Science in Context (2021), **34**, 69–100
doi:10.1017/S0269889722000023

RESEARCH ARTICLE

Representing vulnerable populations in genetic studies: The case of the Roma

Veronika Lipphardt[1]*, Gudrun A. Rappold[2] and Mihai Surdu[1]

[1]Albert Ludwig University Freiburg and [2]University of Heidelberg
*Corresponding author. Email: veronika.lipphardt@ucf.uni-freiburg.de

Argument
Moreau (2019) has raised concerns about the use of DNA data obtained from vulnerable populations, such as the Uighurs in China. We discuss another case, situated in Europe and with a research history dating back 100 years: genetic investigations of Roma. In our article, we focus on problems surrounding representativity in these studies. We claim that many of the circa 440 publications in our sample neglect the methodological and conceptual challenges of representativity. Moreover, authors do not account for problematic misrepresentations of Roma resulting from the conceptual frameworks and sampling schemes they use. We question the representation of Roma as a “genetic isolate” and the underlying rationales, with a strong focus on sampling strategies. We discuss our results against the optimistic prognosis that the “new genetics” could help to overcome essentialist understandings of groups.

Keywords: Vulnerable populations; Roma; sampling; representativity; genetic isolation

Introduction

“All European Roma,” states an article published in 2015 in the *European Journal of Human Genetics*, “appear to descend from a low number of founders, and to have diverged into socially distinct endogamous groups after their arrival in Europe” (Martínez-Cruz et al. 2015, 2). This grand claim may provoke questions in readers not trained in genetics. How can someone make such a general claim about “all European Roma”? And why is this knowledge framed in such puzzling terms? Or, to start off more broadly, what, after all, is known about “the Roma”? How are the boundaries of Europe defined? What would be necessary in order to provide reliable knowledge about the Roma? Who is a Roma, or what criteria would allow us to distinguish Roma from others?[1] And why, or how, can genetics contribute to all of this?

[1] The grouping of Roma in one homogenous category has a long history, and it is beyond the scope of this article to retrace it. The term “Roma“ was introduced in political and academic discourse to replace the term “Gypsies” after 1971, when the first World Romani Congress was held. “Gypsies” is considered a pejorative term by many who identify as Roma, though others self-identify as “Gypsies” in censuses, interviews or other situations. Even though ethnic self-identification is considered state-of-the-art in census taking, some state administrations still employ the term “Roma” for classifying people who would not self-identify as such. For example, some groups who are counted under the category “Roma” would self-identify as Egyptians and Ashkali (in Macedonia, Kosovo and Albania), Boyash or Rudars (in Romania, Serbia, Croatia and Hungary). The Council of Europe’s (CoE) understanding of the term “Roma” is widely used in policy reports and academic publications: ”The term ‘Roma’ used at the Council of Europe refers to Roma, Sinti, Kale and related groups in Europe, including Travellers and the Eastern groups (Dom and Lom), and covers the wide diversity of the groups concerned, including persons who identify themselves as Gypsies” (CoE 2012, 4). The CoE comes close to an essentialist definition of Roma, linking ethnic belonging with ancestral origin: “The term ‘Roma’, as used internationally, denotes all groups sharing a common Indian origin (Roma, Sinti, Kale), and the communities who refer to themselves as Roma, found mainly in the Balkans and central and eastern Europe, but also throughout the world” (2012, 7). Although we consider it problematic to use the term “Roma” to

Far from being able to answer these questions, we instead wish to examine how some geneticists have answered them in DNA studies in the last thirty years. These answers include frequent statements about how genetics can contribute to one's understanding of the Roma, how little had been known about the history of Roma before these studies, and how much knowledge there is still to gain through future genetic investigations of Roma. For example, a press release announced a new genetic study in 2012: "The Romani people," it said, "lack written historical records on their origins and dispersal" (Cell Press 2012). To "fill in the gaps," geneticists had "gathered genome-wide data from 13 Romani groups collected across Europe to confirm an Indian origin for European Romani, consistent with earlier linguistic studies" (ibid). The release quotes co-author David Comas' claim that: "Their marginalized situation in many countries also seems to have affected their visibility in scientific studies." One co-author of this study is quoted with an evaluation of the wider usefulness of genetic data from Roma, arguing that: "Our study clearly illustrates that understanding the Romani's [sic] genetic legacy is necessary to complete the genetic characterization of Europeans as a whole, with implications for various fields, from human evolution to the health sciences" (ibid). An unknown "genetic legacy"; a potential "Indian origin"; a history of marginalization and invisibility; gaps in historical records; the Europeans' desire for "genetic characterization"; human evolution and human health: all those appeals to knowledge are woven together in a single statement that invokes the unknown and emphasizes the potential.

