

Sommersi dai byte dei genomi

Il sequenziamento genomico a basso costo sta inondando il mondo di dati rilevanti. Come contrastare il diluvio, senza rimuoverli?

Mike Orcutt

Il costo del sequenziamento del genoma umano sta crollando; in alcuni dei centri genomici più avanzati, si sta abbassando cinque volte più rapidamente di quello dei computer. Di conseguenza, le persone sempre più di frequente fanno analizzare il loro DNA da aziende e laboratori di ricerca, per ottenere informazioni su varianti e malattie genetiche.

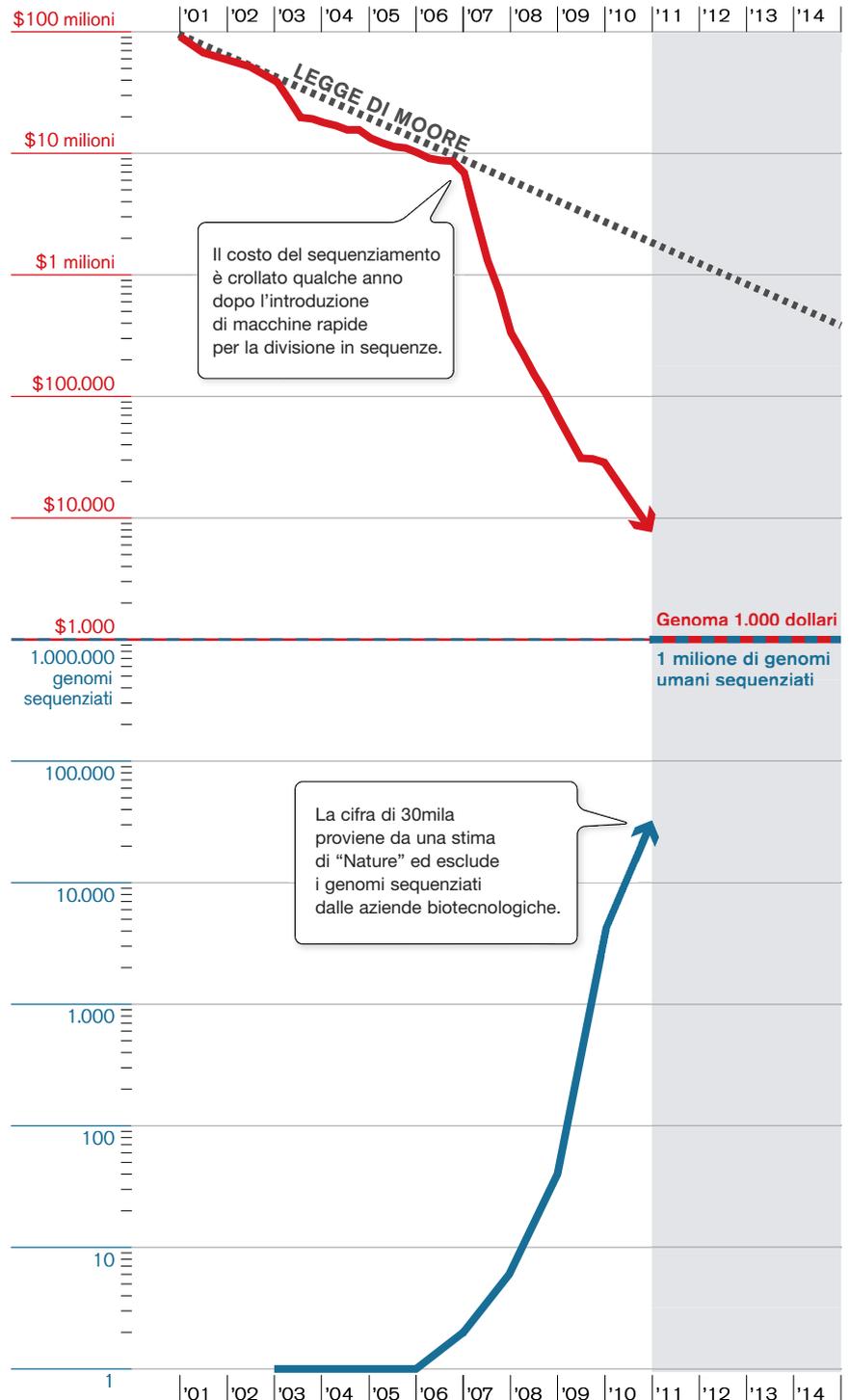
L'industria si trova di fronte al problema di archiviare a costi accettabili questa imponente massa di dati. Ognuna delle 3,2 miliardi di paia di basi del genoma umano può essere decifrata con due bit, per complessivi 800 megabyte per l'intero genoma. Ma in genere si raccolgono i dati significativi per ogni base e i geni vengono spesso suddivisi in sequenze molte volte per garantirne l'accuratezza, per cui è prassi normale salvare circa 100 gigabyte effettuando, per esempio, il sequenziamento del genoma umano con una macchina prodotta da Illumina, il leader industriale di settore.

