THIS WEEK

EDITORIALS

FOR SALE Cash crisis threatens future of London's Royal Institution **p.452**

WORLD VIEW Poor world needs more than simple health solutions **p.453**

EARTH MOVE Marcia McNutt to leave the US Geological Survey **p.457**

Genetic privacy

The ability to identify an individual from their anonymous genome sequence, using a clever algorithm and data from public databases, threatens the principle of subject confidentiality.

How private is private? A study published on 17 January reveals vulnerabilities in the security of public databases that contain genetic data, the latest in a series of similar revelations. So far, research funders that host the databases have responded to such problems on a case-by-case basis, but it is now clear that the research community as a whole must devise a more comprehensive approach.

In the latest study, led by Yaniv Erlich at the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts (M. Gymrek *et al. Science* **339,** 321–324; 2013), researchers showed that they could discover the identity of some men whose genomes had been sequenced as part of a genomics project (see *Nature* http://dx.doi.org/10.1038/nature.2013.12237; 2013). Erlich's team wrote an algorithm that infers an individual's pattern (a haplotype) of genetic markers called short tandem repeats from the nucleotide sequence of his Y chromosome. The team then searched genealogical databases for the names of men with corresponding Y-chromosome haplotypes. The team confirmed the correct names by cross-referencing the possible last names with public records of people of similar ages and locations.

Using this strategy, the team was able to confirm the identity of known individuals whose genomes have been sequenced, such as genomics entrepreneur Craig Venter, and to discover the identities of anonymous research subjects, including five men who participated in both the 1000 Genomes Project and a study of Utah Mormons initiated by the Centre for the Study of Human Polymorphism (CEPH) in Paris. Erlich's team was also able to discern the identity of some of the study subjects' family members, because family pedigrees were collected as part of the CEPH study.

It is important to note that the CEPH cohort is particularly suitable for this method of identification, because of the volume of informative data that has been collected and published about CEPH participants. Their family pedigrees, the places where they lived and their ages at the time of the data collection are all public information. Or at least they were until the US National Institute of General Medical Sciences, part of the National Institutes of Health (NIH), responded to Erlich's study by removing participants' ages from public view on the Human Genetic Cell Repository website that it funds.

It would probably be more difficult to use Erlich's method to identify participants in studies lacking extensive demographic information. And Erlich responded in an exemplary way to his team's findings by contacting the NIH and other genetics researchers with his findings before publishing them. This sets an important precedent for constructively dealing with newly discovered privacy loopholes, and other researchers should take note. Erlich's team is also not publishing the names of the anonymous study participants whose identities they uncovered.

How the genetics community addresses these issues is crucial to how large-scale genetic studies will proceed. Although research participants are already sometimes told that their data might not remain private — as the CEPH study participants were — the fact that their identities could be revealed would seem a remote risk to them, as that has only recently become possible. It is now imperative that participants fully understand that it is unlikely that their identities can be kept hidden if their genetic data are revealed. Some participants might welcome this, such as those with an interest in genealogy. Others — perhaps those with stigmatized diseases, for instance — might not.

Moving data behind a controlled-access barrier lessens their utility to science and to society at large. But researchers need to show the public that they are acting as careful stewards of the data entrusted to them. Erlich argues that the solution is to make sure that participants understand what they're signing up for, and to adopt laws that adequately protect people against the misuse of their genetic information.

"Researchers need to show the public that they are acting as careful stewards of the data entrusted to them."

Geneticists are brainstorming other proposals for balancing data sharing with the need to protect the privacy of research subjects. One is to move more data behind a controlled-access barrier, but to authorize trusted users to access the data from many studies, rather than having to obtain it piecemeal from different studies, as researchers must do today. There are logistical barriers to this — for instance, ensuring compatibility across databases. And it is debatable whether such restrictions might do more harm than good.

But if controlled access is not the right solution, it is up to the research community, in consultation with the public, to devise a better one. A solution should come sooner, rather than later, because this latest revelation of a privacy loophole will be far from the last. ■

Vigilance needed

Experiments that make deadly pathogens more dangerous demand the utmost scrutiny.

The year-long voluntary moratorium on research to engineer strains of the H5N1 avian influenza virus that can transmit between mammals has already borne fruit. Claims of public-health benefits have received thorough scrutiny, and the researchers involved have better explained the biosafety and biosecurity precautions that they take. The debate has drawn attention to, and exposed gaps in, the rules that govern 'dual-use' research — work that can bring public benefit but might also be used for harmful purposes. The row has also, for example, prompted long-overdue national guidelines in the United States and made funders everywhere more aware of the need to assess