Yet Roma have not been "invisible" to scientific study, and a continuous stream of publications focusing on the heredity and genetics of Roma have been published over the past hundred years, resulting in more than 440 publications between 1921 and 2019. About 75% of these (ca. 340) were published in the past thirty years - approximately 220 of them in the field of medical genetics, 75 in population genetics and 45 in forensic genetics.[2] Those studies analyze, compare, discuss or otherwise draw conclusions on DNA data obtained from individuals labeled "Roma." In forensic genetics journals, Roma have been the most intensely studied population in Europe over the past thirty years.[3] It seems the geneticists who pursue those studies have extensive data and knowledge in their field upon which to build.

Approach

In this paper, we examine how exactly those authors make their claims of what is known and what is not known about the Roma, and we examine how Roma groupness and ethnicity is concluded upon in those genetic studies. We understand our approach as firmly rooted in Science and Technology Studies (STS). In this field, numerous publications have tackled issues of genetics and society, and our work draws on and contributes to this strand of research. With this paper, we wish to address a broad audience – geneticists as much as social scientists as well as humanities scholars – some of whom would perhaps not be ready to follow through the kind of theoretical debate usually introduced in an STS paper. Accordingly, we contextualize our findings within the relevant STS scholarship rather sparsely in this section and more towards the end of this paper. Our goal is to involve colleagues from all relevant disciplines in an interdisciplinary debate. Hence, both our approach and vocabulary must take a form that facilitates communication across disciplines.[4]

This, however, is challenging. Already the abstract of this text has probably triggered reactions in some readers. Some scholars from the social sciences and humanities, following their first

ascribe ancestral origins to people who do not self-idenfy accordingly, or as an umbrella category in censuses and expert estimates (see Surdu 2016), we use it in this article to refer to the persons addressed as Roma, "Gypsies" or "Roma/Gypsies" in genetic studies and those who are subsumed as "European Roma population" in these studies.

[2] Not included in this analysis are ca. 70 seroanthropological studies published between 1921 and 1994.

[3] For an overview of ethically problematic aspects of using DNA data of Roma in forensic contexts see Lipphardt, Rappold and Surdu's article (under review), "Ethical Standards in Forensic Genetic Research - a Critical Appraisal of Roma Studies".

[4] Further publications are in the making. In Freiburg, we are closely collaborating with colleagues from the life sciences (biologists, epidemiologists) and mathematicians.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

impulse, might consider any genetic research on minority groups such as Roma as ethically problematic. On a fundamental level, from their perspective, approaching and describing Roma with genetic concepts, terms and methods seems like a reprehensible echo of particularly dark historical moments.

As justified as such concerns might seem to those readers, others who have training in genetics might think differently. Geneticists involved in these studies would think that they are applying the same methodological tools here as they would for any other ethnic group. After all, many of these studies state that they have passed the appropriate ethical procedures or have been approved by relevant institutions, such as ethical committees. After much public debate about ethics in human genetics and human variation research over the past three decades, members of the genetics community would certainly say they seek the ethically most appropriate, or the least offensive and harmful approach they can think of. Within those boundaries, the object of curiosity is justified by its significance, or by its informative value. In the case of the Roma, according to many genetic studies, this value is believed to be especially high, as we shall demonstrate below.

One could approach these differences from the perspective of ethics of science, a field of growing importance, or from the perspective of science policy, highlighting international agreements for ethical standards in genetic research. For the sake of brevity, these perspectives are not taken here, but in another paper that discusses ethical aspects of genetic studies of Roma (Lipphardt, Rappold and Surdu 2022, under review).

Yet between questions about ethical approval procedures and worries about the repercussions of historical moments, there is another level of critical awareness for the potential shortcomings of genetic studies of vulnerable groups. Coming from the perspective of sociology and epistemology, scholars have warned against the essentializing and reifying effects of representing ethnic groups, and have asked pressing questions about how legitimate an object of scientific curiosity ethnic groups can be. Many social sciences and humanities scholars, as well as interdisciplinary author panels, have discussed the risks of geneticization, essentialism and genetic determinism in this context. Most of these critics have not simply rejected genetic studies of vulnerable populations, but oftentimes seek a more nuanced and differentiated approach and call for heightened ethical awareness.[5]

Rogers Brubaker, an UCLA-based sociologist, writes about population genetics and the "newly respectable biological objectivism about race" (Brubaker 2015, 54). He warns against simply "reasserting the usual mantra that there are no biologically significant differences between socially defined racial categories" (Brubaker 2015, 55). Rather, Brubaker explains, "it is not that socially defined racial categories are entirely arbitrary, bearing no relation to biogeographic and biogenetic ancestry. Since social understandings of race and ethnicity emphasize origins and descent, it would be surprising if socially defined ethnic and racial categories did not capture, in a crude way, some information about biogeographic and biogenetic ancestry" (Brubaker 2015, 83).

