



Be prepared for the big genome leak

It is only a matter of time until idealism sees the release of confidential genetic data on study participants, says **Steven E. Brenner**.

Most people in the United States could soon know someone whose genome is held in a research database. Concerns are growing about our ability to properly control access to that information. Also growing among some scientists is the feeling that restricting access to genomic data fetters research. How long will it be until an idealistic and technically literate researcher deliberately releases genome and trait information publicly in the name of open science?

Both the open-access literature and the open-source software movements began with idealists. It seems inevitable that there will be a major leak of genome information in the near future. Individual scientists, institutions and funders should consider now how they will react when this happens.

Some studies already gather the genetic data of more than 50,000 individuals in a single analysis. Although this information is supposed to be highly protected, it is disseminated to various institutions that have inconsistent security and privacy standards. In practice, data protection often comes down to individual scientists. Once leaked, these data would be virtually impossible to contain.

What harm would come from a leak of personal and genomic data? The consent form for the Personal Genome Project (PGP) — which makes no attempt to keep genetic information secret — offers a guide. It lists a range of adverse consequences, from revealing non-paternity to being framed with synthesized DNA planted at a crime scene.

Most research genome data are de-identified, but given progress in re-identification and commercial genetic databases, will they stay that way? De-anonymized genomic data would be most likely to reveal health conditions relevant to the study for which they were collected. The effects might be uncomfortable but would probably reveal less than a typical Google search history. So far, no PGP participant who released genomes and traits has experienced adverse consequences that have been reported to the Institutional Review Board. In the longer term, the risk of harm may rise as our understanding of genetic variation increases.

Then there is the public outcry a genome breach might incite. The public often has an exaggerated perception of the links between genes and personal traits. Lacking contextual information, research participants could wonder whether their own genomes had been leaked and dread implausibly dire consequences.

Thus a genome leak might lead to a backlash. Volunteers might withdraw from research studies and refuse to join new ones. Research might even be subject to moratoriums and prohibitive restrictions. The harm to genetic research could be great, and study participants could be unsettled.

What can be done? Two extreme options offer appealing simplicity. One is for research projects to incorporate unrestricted data release from the outset. This option should be offered more broadly owing to the certainty and research benefits it offers. However, would enough people be willing to share so openly? The second option would be to lock down genomes so tightly that they are virtually impossible to steal, for example by only allowing analyses on central computers through restricted interfaces. Although useful as an alternative, this system would stymie research were it to become the exclusive means of access to data, but it would still remain vulnerable to ingenious ways of eliciting inappropriate genetic information.

Neither option is comprehensively workable, which means that the question is not how to prevent a leak but how to mitigate the fall-out. This requires some specific steps, as well as progress in adapting

concepts already used elsewhere in biological research and in applying principles proposed by groups such as the Presidential Commission for the Study of Bioethical Issues in Washington DC.

Funders should develop rapid mechanisms for notifying study participants, governments and the media when breaches occur and provide informed guidance about scope and probable consequences for those affected. This would require recontacting research participants to warn those whose data were leaked and, implicitly, to calm others whose data remain secure. More research is needed about the possible harm of such leaks to better inform and protect research participants before and after leaks occur.

We should also take steps to minimize the frequency and extent of future genome leaks. Institutions could establish uniform protocols and reviews to ensure the safety of protected genomic data. All researchers using restricted genomic data should be trained regarding the ethics of and the technologies involved in protecting human data. Technical and legal strategies should be proactively deployed to help limit dissemination of leaked data to those who furtively hunt for them.

Augmented legal protections could reduce the harm from inappropriate use of such data. In the meantime, we need to address a quandary: research with leaked data would undoubtedly speed immediate scientific progress, but should scientists exploit them?

Most importantly, we must ensure that the necessary discussion about the risks of a genome leak is balanced with information about the tremendous benefits that collected genetic information has for all of us. Although the acceleration and promise of genomics makes a leak inevitable, it also guarantees medical progress. ■ [SEE EDITORIAL P.137](#)

THE QUESTION IS NOT
HOW TO PREVENT
A LEAK
BUT HOW TO
MITIGATE THE
FALL-OUT.

➔ [NATURE.COM](#)
Discuss this article
online at:
go.nature.com/oybzqm

Steven E. Brenner is a Professor at the University of California, Berkeley.
e-mail: brenner@compbio.berkeley.edu