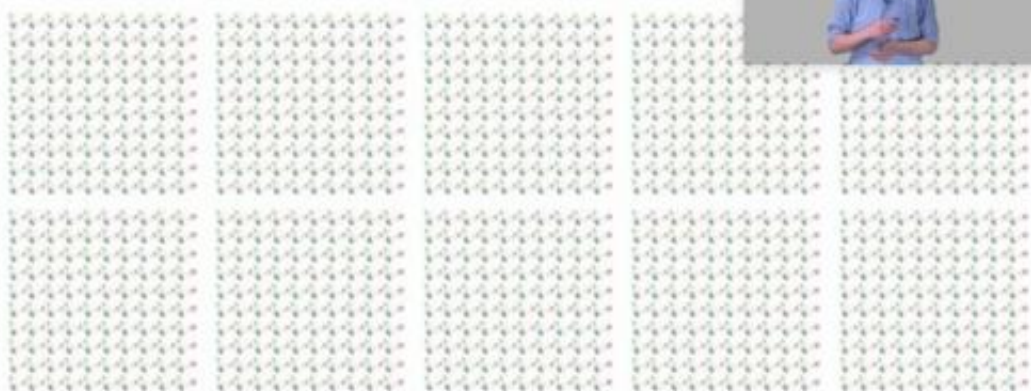


# Google making inroads with Genomics database

November 7 2014, by Bob Yirka

---

Data: Variant Calls for 1,000,000 Genomes



Last March, Google announced that it had developed a database and associated ways for accessing the data it stored, geared towards storing human genome information—named quite naturally, Google Genomics. Since that time, the company has been actively pursuing hospital and university data, so as to have as much genome data stored as possible. The overall objective, the company says is to provide a service to researchers studying genomes as a means of curing diseases, primarily cancer.

Scientists suspect that if everyone (or at least an awful lot of us) had their genome decoded and the results put into a [database](#), then finding cures to things like cancer would become possible. The hope is that by comparing the genomes of large numbers of people that get, say breast cancer, with those that don't, underlying genetic proclivities would be revealed. And if that happened, perhaps a means could be created to cause genetic changes to those with such a proclivity, to prevent it from happening to them.

Meanwhile, the cost of decoding a single [human genome](#) has plummeted in research years, which has resulted in an explosion of decodings being performed. But the data has been stored in many different places, which aren't connected together. Furthermore, storing the data from just one decoding amounts to about 100 gigabytes per person—that can add up fast. Fortunately, researchers have found a way to trim down unnecessary information, allowing a single genome to take up just a single gigabyte, which makes storing it more feasible. But, even at that smaller amount, because of the [large numbers](#) of genomes being decoded, the space requirements begin to pile up. That's where Google comes in with Google Genome—a project that combines massive storage capabilities with world class retrieval technology, courtesy of Google's Internet search engine.

Such a service doesn't come free, of course, Google charges \$25 per full [genome](#), and just twenty five cents for the reduced version. That price point, some have noted, is what research institutes would generally have to pay to store the data on their own servers—without the search capabilities. The database project has met with some success already, as some estimate that data representing thousands of genomes are already in the database, including approximately 2.6 petabyte's worth from the National Cancer Institute.

Google has also developed web based search applications, BigQuery,

MapReduce etc. but so have other third party groups such as Seven Bridges, Tute Genomics and NextCode Health, sensing profits to be made in making it easier for researchers to make connections between genomes as more and more are added to Google's database. Also, Google isn't the only one seeking to become the repository for massive amounts of human [genome data](#), Amazon too has built a database and offers ways to retrieve it in useful ways as well.

Perhaps, in the end, Google, Amazon and others will combine their data allowing researchers everywhere to access the largest number of genomes possible, giving scientists a chance at finally finding a cure for cancer and other diseases that so far have eluded science's best efforts.

**More information:** [cloud.google.com/genomics/](https://cloud.google.com/genomics/)

© 2014 TechExplore

Citation: Google making inroads with Genomics database (2014, November 7) retrieved 14 April 2024 from <https://techxplore.com/news/2014-11-google-inroads-genomics-database.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--