To be sure, Brubaker does not ask us to simply adopt the idea that social groups neatly overlap with racial categories. Rather, he claims that genomics can – at least crudely – infer self-identified ancestry or race from genotype. It cannot, however, make the opposite inference: "If there is information on the self-identified ancestry or race of an individual, one cannot infer their genotype. This is because genetic variation does not take the form of discrete and sharply bounded groups" (Brubaker 2015, 82-3).

[5] To name but a few who have demonstrated that cultural, political and social preassumptions about human groups not only inform the research designs, group labels and collection of DNA data, but also reinforce existing stereotypical group notions concerning ethnic or racial minorities, vulnerable and marginalized groups, see: Fujimura et al 2014; Bliss 2015, 2018; Duster, 2015; Fullwiley 2015; Gannett 2014; Lipphardt 2014, 2019; M'charek, Schramm, Skinner 2014; Radin and Kowal 2015; Munsterhjelm 2014; Rajagopalan, Nelson and Fujimura 2017; Reardon 2017; Schwartz-Marin et al. 2015; Tallbear 2013 and Wade et al. 2014.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Others have warned more strongly against neglecting the disconnects between ethnic or racial labels and genotypes, and pointed to the importance of considering the sampling as a critical moment in genetic research on human variation (Fujimura et al. 2014; Nash 2013). Brubaker, too, is well aware of the risks of the “new objectivism,” noting that: “By providing a natural foundation for social identities, geneticization can essentialize, even absolutize understandings of difference” (Brubaker 2015, 54). But Brubaker hopes that there is also something to gain from embracing the new objectivism – a kind of de-essentialization that could ultimately help to fight racism. “By highlighting the genetic heterogeneity within any collectivity, the dominance of within-group over between-group variation, and the histories – ancient and modern – of migration, gene flow, and admixture,” he argues, “geneticization can undermine understandings of pure, internally homogeneous, externally bounded groups” (Brubaker 2015, 54).

Genetic variation understood as non-discrete, not sharply bounded, not pure and not structured into homogeneous groups: without doubt, many geneticists would subscribe to this understanding of human genetic differences. Some population geneticists may be deeply engaged in a research agenda along these lines. This perception is strongly represented in Brubaker’s groundbreaking book *Ethnicity without groups* (2004), with its influential critique of essentialist and deterministic understandings of groups and ethnicities.

But seen from the perspective of the groups or ethnicities that have been studied as distinct genetic populations, genetics does not look unified in this regard. Non-essentialist understandings of genetic variation have too often not been the guiding principle for the geneticists. Some classical examples for genetically isolated populations have never been framed in other than essentialist ways. They have invariably been described as discrete, sharply bounded, more or less homogeneous groups. The Roma are but one example for which Brubaker’s hope is not justified: the undermining of essentialist understandings of groupness by genetic studies has not worked in their case.

A number of studies from the social sciences and from STS (science and technology studies) on special populations in genetics and genomics have appeared in the past decade. These have addressed, for example, populations in Brazil (Santos, Da Silva and Gibbon 2014), Mexico (Benjamin 2009), and other South American countries (Kent et al. 2015); Iceland (Pálsson 2008), Finland (Tupasela 2016), Quebec (Hinterberger 2012), Singapore (Ong 2016), and Taiwan (Tsai 2010). Reardon (2005) and M’charek (2005) have written about isolated populations in the Human Genome Diversity Project (HGDP). Munsterhjelm (2014) demonstrated how the Karitiana, a small indigenous group in Brazil, “became famous in forensic circles”: not because they were so overtly special, but because their genomes, accessed without their consent, were so readily available to forensic geneticists, and were such effective research tools due to their supposed isolation and inbreeding (Musterhjelm, 290).

Genetic studies on Roma, however, have hardly been the focus of social scientists, or at least not in comparable depth (Cazacu et al. 2013; Myers 2019). Nevertheless, these very few social sciences studies have added new insights: Geneticists have conceptualized Roma as different from other genetic isolates, as a very specific isolate indeed. Similar to Jews, they are seen as a transnational isolate or a diaspora group (Jobling 2014, 448). But in contrast to Jews, the authors of these studies believe, Roma have no “written records” of their own history, and in contrast to religious communities such as the Amish, they have no genealogies (Floersch, Longhofer and Latta 1997). In contrast to “Native Americans” who are viewed as indigenous groups (Tallbear 2013), Roma are depicted as a foreign population, while the comparison groups are seen as “indigenous” or “autochthonous populations”. Unlike the Karitiana, a group with only a few hundred individuals, the Roma in Europe officially count several millions, which allows for very different research designs. The Finns, another so-called genetic isolate in Finland, have attracted much of the geneticists’ interest since the 1980s (Tupasela 2016; Tarkkala and Tupasela 2018), but the Finns are not seen as a “transnational isolate”- for obvious reasons.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Other differences between the Roma and other so-called "genetic isolates" might strike the social scientist much more than the geneticist. The history of Roma being studied as a genetic isolate started a hundred years ago; and much more than the Finns, Roma are considered to be a vulnerable minority that remain heavily discriminated against even today. In contrast to the Saami, who established a Saami council in 1956, there is no Roma constituency that could prompt or preclude research, or successfully claim some of its economic benefits.

