DAVID PAUL MORRIS/BLOOMBERG VIA GETTY

Processing scrambled DNA-sequence data promises safe and fast discovery of disease-linked gene variants.

COMPUTING

Cloud cover protects gene data

Extreme cryptography set to bring 'personalized' medicine a step closer.

BY ERIKA CHECK HAYDEN

The dream for tomorrow's medicine is to understand the links between DNA and disease — and to tailor therapies accordingly. But scientists working to realize such 'personalized' or 'precision' medicine have a problem: how to keep genetic data and medical records secure while still enabling the massive, cloud-based analyses needed to make meaningful associations. Now, tests of an emerging form of data encryption suggest that the dilemma can be solved.

At a workshop on 16 March hosted by the University of California, San Diego (UCSD), cryptographers analysed test genetic data. Working with small data sets, and using a method known as homomorphic encryption, they could find disease-associated gene variants in about ten minutes. Despite the fact that computers were still kept bogged down for hours by more-realistic tasks — such as finding a disease-linked variant in a stretch of DNA a few hundred-thousandths the size of the whole genome — experts in cryptography were encouraged.

"This is a promising result," says Xiaoqian Jiang, a computer scientist at UCSD who helped to set up the workshop. "But challenges still exist in scaling it up."

Physicians and researchers think that understanding how genes influence disease will require genetic and health data to be collected from millions of people. Such a massive task will probably require harnessing the processing power of networked cloud computers, but online security breaches in the past few years illustrate the dangers of entrusting huge, sensitive data sets to the cloud. Administrators at the US National Institutes of Health's database of Genotypes and Phenotypes (dbGaP), a catalogue of genetic and medical data, are so concerned about security that they forbid users of the database from storing it on computers that are directly connected to the Internet.

Homomorphic encryption could address those fears by allowing researchers to deposit only a mathematically scrambled, or encrypted, form of data in the cloud. It involves encrypting data on a local computer, then uploading those scrambled data to the cloud. Computations on the encrypted data are performed in the cloud and an encrypted result is then sent back to a local computer, which decrypts the answer. If would-be thieves were to intercept the encrypted data at any point along the way, the underlying data would remain safe.

"If we can show that these techniques work, then it will give increased reassurance that this high-volume data will be computed on and stored in a way that protects individual privacy," says Lucila Ohno-Machado, a computer scientist at UCSD and a workshop organizer.

Homomorphic data encryption, first proposed in 1978, differs from other types of encryption in that it would allow the cloud to manipulate scrambled data — in essence, the cloud would never actually 'see' the numbers it was working with. And, unlike other encryption schemes, it would give the same result as calculations on unencrypted data.

But it remained largely a theoretical concept until 2009, when cryptographer Craig Gentry at the IBM Thomas J. Watson Research Center in Yorktown Heights, New York, proved that it was possible to carry out almost any type of computation on homomorphically encrypted data. This was done by transforming each data point into a piece of encrypted information, or ciphertext, that was larger and more complex than the original bit of data. A single bit of unencrypted data would become encrypted into a ciphertext of a few megabytes — the size of a digital photograph. It was a breakthrough, but calculations could take 14 orders of magnitude as long as working on unencrypted data. Gentry had rendered the approach possible, but it remained impractical.

Since then, cryptographers have developed systems to address these issues, for instance by encrypting many pieces of data together so that they can be processed in parallel, or by encrypting real numbers directly into single ciphertexts, rather than first converting them into bits. As a result, homomorphic encryption now runs 150,000 times faster than it did in 2009, says Shai Halevi, a cryptographer at the IBM research centre. "The same calculation that took a day-and-a-half in 2012 now takes us five minutes to do," he says. "Now is the time to ask, is this fast enough to be usable?"

