



**POLITECNICO**  
MILANO 1863

## **Il Broad Institute utilizza il sistema per l'analisi di big data genomici del Politecnico di Milano**

### **Il centro di eccellenza per la genomica computazionale ha scelto il sistema sviluppato all'interno del progetto GeCo**

*Milano, 21 giugno 2018* - Il Broad Institute, centro di eccellenza per la genomica computazionale, ha reso disponibile ai suoi ricercatori GenoMetric Query Language, un sistema per l'analisi di big data genomici sviluppato all'interno del Progetto GeCo del Politecnico di Milano.

GeCo (Data-Driven Genomic Computing) è un Advanced Grant ERC, diretto dal prof. Stefano Ceri del Politecnico, che ha per obiettivo lo sviluppo di un nuovo approccio all'analisi dei "big data" derivanti dal sequenziamento del genoma per trovare una risposta più precisa ed efficace a tante domande della biologia e della medicina, incluse le modalità di sviluppo dei tumori.

L'equipe del progetto GeCo ha sviluppato il sistema GMQL (GenoMetric Query Language) per interrogare dati genomici scaricati da grandi banche dati prodotte da Consorzi internazionali. GMQL è un sistema aperto, utilizzabile pubblicamente. È già installato presso il Consorzio Cineca e scaricabile dai server del Politecnico di Milano stesso.

Il Broad Institute è un centro di ricerca di eccellenza per lo studio delle malattie, con particolare enfasi sui tumori e sulle malattie cardio-vascolari, infettive e psichiatriche; coinvolge ricercatori provenienti da Harvard, MIT e da diversi ospedali legati ad Harvard, che operano in discipline diverse, tra cui la medicina, la biologia, la chimica, l'informatica, la matematica e l'ingegneria.

Il gruppo di Data Science del Broad Institute è responsabile di FireCloud, una piattaforma aperta per l'analisi dei dati genomici che garantisce sicurezza ed elevate prestazioni; FireCloud consente l'accesso protetto a dati genomici presenti in TCGA (The Cancer Genome Atlas) e l'uso di GATK, un software sviluppato al Broad Institute per l'individuazione di mutazioni.

Al termine di un percorso di integrazione del software, anche GMQL è disponibile su FireCloud. Usando una workspace pubblica creata per GMQL, i ricercatori possono vedere GMQL "al lavoro" su tre casi di studio di progressiva complessità, e possono quindi utilizzare il codice sviluppato al Politecnico di Milano per integrare e completare le loro analisi dei dati.

Il gruppo del Politecnico ha fatto da "apripista" nella sperimentazione dell'uso delle workspace pubbliche, uno strumento di FireCloud per

facilitare l'integrazione con strumenti software sviluppati da gruppi di ricerca esterni al Broad Institute.

**Maggiori informazioni:**

<https://software.broadinstitute.org/firecloud/blog?id=12125>