The geneticists studying Roma would say that genetics regards them as a genetically isolated population (see, for example, a separate chapter in a text book on human evolutionary genetics by Mark Jobling 2014, 448), and therefore they are justifiably viewed as a genetically bounded group. Yet this view misses an important point: Adopted as a conceptual premise, and then turned into a sampling strategy for genetic studies, the rationale of the "isolated population" becomes a circular logic, a self-fulfilling prophecy. To highlight this tautology is the main aim of this paper.

An advanced social sciences approach would firstly check the geneticists' groupness claims against state-of-the-art academic literature about Roma, and secondly ask for the representation of Roma in the genetic studies. An advanced life sciences approach would seek to reproduce research results with DNA data and then ask for a thorough inspection in terms of scientific standards. Regarding ethical questions, there would likely be a convergence of life science and social science critics of genetic essentialization.

In what follows, we leave aside most of these questions, and concentrate on one aspect we deem to be of interest for all sides: representativity. Whether a phenomenon is captured well in a scientific study depends much on adequate methodological considerations about how to represent it. If the main unit under investigation is "all European Roma," or "European Roma," or, as we may still find in studies even today, "Gypsies," then the obvious question is how to represent this main unit in a scientifically sound way.

In a social sciences methodology course, students learn what a main investigation unit (or population or universe) is, how it should be represented, and what methodological flaws one must avoid. In political science, students would learn how citizens of a nation state are to be represented. For some geneticists, however, this kind of social or political representativity (i.e., the question of how can a small number of people represent a large number of people such as a nation state's population) has not been a center of concern in the past.[6] Asking test subjects to self-assign to an ethnic category has become routine in biomedical research,[7] but this is not the same as representing populations and their history.

What is known and knowable about the Roma through genetic studies, then, depends on how Roma are recruited, sampled and represented in DNA databases. Whether the Roma (the European Roma, or the Roma in any given country such as Bulgaria, Romania, Hungary or Spain) are represented adequately in DNA databases is hence as much an issue for scholars from the social sciences and humanities as it is an issue for geneticists. No discipline alone can come to a conclusive judgement on this issue without consulting the other.

With a qualitative approach[8] to genetic studies of Roma published in the past three decades, we aim to point out the conceptual challenges of representativity. From several hundred, we have selected a handful of studies for a focused analysis, most of them from the field of population genetics with a main interest in the migration history of Roma. (Some medical genetic studies are included and indicated as such; we are aware that the challenges of representativity are not the same in population and medical genetics.) We have selected these studies for their recent

[6] More specifically, some have sought to overcome the limitations of small samples by technological solutions. But these solutions draw on the reconstruction of a supposed *biological* population; that is, a number of people sharing common biological ancestors, and not a politically or socially defined population.

[7] To be sure, this is standardized routine in English speaking countries such as the US, Canada, and Great Britain, but not - or much less so - in other countries.

[8] In subsequent publications, we will use quantitative methods for a statistical analysis of our text collection.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

publication date and their academic and public impact, and not because we think that they belong to the ethically and methodologically most problematic papers in our sample.

We do not discuss the selected genetic studies and their results in depth, but concentrate on sampling practices and representativity. Ethical questions, despite unavoidably popping up in Lipphardt et al. (2021). In spite of the problematic aspects we are going to point out, it is still important for us to state that in studies from the past ten years, we have noticed a trend towards more transparency regarding ethical procedures, self-assignment in recruitment, more cautious wording and more balanced methodologies. The papers we discuss embrace up-to-date critical awareness to varying degrees, but nevertheless reveal a lack of societal awareness that has consequences for research designs, methodologies and findings.

For simplicity, in what follows, sometimes the geneticists involved in the genetic studies on Roma are called "authors."

What do authors of genetic studies claim to know about the Roma?

Our examination of the authors' epistemic claims about Roma begins with numbers. How many Roma are there in Europe? Nobody knows, and there is no good way of knowing (Surdu 2016, 2019). States count their citizens, which leads to more or less accurate numbers, but there is no such count for all European Roma. Where Roma census numbers exist, state authorities (but also scholars, international organizations or Roma leaders) do not trust them to be correct: Roma, census takers claim, often do not self-identify as Roma because they fear discrimination (Surdu 2016).[9] That is why their total number is said to be uncertain, and estimates vary widely between four and twelve million. However, this lack of certainty about Roma population size must necessarily affect claims to representativity, as well as the reliability of figures for the prevalence of rare disease mutations in this group.

