

# Privacy, anonymity and subjectivity in genomic research

## Summary

The use of non-anonymized human genome data is becoming increasingly popular in research. Here we review the proceedings of a special meeting on this topic that took place at European Molecular Biology Organization (EMBO) in December 2014. The main points discussed centered on how to achieve ‘anonymity,’ ‘trust,’ and ‘protection of data’ in relation to new genomic technologies and research. Following our report of this meeting, we also raise three further issues for future consideration: the harmonization of international law in relation to genetic data protection; the complex issues around the ‘dividual’ nature of genetic data; and the growing commercial value of personal data. In conclusion, we stress the importance of scientists working in the area of genomic research engaging in interdisciplinary collaborations with humanities and social science scholars and addressing these complicated issues.

## Introduction

The European Science Foundation sponsored a workshop on the use of non-anonymized human genome data in research, which met in Heidelberg, Germany, on 10th and 11th December 2014 (European Science Foundation, 2014). The meetings were located in the European Molecular Biology Organization (EMBO) building, next to the European Molecular Biology Laboratory, lending an historical weight to the presentations of the 14 participants, who came from 12 countries. The core aim of the meeting was to evaluate the current state of knowledge relating to anonymity and privacy in genomic research, and to identify areas for development in respect to both the technical science and the social policy.

In this article, we have two aims: to report on the main issues discussed at the Heidelberg meeting, highlighting the key suggestions for further consideration; and to raise three additional issues of our own for future studies. This report is a collaboration between the authors, Noam Shomron (*Genetics Research*, Editor-in-Chief), who attended the Heidelberg meeting and contributed to the proceedings, and Ian McGonigle, who is a neuroscientist and anthropologist working on issues in bioethics and science and society. This article is therefore both a report of the event proceedings and an effort to continue the conversation and raise further issues for bioethical consideration. These perspectives should be of interest to practitioners and analysts alike.

## Heidelberg report

### *Protection of data*

The first issue discussed at Heidelberg related to the dissemination of genomic data outside of the initial

research setting, in other words, passing on data to third parties. Issues of privacy and disclosure were discussed, especially in relation to publicly available resources such as the Personal Genome Project (PGP) databases. Further issues raised were around ‘civil rights’ and the law, specifically the ‘right to know’ genetic analyses, as compared to the ‘right not to know.’

From the technical side, technologies for sequencing and storing human genomic data and for analyzing genetic information are rapidly increasing in speed and power. These technical advances, however, necessitate appropriate governance and ethical policies, so that individuals and groups can be sufficiently informed about the stakes, and protect their genetic privacy accordingly. A major challenge, and an important ethical consideration in the development of personalized medical models, however, is the establishment of databases that couple genetic and phenotypic data (clinical information about the person), which at this point is considered sensitive data. This raises questions as to how to protect the genetic privacy of volunteers consenting to their personal sensitive data. Moreover, the significance of those data may change as technologies and analytics capacities increase in power, making it important to have long-term security measures in place for data stored.

### *Anonymity*

Facing the possibility of only being able to work with participants’ genetic data, the conference participants questioned whether researchers could proceed much further without more useful phenotypic information to amalgamate their analyses, or whether they have reached the ‘end of the line’ with just genomic data.

The idea of anonymizing research subjects as a way to circumvent ethical issues of privacy was discussed, and it was thought that more stakeholders would need to be included in such decisions and policies.

Genetic data may hold different levels of importance in different contexts. Some communities have a low threshold for sharing their clinical and genetic data. For example, some patients might desire to share their data to help further medical research, especially in cases of rare genetic diseases where personal family history is involved. Such cases may in fact see patients insist on contributing information that is more personal in order to enhance the genomic data rendered. It was noted that with time, major technical developments might lead to further issues requiring ethical analysis. If sequencing, for example, moves from sequencing DNA to sequencing RNA, this would require re-evaluation of the information yielded from each technology, and also the degree of vulnerability of each dataset. This might also refer to modifications on the DNA itself, such as chemical modifications (e.g. methylation).

### *Trust*

Trust was raised as a key theme of discussion at the meetings. Trust, however, is not only important between researchers and participants, but also between governing bodies and scientific communities. Developments in data encryption and privacy legislation could help in this regard, but good governance would require a meeting of a wide range of stakeholders.

As genomics shifts towards personalizing the field, our ideas about openness, privacy and anonymity, may, however, change. Patients may start to come to the clinic or to research partnerships already with more genomic knowledge than the researchers, investigation will produce (they may have sequenced their own genome already), raising the issue of the symmetry of the levels of information and power between researcher and participant. Though we see a rise in doctors, genetic counsellors and for-profit companies interested in genomic data, it is still not clear how much solid, actionable, medical truth can be derived from the technologies presently available. With the harmonization of databases, however, the interpretive power of genomics will likely improve.

At this point we would like to conclude our report on the Heidelberg meeting and present three additional important concerns that should be considered when forming policy or governance structures in relation to the protection of genetic data. These are legal harmonization across states; the 'dividual' (shared, divided, distributed) nature of participants' personal data; and the emerging valuation/commodification of genetic data. These points relate to the bigger

picture of the development of science and society, but they are also important for consideration by practitioners.

### **Issues for further consideration**

#### *Legal harmonization*

First, the legal protections afforded to participants' genetic data could become compromised if, for example, there are criminal charges against a subject, and where a prosecuting body, like a state, wants to acquire the genetic data of the subject. The question that arises is whether genetic databases or genetic information used in research are subject to a legal disclosure by subpoena, or whether genetic data given in confidence can be guaranteed legal protection from outside agents. Further, in cases where foreign states petition for the genetic data of another state's citizen, questions about the harmonization of international law arise.

