell'intera sequenza del genoma umano a costi molto bassi (circa 1000 dollari). Parallelamente sono stati sviluppati algoritmi specializzati per estrarre le caratteristiche salienti del genoma che si vuole studiare, per evidenziare ad esempio le mutazioni o l'espressione dei geni, cioè la loro attività di trascrizione. La grande lacuna da colmare rimane però l'ideazione di un sistema capace di integrare i dati genomici estratti da tali algoritmi ottenendo un "senso biologico" interpretabile dai medici per comprendere meglio, ad esempio, lo sviluppo di gravi malattie o la loro dipendenza da fattori ambientali.

Gli obiettivi di GeCo per migliorare la Genomica Computazionale.

L'equipe del progetto GeCo, partendo da un modello di dati astratto che garantisce interoperabilità fra i vari formati potenzialmente utilizzabili, **ha già sviluppato un sistema per interrogare dati genomici scaricati da grandi banche dati prodotte da Consorzi**

**internazionali.** Un risultato che si colloca all'avanguardia mondiale della ricerca di settore e che è pubblicamente utilizzabile presso il Consorzio Cineca, oppure scaricabile dai server del Politecnico ([http://www.bioinformatics.deib.polimi.it/genomic\\_computing/](http://www.bioinformatics.deib.polimi.it/genomic_computing/)).

Nel corso del progetto, il sistema sarà arricchito di strumenti per l'analisi dei dati e verrà reso sempre più efficiente, utilizzando vari framework per la gestione di dati disponibili su server paralleli e in ambiente cloud.

Tra gli obiettivi del progetto vi è anche la costruzione di un open source messo a disposizione dei ricercatori biologici e clinici, che potranno usare servizi offerti dal sistema oppure scaricarlo e installarlo presso i loro centri. Mentre i servizi realizzati dal Politecnico di Milano useranno esclusivamente dati pubblici, messi a disposizione per "uso secondario", cioè per attività di ricerca, l'installazione protetta del sistema in un contesto clinico potrà essere utilizzata per la cosiddetta "medicina personalizzata", cioè l'adattamento delle terapie ai dati genomici di specifici pazienti.

L'obiettivo più ambizioso del progetto è la realizzazione di un "**Internet per la genomica**", cioè di un modo di raccogliere dati genomici pubblicati da consorzi internazionali e dai ricercatori, e di un "**Google per la genomica**", cioè un sistema di indicizzazione e ricerca su grandi raccolte di dati genomici pubblici. Questi strumenti potranno essere usati per facilitare in futuro lo studio approfondito di gravi malattie.

Stefano Ceri terrà un seminario sul tema "Data-Driven Genomic Computing: Making Sense of the Signals from the Genome" il 18 maggio 2016 alle ore 16.30, nella Sala Conferenze del Dipartimento di Elettronica Informazione e Bioingegneria del Politecnico di Milano, in Via Ponzio 34/5, Milano.