The study by Martínez-Cruz and colleagues, for example, admits that "social and political factors preclude the collection of precise census on the Roma," but then adds that "they are acknowledged as the largest ethnic minority of Europe, with a population of up to 10 million people spread across the continent and mostly concentrated in Central and South-Eastern Europe" (Martínez-Cruz et al. 2015, 1). The authors represent this group of humans within a genetic framework, which, they suggest, extends to up to ten million people, distributed over thousands of kilometers. All this, however, is based on estimates that are not produced by any academically recognized methodology, which is indeed a challenge in terms of representativity.

The central premise, the pre-assumption or starting point for most of the studies is that Roma in Europe are an *isolated* population. Some authors call them a "diaspora"; others, a "genetic isolate." In the past decade, some authors in population genetics studies have admitted a certain degree of "admixture" with the majority society, but the overall notion of a rather more than less genetically isolated population still holds implicitly and explicitly. For example, the press release mentioned above quotes a claim that "from a genome-wide perspective, Romani people share a common and unique history that consists of two elements: the roots in northwestern India and the admixture with non-Romani Europeans accumulating with different magnitudes during the out-of-India migration across Europe" (Cell Press, 2012).

This description speaks of a well-defined process that can be modeled and quantified. In the December 2012 volume of *Nature*, the same study is described in the section "Research highlights" under the title "Romani have Indian ancestry":

[9] An overwhelming number of scholars, representatives of Roma NGOs and international organizations, policy makers and politicians consider that a census based on self-assignment cannot produce a reliable count of Roma. Some argue that Roma "hide" their "true" identity and choose to self-identify with other ethnic labels. This, however, undermines the concept of self-identification as such and implicitly subscribes to an essentialized perspective of Roma ethnicity based on allegedly objective identification criteria.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

> The 11 million members of Europe's largest minority group, the Romani . . . , are descended from a single population that left India some 1,500 years ago and dispersed across Europe through the Balkans. [The research team] analysed the genomes of 152 Romani individuals from across Europe and compared them with those of populations worldwide. European Romani probably originated from northern and northwestern India. Genetic analysis suggests that, after leaving India, Romani ancestors interbred with local populations on the way to the Balkans before beginning to spread throughout Europe around 900 years ago. Since then, Romani have interbred with local populations in Europe.

While this text does not give any indication of the extent of admixture, and while a large extent would contradict the assumption that they are "descended from a single population," the study itself finds considerable evidence for admixture, recent and long ago, but nevertheless depicts it as rather limited. The conclusions state:

> Our data suggest that European Romani share a common genetic origin, which can be broadly ascribed to north/northwestern India around 1.5 kya. After a modest genetic contribution from the populations encountered through their rapid diaspora from India toward the European continent, our data indicate that the Romani dispersed from the Balkan area around 0.9 kya. We further observe evidence of secondary founding bottlenecks and small population sizes, together with isolation and strong endogamy. (Mendizabal et al. 2012, 2347)

In this description, the Roma are still a sharply bounded, discrete genetic group, isolated and strongly endogamous, yet not pure and homogeneous. The extent of admixture, however, is depicted as quite limited, temporally, geographically and dimensionally:

> Our data further imply that in more recent times, temporally and geographically variable admixture events with non-Romani Europeans have left a footprint in the Romani genomes. Overall, our analyses suggest that despite the relatively short time span, the demographic history of the Romani is rich and complex. Further studies with more dedicated geographical sampling and resequencing data would help in defining the Indian parental population of the Romani, as well as further details of their migration and subsequent history in Europe. (Mendizabal et al. 2012, 2347)

According to this account, the Roma have "Romani genomes" with a "footprint" of admixture in "recent times." What has priority for the authors, however, is to define the "Indian parental population."

Interdisciplinary input

Geneticists writing about Roma history usually rely on prior literature that assumes a particular grand narrative about their origin,[10] and they choose hypotheses to test that agree with that historical narrative.[11] If the knowledge geneticists obtain from DNA analysis were not in accordance

[10] Such grand "biohistorical narratives" can be understood as stories constitutive of social formations such as ethnic groups and nations and described in evolutionary biology language with concepts such as mutation, selection, drift, founder events and admixture (Lipphardt and Niewöhner 2007). These are intertwined with personal family stories of heritage and kinship.

[11] In genomics, so-called hypothesis-free methods are highly valued for achieving novel and unexpected insights. Analyzing the genomes of patients affected by the same disease symptoms for commonalities, one hopes for a significant finding or correlation. STS-scholars would argue that the sampling of a patient group (i.e. before the experiment is run) cannot be hypothesis-free – as the patients are hypothesized to represent a group affected by the same condition. For a genetic history study, the hypothesis is to be found in the sampling as well, but also in the assumptions about the "ancestral" population; the equivalent to the "experiment" is the population's history. The recounted narrative of that population history is also a hypothesis about how the observed genetic structure has emerged over time.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

with widespread societal and cultural narratives, then genetic studies on Roma might find it more difficult to find public resonance.