#### *Genetic dividuality*

An interesting question arises in relation to the proposition that genetic data might be considered 'individual personal property', which can be legally protected as such. Since humans gain their genetic signature through biological inheritance from two parents, which they share with siblings, much information about an individual can be extrapolated by examining the genetic data of close relatives. The fact that genetic data, or metadata, could be easily acquired by investigating a person-in-question's close relatives challenges the notion that genetic data is 'individual' in any meaningful way. Rather, it would seem that personal genetic data is precisely 'dividual' in nature. Dividuality is an anthropological concept emerging from the study of kinship that describes the intersubjective nature of personhood in contextualized social relations. In relation to genetic personhood and notions of the limits of personal privacy, human genetic personhood and identity might be better considered as being 'dividual' in nature, rather than individual, in the sense that genetic data are usually partially shared with close kin, who may also share relevant family, health and life experiences. The proposition that genetic data, and the associated personal medical and personal data, are precisely 'dividual' in nature might be considered when developing ethical standards, legislation and governance structures. Indeed, legal 'individual' citizens should recognize that when they disclose their personal genetic data publicly they also inadvertently share data about their biological kin. The disclosure of genetic data may therefore entail damages to related individuals who could suffer as a consequence. This

potentiality raises more complex questions about collective consent, responsibility to disclose or restrict data, and the limits of personal and family privacy, of course.

### *The value of data*

We are facing the real prospect that very soon personalized, precision, medicine will be a viable option for many patients. Indeed, with advances in the speed and ease of complete genomic sequencing, and in-the-clinic sequencing of other molecules and states (RNA, methylation of DNA and more), it will likely be possible to make better diagnoses and design more effective, tailored, treatments for patients. This development will impact the commercial value of data. Individuals with certain genetic markers could be identified in advance as a potential customer for certain drugs or therapies. And while this kind of tailored, targeted, treatment could improve outcomes and prolong healthy life, it also could entail more focused direct-to-consumer marketing of medications and therapies, particularly in relation to long-term prophylactics, such as drugs for treating lifestyle diseases (such as hypertension, hypercholesterolemia or obesity), or indeed common psychiatric drugs (such as antidepressants, anxiolytics and drugs for attention deficit disorder). Populations that have volunteered their data could easily become potential customers to pharmaceutical companies, and as such, they may want to be able to restrict the access other agents have to their genetic data. On the other hand, as bearers of valued data, they may also wish to capitalize on this. By sharing their genetic data with insurance companies, for example, individuals could potentially benefit from being classed at a lower risk for some conditions, and may perhaps even benefit from a lower insurance premium. Alternatively, by sharing data when participating in clinical trials or regular therapies, patients may be in a position to be remunerated, or treated at a reduced cost.

From the perspective of consumers wishing to extract value from their personal data, however, it should be noted that the public has been paying billions of dollars in annual health insurance policies, thus allowing the companies to become large and influential. It is arguably the responsibility of the companies to reciprocate and pay back something to the community, perhaps by sponsoring data sharing initiatives. This could be seen both as a pro-active step to prevent diseases, to help detect them early on, and to categorize patients at risk and carefully monitor their health. In fact, an insurance company in South Africa and the UK made the pioneering move and started offering subsidized genetic tests to its policyholders, as long as they opt to take actionable decisions that can defer potential sickness

(<http://www.humanlongevity.com>). This might be a way that insurance companies and their customers can begin to work together in the future to improve human health outcomes and lower disease risks.

Such potential in data sharing and personal genetic medicine will probably lead to the development of algorithmic systems that can measure the relative value of the data relating to specific genes, groups of genes or RNA. With the huge market value that is created by these sequences and their complex relations, the importance of further research on privacy, protection and anonymity in relation to genomic research only becomes greater.

### **Conclusion**

Here we have presented the key points of discussion at the 2014 European Science Foundation meeting at Heidelberg regarding the use of non-anonymized human genome sequence in research, which centered around how to achieve ‘anonymity,’ ‘trust,’ and ‘protection of data’ in relation to new genomic technologies and research. We have summarized the discussions and we have also raised further issues for consideration in future research. These are international legal harmonization in relation to genetic data protection; the complex issues around the ‘dividual’ nature of personal data; and the growing commercial value of personal data.

In conclusion, we stress the importance of scientists working in the area of genomic research to engage in collaborative research activities with bioethicists, social scientists, philosophers, lawyers, humanities scholars, historians of science and medicine, and relevant others, as these complicated issues necessitate interdisciplinary efforts.

### **Declaration of Interest**

None.

IAN MCGONIGLE\* AND NOAM SHOMRON\*  
*Faculty of Medicine, Tel Aviv University,  
Tel Aviv, Israel*

\*Corresponding authors: Noam Shomron,  
[nshomron@post.tau.ac.il](mailto:nshomron@post.tau.ac.il) and Ian McGonigle  
[mcgonigle@fas.harvard.edu](mailto:mcgonigle@fas.harvard.edu)

### **Reference**

European Science Foundation (2014). Exploratory Workshop on Use of non-anonymized human genome sequence in research: Science and Policy, Heidelberg, Germany, 10–11 December 2014. Available at [http://www.embo.org/documents/science\\_policy/non-anonymized\\_human%20genome\\_sequence\\_ESF\\_EMBO\\_2014\\_workshop\\_report.pdf](http://www.embo.org/documents/science_policy/non-anonymized_human%20genome_sequence_ESF_EMBO_2014_workshop_report.pdf) (accessed 18 November 2015).