

Comment



A Roma woman on her balcony in an apartment building in Kosice, Slovakia.

SEAN GALLUP/GETTY

Europe's Roma people are vulnerable to poor practice in genetics

Veronika Lipphardt, Mihai Surdu, Nils Ellebrecht, Peter Pfaffelhuber, Matthias Wienroth & Gudrun A. Rappold

Analysis of how papers and databases are handled and interpreted shows that geneticists in Europe must stamp out unethical research practices at home, not just abroad.

In the past few years, several media and scientific reports have raised awareness about unethical uses of DNA databases. Perhaps the most alarming is the Chinese government's use of DNA to monitor the Uyghur minority ethnic population, which is predominantly Muslim, in Xinjiang province.

Yet problems with DNA databases are more widespread and entrenched than many geneticists either realize or want to acknowledge.

For many samples, either there is no record of consent being obtained from individuals whose DNA was collected, or the procedures used to obtain consent were inadequate. This applies to numerous studies involving Indigenous communities, including Australia's Aboriginal and Torres Strait Islander people,

Native American communities in the United States and the San people in southern Africa¹. Moreover, people often have little or no say in how their DNA will be used, and rarely benefit from the studies¹.

Now, our analysis of several hundred publications and five databases points to multiple issues with the handling and interpretation of DNA data from Roma people. The Roma are the largest minority group in Europe.

In our view, research and peer-review practices must change across a broad array of disciplines, from forensic genetics to molecular anthropology. Failure to correct past and ongoing mistakes puts more people at risk of harm from the collection of DNA. It also

threatens the reputation of human genetics – and of science in general.

The Roma in Europe

Around 10 million to 12 million Roma people currently live in Europe. The term Roma was introduced in the 1980s to replace labels such as ‘Gypsy’ or ‘Zigeuner’ (used in Germany) – words perceived in many European countries to be extremely insulting. Here, we use ‘Roma people’ to describe individuals who self-define as Roma or who are referred to as Roma by the European Union and European nation states. Yet we acknowledge that the term is problematic and can have reifying effects².

The same holds for the history and ethnicity ascribed to this group. Many scientists claim that the ancestors of Europe’s Roma originated in India, and that Roma people have largely remained genetically isolated for the past 300–600 years^{3–5}. Many Roma people, however, do not see themselves as having a separate ethnicity from Europeans, and today their ancestry, cultural practices and history are extremely diverse⁶.

We chose to focus on genetic studies of Roma people because they have suffered from social discrimination for hundreds of years. Between 1935 and 1945, hundreds of thousands of Roma people were deported, sent into forced labour or killed⁷. Today, in many countries of the European Union, particularly Bulgaria and Slovakia, many Roma people live in segregated settlements. They have lower incomes and less access to quality education, housing, food and health care than does the rest of the population⁸. In 2016, one in three Roma people across nine EU member states lived without drinking water, and one in ten lived without electricity⁹.

DNA has been collected from thousands of Roma people across Europe, mainly since the 1990s. (The collection of blood samples from which DNA could in principle be extracted goes back to the 1970s.)

Over the past 5 years, we have assessed more than 450 papers, published between 1921 and 2021. Roughly two-thirds of these publications appeared in the past three decades. We have also checked DNA data from Roma people in five public databases. These are the Y-STR Haplotype Reference Database (YHRD; a Y-STR, or short tandem repeat, is a repeated nucleotide sequence on the Y chromosome), the Allele Frequency Net Database (AFND), the Allele Frequency Database (ALFRED), the Estonian Biocentre Human Genome Diversity Panel (EGDP) and the European DNA Profiling Group’s Mitochondrial DNA Population Database (EMPOP). In the case of EMPop,

users must register before they can access the data, but they are required to provide only an e-mail address, name and affiliation. YHRD and EMPop are also accessed by (but not owned by) law-enforcement agencies.

Our aim was to better understand how geneticists, medical researchers and molecular anthropologists, among others, have obtained this DNA. We also wanted to interrogate how researchers have conducted and interpreted their analyses. As part of our analysis, we interviewed and e-mailed 10 researchers, 3 ethics committees and 13 research and funding institutions and journal editors about their methods and policies. Throughout our study, we sought guidance from Anja Reuss, a political adviser at and spokesperson for the Central Council of German Sinti and Roma, an advocacy group based in Heidelberg.