Accordingly, these articles sometimes build on unsubstantiated evidence (e.g. medieval chronicles or folk myths, as in Kalaydijeva et al. 2005, 1086), or often on academic knowledge from the humanities. Linguistic and anthropological studies are cited frequently and cursorily, as in the public release quoted above, and mostly as evidence for the Indian origin of Roma. Rarely are these references based on cutting edge research, but rather to articles and books published some decades ago.[12] In most cases, these accounts are used as a starting point or as a historical source for the Indian origin of Roma and for their migration routes in Europe.[13]

To build upon their central hypotheses of isolation, these geneticists require additional information on the lifestyle of Roma, which is generally drawn from cultural and ethnographic studies. These are often referenced only cursorily and without specific citations, as providing knowledge of their cultural characteristics.[14] Thus, for example, Melegh et al. (2017) notes that "Studies investigating Roma culture revealed significant similarities between Roma and Indian culture including the caste system and endogamic habits that means exclusive marriage within Roma sub-ethnic groups (clans)" (1). Salihovic et al. (2011) claim that "Traditional social organization based on strict and complex rules of endogamy and particular population history of the Roma have similarly shaped Romani population structure as well as epidemiology and molecular architecture of single-gene disorders" (263). Finally, Plášilová et al. (1999) argue that "The majority of Roma still preserve their language, traditions, and lifestyle, and their communities remain almost totally genetically isolated not only from the surrounding population but also from one another. Endogamy is a strict rule, consanguineous marriages are frequent (15-45%), and the inbreeding coefficient ranges among the highest worldwide" (293).

All three of these examples highlight a cultural tradition of endogamy. In this interpretation, any separation of Roma from the rest of society is self-inflicted, voluntary, and precisely what Roma culture dictates. The genetic isolation is hence depicted as a consequence of self-determined social separation, implying that, typically, Roma have offspring with Roma because they prefer to choose their marriage partners among themselves. Other factors that have also contributed to the societal isolation of Roma, such as discrimination, ghettoization, stigmatization, exclusion or persecution, are rarely taken into account in this narrative.

Some geneticists studying Roma could now say, well, there might be different reasons for isolation, but the reasons do not matter. Isolation is just isolation, and the result will always be an isolated population. However, factors like discrimination or ghettoization would not only lead to societal isolation followed by genetic isolation – they would contribute to genetic isolation *differently*. If voluntary endogamy is the causal factor for isolation, then the criteria of the community determine who is considered an acceptable marriage partner and who is not. If discrimination is the causal factor of isolation, the majority society determines the criteria for exclusion. These two causal factors are not the same. For example, in many Jewish families and communities, being Jewish requires one to be born to a Jewish mother, but this criterion is not shared by all majority societies of the Jewish diaspora. Also, some majority societies have excluded and ghettoized Roma

[12] Some genetics papers include references to articles published by the Gypsy Lore Society, mostly before 1945. Morar et al. (2004), Kalaydjieva et al. (2005), Gresham et al. (2001) and Tournev (2016) cite a publication from 1915-1916; Moorjani et al. (2013) cites a publication from 1927; de Pablo et al.(1992) and Ramal et al. (2001) cite a publication from 1923; Regueiro et al. (2011) cites a publication from 1941.Historically, the publications of the Gypsy Lore Society were a major source of "scientific racism" until the late 1970s (Acton 2015).

[13] Angus Fraser's book *The Gypsies* (1992) is the most often cited publication from the humanities; in most cases as supporting the claims about endogamy as a cultural tradition among Roma. However, though Fraser also suggests that "mixing" was very frequent, those of Fraser's statements that contradict the conceptualization of Roma as a genetic isolate are not cited in the genetic studies.

[14] A noteworthy exception are the co-author contributions by ethnographers Marushiakova and Popov to the genetic studies of Gresham et al. (2001) and Martínez-Cruz et al. (2015).

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

together with other groups of “undesired” people – but these groups would not necessarily have been among the acceptable marriage partners for Roma families and communities. In some countries or regions, Roma were not the only ones affected by exclusion, and excluded groups were relocated to separate settlements – ghettos – together. Also, a suppressed minority can typically not maintain its own “traditions” with regard to marriage and reproduction. Being enslaved, for example, strongly limits one’s reproductive freedom.[15] Such complexities, however, which vary considerably from country to country and from region to region, are not considered in the genetic studies.

Representativity

What does representativity mean, and why does it matter for genetic studies of populations? For each of the subfields – forensic, medical and population history genetics – representing a group comes with different challenges, particularly with regard to the application contexts. For example, a medical geneticist needs to know how prevalent a mutation is in a population. A forensic geneticist, who wants to build up a reference database for checking allele frequencies in order to estimate how frequent a profile of a suspect is in a given place, needs to know whether she has tapped into substructure, or whether there might be real-world populations unrepresented in the database. A population history geneticist wants to collect DNA samples from individuals who most likely represent a supposed historical group, that is, whose ancestors have only married within their own group since historical times. Yet ultimately, in any of these cases, the geneticists who want to make claims about “all European Roma” must consider questions of representativity. Otherwise, they risk making claims that only hold for a small subgroup, or for that matter, not at all.

