

# Google strikes deal to preserve DNA data online

October 27 2011, By Peter Delevett

---

Concerned that the federal government might not keep funding the world's largest free database of genetic data, Google Inc. has forged a deal with a Mountain View, Calif., startup to keep the information online - and free for researchers.

The Internet giant began talks with DNAnexus last spring, when the National Institutes of Health announced it might have to drop support of the Sequence Read Archive due to funding cuts.

The database is a vast repository for short snippets of [genetic data](#) decoded by sequencing machines, which spell out the unique combinations that make up a specific person's DNA. Researchers can compare data in the archive to look for similarities and differences between people and better unravel how genetics affects health.

While NIH officials recently announced they'd keep funding the database, "We wanted to make sure we had a Plan B," said DNAnexus Chief Executive Andreas Sundquist. Using the [Google](#) Cloud Storage service, DNAnexus will maintain a "mirror" of the public archive, together with tools Sundquist's company has developed to make it easier for scientists to search the database and share their findings.

Google officials said the Sequence Read Archive is one of the largest datasets ever deposited in Google Storage, but Sundquist predicts there will be thousands of similar databases in the future. That's due both to the growing speed and power of gene sequencers and to the decreasing

cost of storing and sharing information in the remote network of web servers known as the cloud.

"DNA sequencing becomes 10 times cheaper every 18 months thanks to hardware improvements," said Sundquist. "It's sort of like Moore's Law on steroids."

A year ago, he says, it cost about \$30,000 to sequence a person's entire DNA. Today, that number's down to \$4,000. Sundquist believes researchers eventually will sequence everyone on earth and make that data part of each person's [medical record](#).

But given that each genome is about 3 billion letters long, improvements in gene sequencing are creating a massive data management challenge, said Krishna Yeshwant, a partner at Google Ventures who's joining Sundquist's board as part of a separate \$15 million investment in DNAnexus.

The SRA database alone, he said, is hundreds of terabits in size, referring to the unit of measure for a trillion bits of computer data.

The decreased cost of gene sequencing is making it possible for genomics to move out of the research lab and into clinical settings, added Yeshwant, a Harvard-educated physician who also practices at Brigham and Women's Hospital in Boston.

"It feels like we're on the cusp of a revolution in genomics and how we think about health care," he said.

Google has a long history of investment in genomics. One of the first companies to join its venture arm's portfolio was Adimab, a New Hampshire startup that helps discover how antibodies can be turned into drugs. Google - and co-founder Sergey Brin - also has poured millions

into Mountain View's 23andme, co-founded by Brin's wife to help consumers better understand their own DNA while building a massive database for researchers to study the genetic underpinnings of disease.

Still, the head of another local genomics startup said he was underwhelmed by Google's announcement that it would create a mirror of the SRA database.

"Part of the reason it was being discontinued is that NIH prioritized it as low value," said John West, CEO of Palo Alto, Calif.-based Personalis. "It's the most comprehensive database of really raw DNA sequencing data, but it's not very organized, and it's not easy to use."

Sundquist, in fact, would agree, which is why he's hoping the data management tools his team of 25 has developed will make the repository more useful.

He says his version of the database will include a more user-friendly interface to help scientists browse, download, analyze and share data. He envisions researchers plugging data into his virtual data center, finding out if a given person's genetic mutations exist elsewhere in the [database](#) and using those discoveries to figure out how the mutation impacts health - or how a given drug may affect certain people.

Sundquist, 32, founded the company two years ago while working on a Ph.D. in computer science at Stanford. Although he had no prior background in medicine, he was fascinated by the data challenges posed by the rapid improvements in gene sequencing.

"In less than five years," he predicted, "the cost of DNA sequencing will be on par with the cost of other routine lab tests."

(c)2011 the San Jose Mercury News (San Jose, Calif.)

Distributed by MCT Information Services

Citation: Google strikes deal to preserve DNA data online (2011, October 27) retrieved 23 April 2024 from <https://phys.org/news/2011-10-google-dna-online.html>

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.