Consent and labels

In many cases, especially in the late twentieth century, samples have been collected from people (including prisoners) without adequate consent or any record of consent, then shared across research groups or deposited in public databases. In others, participants seem to have given some kind of consent, but it is unclear whether they understood exactly what their DNA would be used for. From two inter-

“Geneticists in Europe need to face up to the fact that unethical research practices are happening on home soil.”

views with geneticists, we even learnt that, in some medical studies, various incentives were offered to Roma people – a practice considered unacceptable by most human geneticists. Participants, who in some cases gave only their spoken consent, were told that their data would reveal whether they were carriers of genetic diseases – but not that their genetic information would end up in public databases (such as EMPop and YHRD) that can also be accessed by law-enforcement agencies, which is what happened in some cases.

In other studies, Roma people were recruited by medical practitioners who gave individual data sets an ethnic label and then shared people’s personal data with researchers. Such secondary usage violates the ethical norms set out in Article 31 of the United Nations Declaration on the Rights of Indigenous Peoples, as well as the research regulations and legislation of the European Union and several countries,

including Switzerland, the United States and Canada.

In tens of publications on the genetics of Roma people, researchers use words such as ‘Gypsies’, ‘inbred’ or ‘consanguineous’, or refer to Roma people as a ‘genetic high-risk group’. For Roma people, these are disrespectful and pejorative terms in themselves. Moreover, such broad extrapolation is stigmatizing for such a large population. Use of these terms has declined over the past ten years, but ‘Gypsy’ continues to be used in a few academic publications. One of the public databases we looked at removed this label only in 2020.

Even the methodological approaches used in many of these studies are questionable. Recruiting individuals from the most isolated communities or from the patient registers of medical geneticists runs the risk of biased sampling, which distorts scientific results. In fact, using various methods, researchers have frequently tried to avoid sampling people whom they consider to have ‘non-Roma’ and mixed ancestry¹⁰. Even today, some researchers remove individual data sets if an analysis indicates mixed ancestry. This might be appropriate for certain research questions concerning a specific community. But, often, such data are used to support claims made about all European Roma people.

Forensic genetics

Perhaps most problematic is the use of these data in forensic genetics research.

Only since 2010 have leading forensic genetics journals required publications to include evidence of appropriate procedures, such as the use of written informed consent or approval from an ethics committee (see, for example, refs 11,12). Yet data collected even decades earlier continue to be widely used. Also, if the police or military forces have helped to collect them, the data might not be published in a journal at all – and so not be subject to editorial checks. A German law-enforcement institution, the Baden Württemberg State Office of Criminal Investigation in Stuttgart, for example, collected data from dozens of people from Afghanistan and Romania and uploaded them in 2017 to the YHRD public database without indicating whether individuals had consented to their data being used in this way¹³.

Roma people are over-represented in the databases accessed by law-enforcement agencies – both because of biases in criminal-justice systems and because geneticists have sought data from communities thought to be genetically isolated. In the ‘national database’ of Bulgaria held in the YHRD, for instance, 52.7%

Comment

of data sets are categorized as ‘Romani’, 36.9% as ‘Bulgarian’ and 10.3% as ‘Turks’, even though Roma people make up only 4.9% of the Bulgarian population (see ‘A biased picture?’).

Some forensic geneticists argue that this over-representation might actually benefit members of minority populations. According to them, reducing the rarity of any one person’s DNA profile in a database increases the chances of that person being exonerated in court. But such a claim cannot be made without assessment of the relevant technology – the use of methodological, sociological, ethical, philosophical and legal analyses to evaluate the impacts of implementing a technology in society. And this evaluation would have to include all possible uses of the databases, such as genetic-ancestry testing or the de-anonymization of families, in which relatives’ identities can be revealed through cross-referencing using other available data.

For several reasons, many people from genetically isolated communities are vulnerable when it comes to de-anonymization – especially those who have rare genetic diseases¹⁴. Users of the YHRD or the AFND cannot easily search for individual data. But the YHRD, for example, displays allele frequencies in specific geographical locations (sometimes down to village names) and further cultural information is provided in the referenced publications. Indeed, one can make inferences about certain communities and families even if genetic markers are used rather than full DNA sequences.