At the 16 March iDASH Privacy & Security Workshop 2015 (iDASH stands for Integrating Data for Analysis, Anonymization and Sharing), five teams revealed homomorphic encryption schemes that could examine data from 400 people within about 10 minutes, and that could pick out a disease-linked variant from among 311 spots at which the genome is known to vary. It took up to 30 minutes to similarly analyse small chunks of genome a little larger than the size of a typical gene, about 5,000 DNA base pairs. For longer stretches of sequence data — 100,000 base pairs, or about 0.003% of the overall genome — analysis was not always possible, or took hours, and consumed up to 100 times more memory than computing unencrypted data. Even so, cryptographers say that the results indicate major progress: "Our challenge shows this is

not impossible, compared to three years ago, when people were thinking this computation was infeasible," Jiang says.

But some data custodians remain sceptical about encryption. Steven Sherry, chief of the reference collections section at the US National Center for Biotechnology Information in Bethesda, Maryland, manages dbGaP. He says that cryptography, even if it worked, would not necessarily protect data on researchers' computers or give them enough analytical flexibility. He instead favours restricting access to a small circle of scientists and asking them to certify that they will abide by rules and regulations on how the data can be used. "We haven't looked at cryptographic methods," says Sherry, "because it hasn't been demonstrated to us that they're both secure and useful." ■

SPACE

Bright spots hint at active ice on Ceres

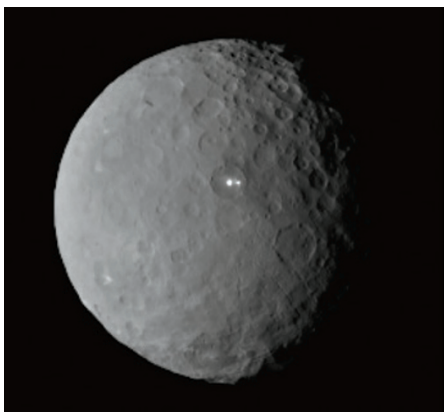
Early data from Dawn spacecraft bring scientists closer to clearing up mystery about dwarf planet.

BY ALEXANDRA WITZE, THE WOODLANDS, TEXAS

A pair of bright spots glimmering inside an impact crater on the dwarf planet Ceres, mystifying scientists, could be coming from some kind of icy plume or other active geological feature.

Images from NASA's Dawn spacecraft show the spots, known as 'feature number 5', at changing angles as the dwarf planet rotates into and out of sunlight. The pictures reveal the spots even when they appear near the edge of Ceres, when the sides of the impact crater would normally block the view of anything confined to the bottom. That something is visible at all from that angle suggests that the feature must rise relatively high above the surface.

"What is amazing is that you can see the feature while the rim is still in the line of sight," said Andreas Nathues, a planetary scientist at the Max Planck Institute for Solar System Research in Göttingen, Germany. Nathues, who leads the team for one of the Dawn cameras, revealed the images on 17 March at the Lunar and Planetary Science Conference in The Woodlands, Texas.



NASA/JPL-CALTECH/UCLA/MPS/DLR/IDA

The Dawn spacecraft captured this image of Ceres' twin bright spots on 19 February.

At dawn on Ceres, feature number 5 appears bright. By dusk, it seems to fade. That could mean that sunlight plays an important part — for instance, by heating up ice just beneath the surface and causing it to blast off some kind of plume or other feature.

Ceres, one of the largest unexplored worlds in the Solar System, is believed to be at least one-quarter ice — a greater proportion than most asteroids. Dawn, which launched in 2007, aims to work out where that ice resides and what role it has in shaping the dwarf planet's surface. One idea is that the ice is blanketed by a very thin layer of soil. The ice may occasionally squirt up in towering 'cryovolcanoes', thanks to internal pressures in the asteroid.

Dawn is currently looping back towards Ceres after being captured by its gravity on 6 March. As the spacecraft gets closer to the dwarf planet, it will take more pictures to see how its surface might be changing. "The big question is whether Ceres has an active region — or more than one," Nathues says.

Christopher Russell, a planetary scientist at the University of California, Los Angeles, and Dawn's principal investigator, says that towards the end of its mission the spacecraft will map Ceres at high enough resolution to see features that are just 30 metres across. The hope is that the possible icy plume will come into focus and reveal its true nature.

"We hope to show that Ceres is every bit a planet, as much as its terrestrial neighbours Mars, Earth, Venus and Mercury are," Russell says. ■