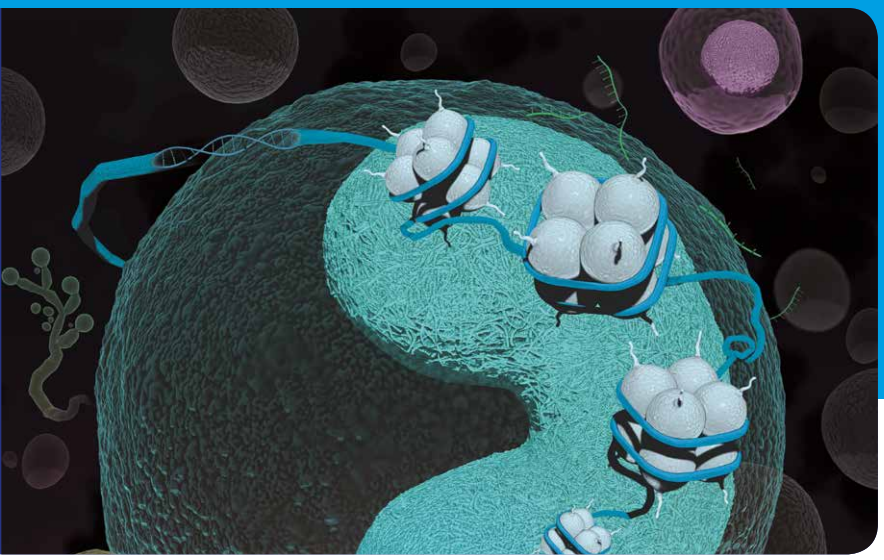


BLUEPRINT epigenome

Balancing accessibility and privacy for epigenomic data

Individual epigenomes hold great promise for medicine - if the data can be widely used without encroaching on privacy

Rapid technological progress and sharply falling costs have made it feasible to obtain genome-wide information from individuals. DNA sequences and epigenomic data promise to provide new insights into our health and the mechanisms that cause disease, particularly if they are made widely accessible to researchers.

Before such access can be granted, methods are needed to protect the privacy of those who have contributed samples. Until now, it has been unclear how difficult it might be to deduce an individual's identity from epigenomic data - which needs to be resolved before such information can be made widely available to scientists.

A work group of the International Human Epigenome Consortium (IHEC) including a number of BLUEPRINT partners has now addressed this question in a study published in the July 17 issue of Genome Biology. The work proposes a new set of guidelines to help protect identities while granting access to data obtained from research participants.

Every human (with the exception of identical twins) has a unique DNA sequence, which has made it straightforward to identify a person by matching one sample to another or to information in a database. The situation is somewhat different for epigenomic data, which captures chemical modifications to DNA sequences – or proteins bound to them – over a person's lifetime. Generally speaking, an individual has one genome that produces many epigenomes, which reflect changes in different types of cells during the process of aging, over the course of a disease, and throughout other events related to a person's environment and lifestyle.

So depending on what type of epigenomic information has been collected, matching a new sample to a second one taken from the same person or to an entry in a database probably wouldn't be as easy as comparing sequences. The scientists carried out a test study to investigate the feasibility of doing so, and also investigated whether unique sequences in a person's DNA were carried over into the epigenome. In other words: could the genome sequence be recovered from epigenomic data, in a way that would allow an individual to be identified?

Another issue had to do with the fact that interpreting and using epigenome data requires extra information - such as people's ages or ethnic backgrounds - that might also provide hints as to their identities. If so, extra measures might need to be taken when giving access to the information. That is also true for data from individuals with rare diseases, which sometimes have unusual effects on epigenetic processes and are studied in hopes of understanding the mechanisms that control them.

The study revealed that in many cases, it was theoretically possible to recover the identity of an individual by matching sets of epigenomic data. In practice, the amount of effort and the feasibility of doing so would depend on many factors that are hard to determine.

As a general rule, the authors state, the measures that are taken to protect the data should be determined by two factors: the likelihood that a person could be identified and the amount of harm that might arise from this happening. Since some types of data were easier to match, and discovering some types of information about a person might be truly distressful, this implied that different strategies would be needed to secure them. The authors propose guidance for sharing open access epigenome data in a way that reduces the risk of reidentification.

One problem is that scientists rarely can predict how the data they collect may ultimately be used, or the ways this might affect privacy issues. In most cases it is impossible to ensure an absolute long-term protection of information, and to claim otherwise to research participants would be a misrepresentation. The authors advise clear, transparent education for patients about the potential uses of data and presenting a balanced impression of the real risks of misuse.