How would one plan to represent “all European Roma,” or ten million Roma, in a DNA database? According to many social scientists, the first thing to do would perhaps be defining criteria to decide who qualifies for the sample and who does not, while aiming at an appropriate sample size. This would cause considerable discussions, as the task is complex and raises fundamental questions.

To provide an example from biomedical research, “all Germans” would become defined as “all German citizens,” hence, only people with a German passport would qualify. In fact, biomedical large scale studies such as the “National Cohort” do consider representativity issues, as they recruit through the registration offices in order to ensure that the “results of the investigation will be transferable to the overall population [of Germany]” (NaKo Gesundheitsstudie website). This may still prompt discussions about representativity, however, there is a clear framework with statistical data available for contextualization and comparison.

There is, however, no Roma nation state citizenship to turn to for that criterion. Language would not be a reliable identifier either. Romani is a language spoken by some, but not by all people considered (or self-assigning as) Roma. Other criteria are even more questionable. A homogenous “Romani culture” is impossible to nail down; Romani surnames do not work either. Being discriminated against as a Roma or “Gypsy” by others is also no solid criterion. In addition, these criteria do not overlap.

Cutting edge historical and sociological evidence demonstrates that Roma have no common language, territory, religion, cultural practices or social status. Some scholars call them a “superdiverse” group (Tremlett 2014). Scholars from the social sciences and humanities today largely agree on the fluidity, complexity and situatedness of Roma identity (e.g. Bogdal 2011; Jonuz 2009; Kovats 2013; Law and Kovats 2018; Plájás, M’charek and van Baar 2019; Stewart 2013; Surdu 2016; Surdu and Kovats 2015; Veermersch 2005). The self-assignment as “Roma” is not a good proxy for external assignment, and vice versa. “Gypsy” cannot be viewed as synonymous

[15] In the territories of present-day Romania, the country with the largest Roma population, Roma were enslaved from the thirteenth until the mid-nineteenth century.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

with "Roma." Social historians consider the term "Gypsy" a construction imposed with differing rationales by national governments and administrations, through a long history of labeling, stigmatizing and repressive control, up until the genocide under the National Socialist (NS) regime and beyond (Lucassen 1991, 1997; Lucassen, Willems and Cottaar 1998; Mayall 2004; Willems 1997). In a number of case studies, scholars have shown that marrying partners from outside of the group is relatively common (e.g. Okely 1983; Fraser 1992; Achim 1998; Stewart 1997). Framing Roma as one generic group is therefore seen as a form of racialization, or essentialization (e.g. Law and Kovats 2018; Surdu 2016; Yıldız and De Genova 2017).

On the one hand, in many countries or regions, Roma have experienced long and repeated phases of integration – leading to what the geneticists would call "mixing." On the other hand, in different places, Roma were (and are) segregated, ghettoized and forced into societally and geographically marginal places by decision makers and authorities (About 2012; Donert 2008; Filhol 2013; Berescu 2019; Kóczé 2018; Picker 2017; van Baar 2015, 2018; Vincze 2019).

The vast existing scholarship on past and present integration and exclusion of Roma in different countries suggests that it is very difficult to sample or represent the Roma as a group. In census taking and in the social sciences, self-assignment is viewed as the most advisable method for the purpose of data collection and research about identity building. Hence, many social scientists would only admit people to the sample who self-identify as Roma. The same holds true for much of biomedical research today, since the US has introduced census categories (based on self-identification of race, including multiple racial belongings) for test subject recruitment (Epstein 2007).

However, self-identifying as Roma rarely comes with benefits in societal contexts where Roma are discriminated against (Jonuz 2009). This is also the reason why census takers distrust the data they have collected on Roma. Roma are believed to "hide" their "true" identity – that is, the identity census takers would have ascribed to them. Yet, individuals may identify with various population labels due to being culturally well integrated into majorities or surrounding populations; self-assignment may situationally emphasize one or the other category of belonging. For geneticists, this makes the sampling criteria of self-assignment problematic.[16]

External identification, that is, the identification of Roma by others, such as doctors, nurses, social workers, teachers, police officers, community leaders, neighbors etc., would be seen as problematic by most social scientists on both methodological and ethical grounds. In their influential empirical research, which drew upon a large set of data, Ladányi and Szelényi (2001) demonstrated that self-identification as Roma and external identification do not overlap, or are not equivalent to each other. Instead, understandings of Roma ethnicity vary greatly across cultural contexts. There is also considerable variation depending on the classificatory work invested by experts and fieldworkers in survey practices. For example, two sets of fieldwork are likely to produce incongruent classifications of Roma (Ladányi and Szelényi 2001).