Archiviare i dati relativi a tutti gli abitanti del pianeta richiederà tanta memoria digitale quanta quella che è stata utilizzata globalmente nel 2010. La soluzione più logica consiste nel salvare un numero limitato di dati. George Church, genetista di Harvard, avanza l'ipotesi di memorizzare solo le differenze tra una nuova sequenza genomica e un genoma di riferimento. Questa informazione si potrebbe codificare in 4 megabyte, non più di un allegato di una e-mail. **IR**

La mappa, nella pagina accanto, tiene conto di informazioni estratte da una banca dati generata dagli utenti, basata su statistiche disponibili pubblicamente, che prendono in considerazione il 60/70 per cento delle macchine; sono escluse le aziende farmaceutiche e biotecnologiche, oltre ad alcuni fornitori di servizi per la divisione in sequenze.

Crollo dei costi del sequenziamento, esplosione di richieste

Costo per genoma e quantità di genomi sequenziati

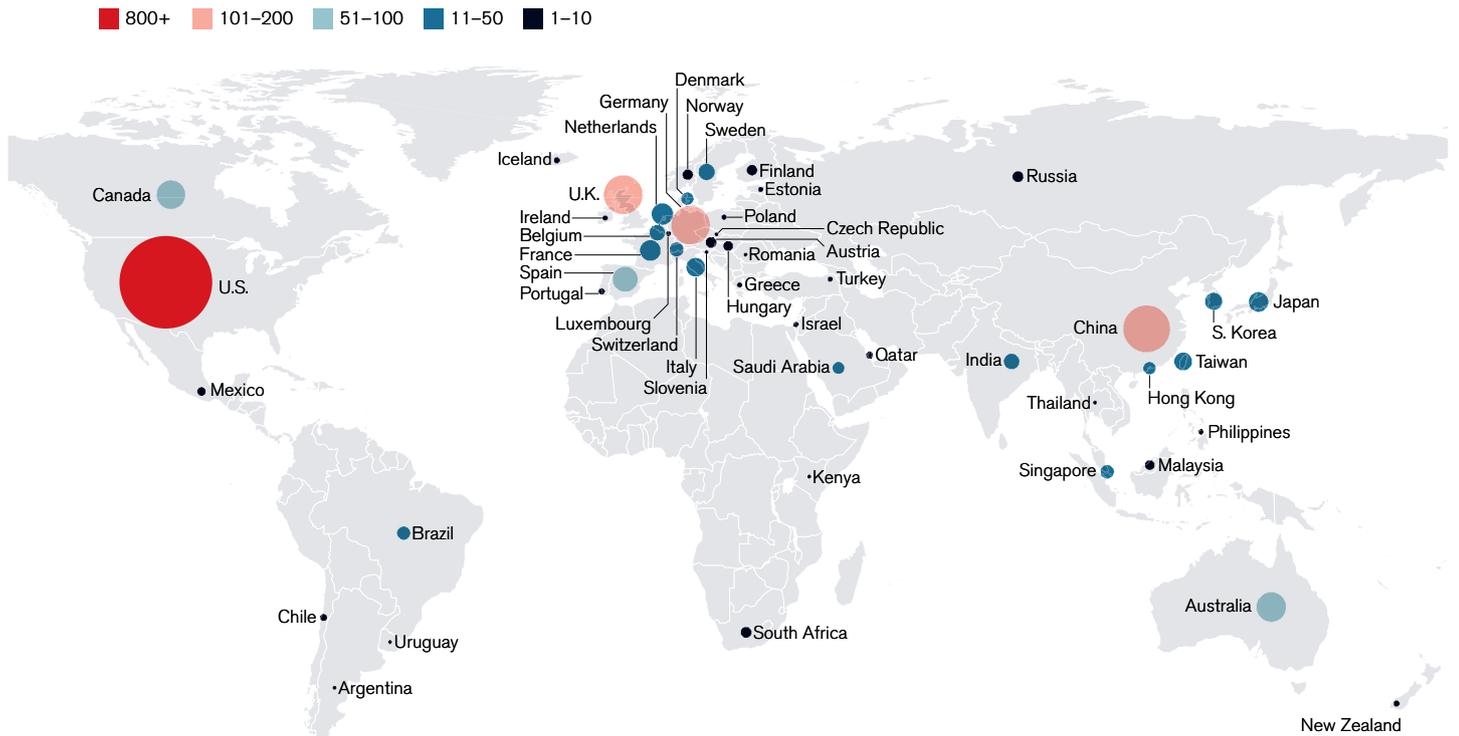


Fonti: US National Human Genome Research Institute, "Nature"

INFOGRAFICA: INFOGRAPHICS.COM

Il sequenziamento veloce diventa globale

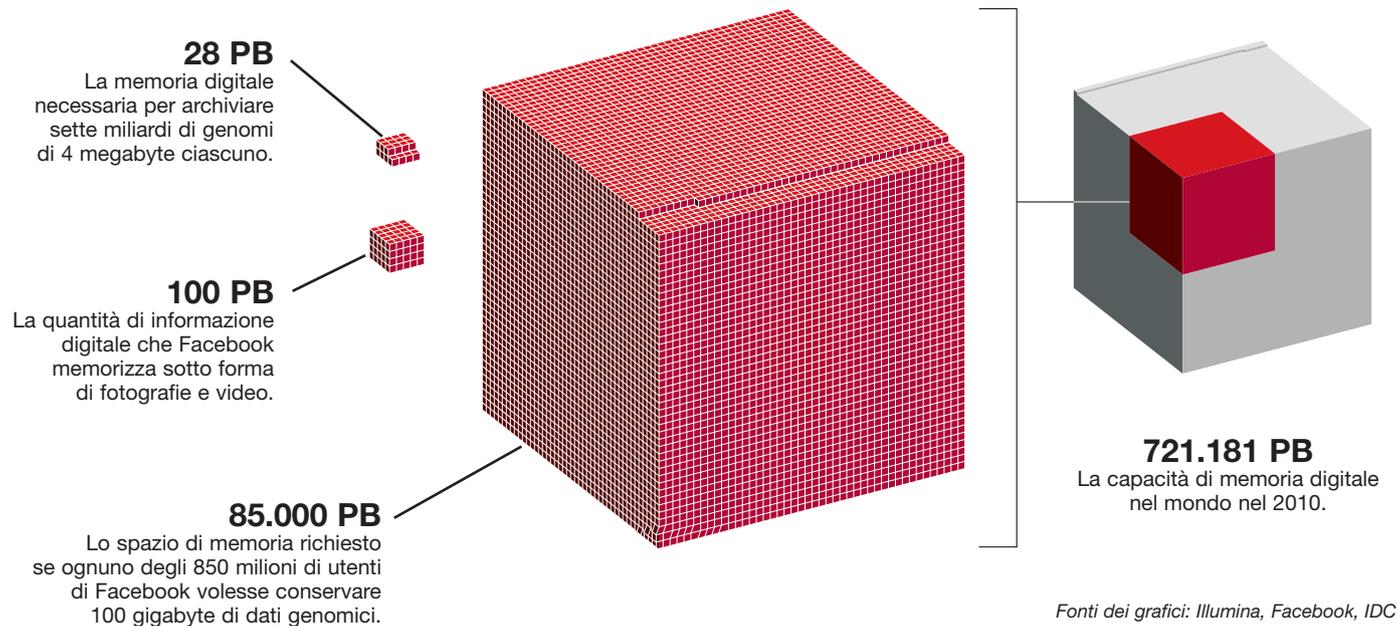
Numero approssimativo di macchine (per paese)



Fonti delle mappe: omicsmaps.com;

Come archiviare i dati?

Le dimensioni dell'informazione digitale in memoria (in petabyte).



Fonti dei grafici: Illumina, Facebook, IDC