It seems unlikely that Roma people, along with many vulnerable groups, will benefit from their DNA being collected¹⁵. The development of medications for rare diseases on the basis of data from genetically isolated communities could, in principle, benefit members of those communities^{16,17}. Yet, in the case of Roma people, we have not been able to find an example of research that has been conducted in a truly cooperative way – such as involving members of the community or efforts to improve the community’s access to health services, including therapies that might already be available.

The problems we have identified with respect to Roma people are highly likely to apply to other groups. From looking at publications and following the data, we know that large genetics projects or databases such as the Human Genome Diversity Project and the YHRD, and the Kidd Lab private database, run by geneticist Kenneth Kidd at Yale University in New Haven, Connecticut, include data and samples taken decades ago from Indigenous peoples and populations considered genetically isolated, such as the San people and the Karitiana of western Brazil¹⁸ (see ‘What the database keepers say’). These data and materials have been used and shared by researchers around the world for more than 20 years. A broad verbal consent for research uses, taken and considered acceptable 30 years ago,

cannot cover all reuses of data and samples that are technologically feasible today.

Course correction

Some researchers and journal editors are trying to make changes, owing in part to increased awareness worldwide of the injustices experienced by minority populations. In the past year, two journals – the *International Journal of Legal Medicine* and *Human Genetics* (both published by Springer Nature, the publisher of *Nature*) – have retracted six papers that use DNA from Chinese minority ethnic groups. We know of another journal that is currently investigating a study that uses DNA from Roma people.

These are welcome steps. But much more must be done. In our view, resolving these problems requires four actions.

Establish an international oversight board.

Human and forensic geneticists, bioethicists, medical scientists, anthropologists and scholars from the social sciences and humanities – as well as community advocates – need to investigate all the DNA data held in public databases that has been obtained from oppressed groups.

WHAT THE DATABASE KEEPERS SAY

Four out of six database coordinators responded to requests for comment.

Geneticist Kenneth Kidd at Yale University in New Haven, Connecticut, who runs the Allele Frequency Database (ALFRED), agrees that “in the past, inadequate consent was a problem”.

Jean-François Deleuze, scientific director of the Human Polymorphism Study Center (CEPH) in Paris, which holds the cell lines and data for the Human Genome Diversity Project, notes that since Europe’s General Data Protection Regulation came into force, “CEPH only distributes global allelic frequencies of genetic markers”, making “the re-identification of samples now impossible”.

Andrew Jones at the University of Liverpool, UK, who co-runs the Allele Frequency Net Database, accepts that the organization has “a responsibility to ‘curate’ data or metadata where there are problems”.

Kristiina Tambets, head of the Estonian Biocentre Human Genome Diversity Panel, states that those working for the database always make sure that they “have the ethical permits in place when [they] deal with human subjects”.

No responses were received from the European DNA Profiling Group’s Mitochondrial DNA Population Database or the Y-STR Haplotype Reference Database.

As a first step, an oversight board could create a list of ‘at-risk’ populations for which problems with DNA data have been identified. Researchers, editors, members of the communities, forensic investigators and so on could then check to see whether the population they are working with or concerned about is on the list.

The European Society of Human Genetics could lead this effort, joined by societies from around the world. Such a board could establish how DNA has been collected, analysed and interpreted (much as we have done for Roma people, but more systematically); the nature of the consent given (if at all); and any resulting harms or benefits affecting the groups from which the data have been collected. In other words, it would extend the ethical diligence that is better established in medical genetics to research on all human genetic data.

The International Society for Forensic Genetics is already setting up an oversight board to examine cases in which consent is unclear. This is promising. But what we are calling for would be broader. Because different ethical standards between different research communities is part of the problem, forensic geneticists cannot solve the problem alone; they need guidance from other disciplines and stakeholders. Such analyses must be co-produced with members of the communities affected, as well as with scholars who understand the political and societal contexts facing these populations.

Retract unethical work and improve publishing practices.

More pressure must be put on journal editors and publishers to investigate and, if necessary, retract problematic studies. In principle, researchers could flag ethically troublesome research to the oversight board, which could then take up the issue with the journal. For new submissions involving DNA data from at-risk populations, reviewer panels must include bioethicists or other experts who know the communities involved and the societal challenges they face. If a reviewer has concerns, the communities must be consulted. It should also be mandatory for researchers to publish blank versions of the informed-consent forms (or equivalent) used for DNA data collection. Institutions, funders and researchers can put further pressure on journal editors and publishers by refusing to support, peer review or reward studies that fail to meet agreed standards.