The same issues have arisen in the context of other sequencing projects and genetic testing. Many countries have adopted anti-discrimination laws that protect patient data in these types of studies, but it is often unclear how they will apply to epigenomic information. While there have been very few cases of abuse, rare events have generated substantial media coverage that likely inflates public perception of risk. When that happens, researchers have found measurable drops in the willingness of potential subjects to participate in data-collection projects.

Several solutions to these problems are being developed by IHEC and BLUEPRINT members. One is to find better ways to control access to DNA sequence data from patients in epigenomic studies, which is easier to link to individuals than epigenetic information itself. If this is done, the authors estimate that re-identifying patients from epigenetic studies would require an enormous effort.

More efforts are being directed at improving methods of educating participants and obtaining informed consent.

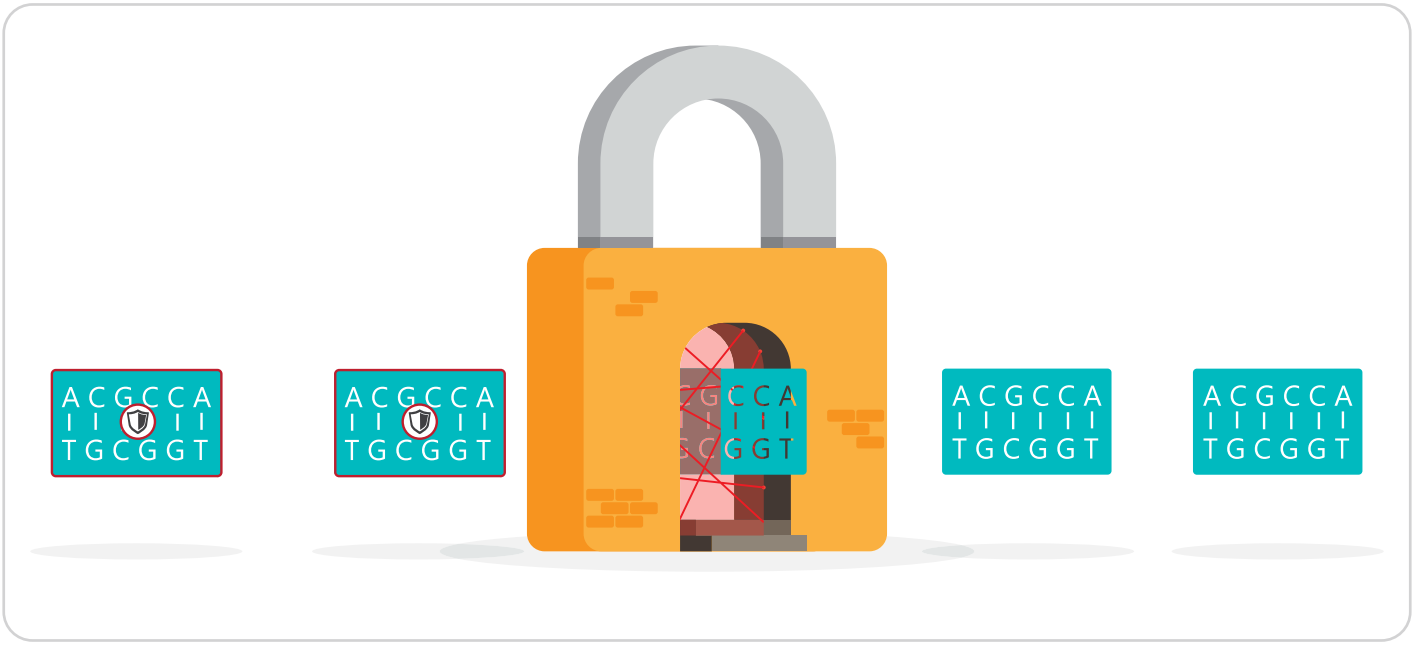
The issue of ensuring privacy is likely to arise again and again as scientists collect more types of data and use it in new ways to discover how genomes and lifestyle factors contribute to disease. Generally, the authors state, „Society may have to re-conceptualize and contextualize medical confidentiality and personal privacy so that they remain relevant in the context of IT developments and the sharing of health information through social media and the World Wide Web."

Reference:

- *Dyke SO, Cheung WA, Joly Y, Ammerpohl O, Lutsik P, Rothstein MA, Caron M, Busche S, Bourque G, Rönnblom L, Flicek P, Beck S, Hirst M, Stunnenberg H, Siebert R, Walter J, Pastinen T. Epigenome data release: a participant-centered approach to privacy protection. Genome Biol. 2015 Jul 17;16:142. doi: 10.1186/s13059-015-0723-0.*

Direct link to full text of the article: www.ncbi.nlm.nih.gov/pmc/articles/PMC4504083

www.blueprint-epigenome.eu



To protect the privacy of patients who have contributed samples is crucial in epigenomic research.