Encouragingly, a statement this year by the Committee on Publication Ethics, a non-profit advisory organization, emphasizes the need for editors and publishers to give special protection to “vulnerable populations”. The CARE Principles for Indigenous Data Governance also offer some guidance on this, but editors and reviewers need to apply these principles to all at-risk populations, not just to those described as Indigenous.

Numerous non-governmental groups,

lawyers and scholars now advocate for many Indigenous groups regarding their DNA rights, particularly in the United States, Australia and Canada. This is not the case for Roma people and other migrant, stateless, nomadic or displaced populations around the world, including Tibetan people in China, Kurdish communities in Turkey or Ethiopian Jewish individuals (all of whom are represented in DNA databases). People in these groups are perceived by many to be foreign in their home countries.

Improve scientific training. In our analysis, we were surprised by the patchy awareness among researchers and institutions of the ethical problems of collecting genetic data from marginalized communities. Some were quick to realize the issues. Others were less willing to engage. In one e-mail, a journal editor joked to the employee of a publisher that they would need to “organise a time traveling machine and go back in time and make these better”.

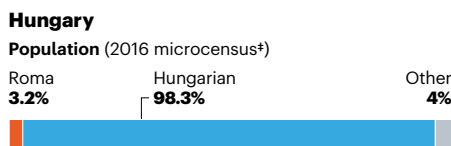
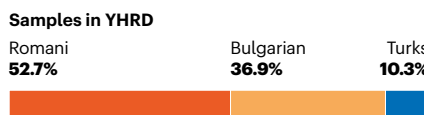
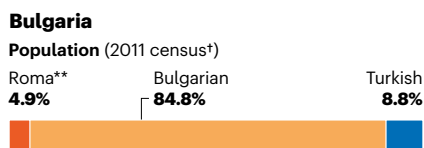
Undergraduates and postgraduates studying human genetics should be taught about potential harms to participants of genetics studies, and how to avoid such damage. PhD students should be required to take courses, ideally involving members of oppressed communities. And workshops to bring senior researchers up to date with current best practice should be mandatory. Several scholars have demonstrated how this training could be achieved, including anthropologists Kim TallBear at the University of Alberta in Edmonton, Canada, and Emma Kowal at Deakin University in Melbourne, Australia, as well as geneticists Deborah Bolnick at the University of Connecticut in Storrs and Keolu Fox at the University of California, San Diego.

Encourage community participation. Individuals whose DNA might be studied must be involved in research projects from the outset. At the very least, this means researchers engaging in a two-way dialogue with people about the benefits and returns they can personally expect (or not), and about the risks of DNA donation. It also means providing community members with ways to stay informed about the uses of their data (perhaps through a smartphone app), or to withdraw their DNA from a project at any time. The international board we are proposing could help to oversee this. Even better would be to train community members in genomics so that people in marginalized communities can identify research questions that are relevant to them.

Again, examples of such cooperative approaches already exist. Kowal established the world’s first Indigenous-governed genome facility – Australia’s National Centre for Indigenous Genomics, hosted at the Australian National University in Canberra. There, members of Indigenous communities decide what

A BIASED PICTURE?

Roma people are over-represented in databases such as the YHRD*, which is accessed by law-enforcement agencies. Without thorough evaluation, the effects of this over-representation are unclear.



*YHRD, Y-STR Haplotype Reference Database. **All ethnic-group labels are those used by the YHRD and population censuses. YHRD data as of 11 November 2021. †Total is <100%: some respondents did not identify with an ethnic group. *Total is >100%: respondents could select more than one ethnic category.

research questions should be asked and how data should be handled.

Sensitive approach

Over the past decade or so, several scientists have urged researchers to collect more DNA data from minority populations, warning that genomics medicine could benefit only a privileged few if this doesn’t happen^{19,20}. We commend these calls. Yet minority populations will be harmed in other ways if DNA collections and analyses are not methodologically sound, or are conducted without awareness of and sensitivity to the societal challenges people face.

Geneticists in Europe need to face up to the fact that unethical research practices are still happening on home soil – not just on other continents. Indeed, political actors have been using genetic studies on Roma people to bolster discriminatory policies. For example, in 2015, the European Commission launched infringement proceedings against the Slovakian government for its policy, established in the 1970s and reinforced after 1990, of segregating Roma children in schools for those with “mild mental disabilities”. In its response, the Slovakian government cited “genetically determined disorders” associated with “inbreeding”.

Such policies are concerning for two reasons. Policies for many Roma children might be being shaped by the health conditions of a few. Also, any child with additional needs requires more educational and emotional support, not less. Slovakia’s schools for those with “mild mental disabilities” are notorious for providing a poor standard of education²¹. Only last year, after a change of government,

did Slovakia acknowledge that this segregation is a problem and begin an investigation.

Meanwhile, more human geneticists globally must take on truly collaborative work across disciplinary and societal boundaries. This would ensure that communities or families whose members experience disproportionate rates of rare genetic diseases are treated with care and respect – not just as a ‘unique research tool’ or ‘precious resource’, as some geneticists write in their publications. Given our long history of misrepresenting human genetic variation, these challenges must be met if people’s trust in science, as well as in health care, policing and criminal justice, is to be retained – or, in some cases, restored.

The authors

Veronika Lipphardt is professor of science and technology studies at the University of Freiburg, Germany. **Mihai Surdu** is a visiting researcher at the University of Freiburg, Germany. **Nils Ellebrecht** is a research scientist in the Department of Sociology, University of Freiburg, Germany. **Peter Pfaffelhuber** is professor of probability theory and modelling in the life sciences at the University of Freiburg, Germany. **Matthias Wienroth** is senior fellow at the Centre for Crime and Policing, Northumbria University, Newcastle upon Tyne, UK. **Gudrun A. Rappold** is professor of human genetics and director of the Department of Molecular Human Genetics at the University of Heidelberg, Germany. e-mail: veronika.lipphardt@ucf.uni-freiburg.de

- Claw, K. G. et al. *Nature Commun.* **9**, 2957 (2018).
- Yildiz, C. & De Genova, N. *Soc. Identities* **24**, 425–441 (2018).
- Gresham, D. et al. *Am. J. Hum. Genet.* **69**, 1314–1331 (2001).
- Mendizabal, I. et al. *Curr. Biol.* **22**, 2342–2349 (2012).
- Martinez-Cruz, B. et al. *Eur. J. Hum. Genet.* **24**, 937–943 (2016).
- Surdu, M. *Those Who Count: Expert Practices of Roma Classification* (Central European Univ. Press, 2016).
- Müller-Hill, B. *Murderous Science: Elimination by Scientific Selection of Jews, Gypsies, and Others, Germany, 1933–1945* (Oxford Univ. Press, 1988).
- van Baar, H. *Social Identities* **24**, 442–458 (2018).
- European Union Agency for Fundamental Rights. *Second European Union Minorities and Discrimination Survey. Roma – Selected Findings* (European Union, 2016).
- Lipphardt, V., Rappold, G. & Surdu, M. Preprint at ResearchGate <https://doi.org/10.13140/RG.2.2.13286.04165> (2020).
- Carracedo, A. et al. *Forensic Sci. Int. Genet.* **4**, 145–147 (2010).
- Parson, W. & Roewer, L. *Int. J. Legal Med.* **124**, 505–509 (2010).
- Lipphardt, V. & Surdu, M. Preprint at ResearchGate <https://doi.org/10.13140/RG.2.2.16641.48484> (2021).
- Mascalzoni, D. et al. *Eur. J. Hum. Genet.* **18**, 296–302 (2010).
- Kalaydjieva, L., Gresham, D. & Calafell, F. *BMC Med. Genet.* **2**, 5 (2001).
- Reardon, J. *The Postgenomic Condition: Ethics, Justice & Knowledge After the Genome* (Univ. Chicago Press, 2017).
- Fox, K. N. *Engl. J. Med.* **383**, 411–413 (2020).
- Munsterhjelm, M. *Int. J. Semiot. Law* **28**, 289–316 (2014).
- Peterson, R. E. et al. *Cell* **179**, 589–603 (2019).
- Popejoy, A. & Fullerton, S. *Nature* **538**, 161–164 (2016).
- Amnesty International and European Roma Rights Centre. *A Lesson in Discrimination* (Amnesty & ERRC, 2017).

The authors declare no competing interests.