Representativity also has a strong technical dimension (Fujimura and Rajagopalan 2011). If one is to use some of the standard software on two populations in order to examine their genetic relations, one needs to make sure the two populations have been sampled in a similar way, that their sizes are of the same order of magnitude, and that the two populations were sampled under the same conceptual framework (in our case, that of a genetic isolate). None of these conditions seems to be fulfilled in genetic studies of Roma. For studies from Hungary, for example, reference data of Hungarians are drawn from a national database, not from some isolated rural settlements.[17]

[16] In the past, geneticists have found various ways to overcome these problems, such as offering incentives. If there were effective incentives to self-identify as Roma, the geneticists would need to consider what this implies in terms of building a sample, as these incentives would perhaps attract people they did not expect to show up.

[17] We thank Peter Pfaffelhuber, Department of Mathematical Stochastics, Freiburg University, for his insightful comment on the non-comparability of the samples.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Beyond these methodological issues, using external appearance as a criterion for recognizing Roma – a criterion used sometimes by geneticists – would be considered much more problematic and even racist by social scientists.

How do the geneticists handle this complicated issue? After all, their investigations and the validity of their results completely depend on the samples they choose.

Sampling procedures

The population genetic studies on Roma from the past three decades are strikingly tight-lipped about their sampling criteria and practices; sampling schemes are not made explicit. Issues of representativity are rarely mentioned, and if so, not in an informative way. "The donors were real representatives of the entire population, as they were collected in a nationwide project," one study of allele frequencies says (Magyari et al. 2014, 149). Mendizabal et al (2012) state: "Alternatively, mixed couples may leave the Romani communities and integrate into the non-Romani societies, and thus would not be sampled from Romani groups in these countries." This suggests that "mixed" and "unmixed" couples segregate neatly. In doing so, it marginalizes cases of "unmixed" couples leaving the community and "mixed" couples staying within the community.

The silence on sampling is a relatively recent phenomenon. In their seroanthropological publications from the 1920s to the 1980s, geneticists were much more explicit regarding their sampling procedures. Many of their efforts were aimed at avoiding "mixed" individuals because they were interested in "pure Gypsies," who were, as they admit, hard to find and recruit. The idealized test subject was *the nomad*, even though nomadism was a marginal phenomenon. Yet nomads were seen as the most isolated from the society and therefore optimal for genetic studies, however rarely willing to cooperate. Towns of all sizes, where people tend to "mix," were avoided. Potential individuals were excluded if, upon being interviewed for recruitment, they said they were born to a mixed couple, or if the recruiters' expectation regarding Roma life style, culture or outlook were otherwise not met. Two studies (Clarke 1973; Rex-Kiss et al. 1972) explicitly state that one crucial criterion of selection was the visual inspection of the "external somatic features" of the recruited subjects (Rex-Kiss et al. 1972, 358). Rex-Kiss et al.'s sampling strategy led them to prisons; other researchers turned to other institutions that, for one purpose or another, classified and treated Roma separately from other citizens.

We firmly assume that sampling would be done differently today. How exactly it would be done though, remains unclear. If sampling information is given, it is vague and abstract. Many studies rely on data shared by other teams. When it comes to describing the sampling scheme, the authors point to the team that has collected the data.[18] Following up on the latter's publications, one can sometimes not find any information on sampling there either. Rather, the information given there is sometimes even more vague. In some cases, following those references back through the literature shows that the data has in fact never been published.

Yet sampling practices are not completely opaque. Approaching individuals for recruitment in population genetic studies, the DNA collectors still want to make sure they do not include people with too little or no Roma ancestry. From some studies, single hints can be gathered as to how this might have been ascertained, and we have checked these observations with two expert interviews. It seems that, in some cases, questionnaires may ask for an individual's self-assignment, for the ethnicity or self-assignment of their four grandparents, for lifestyle parameters, for their mother tongue, for cultural traditions, certificates or registries. It also seems not unusual for scientists to rely on external assignment by a third person or institution, such as a doctor, a community senior, a state official. Physical appearance is yet another selection criterion that is still used today, even if, perhaps, not systematically and not explicitly.

[18] The dataset from the study of Gresham et al. (2001) has been shared with other research teams at least 20 times. In some cases, we observed unexplained attrition of data.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Self-assignment is mentioned in some studies as a sampling criterion. For example, in one study, samples are described to derive from “27 self-declared Romani” (Gomez-Carballa et al. 2013, 2). But on the other hand, it is seen as rather unreliable information. Mendizabal et al (2012) state in the supplement: “All individuals included in this study were self-identified as Romani [sic]. Importantly, the self-identification as Romani is a delicate matter in some European countries due to the social stigma attached to Romani identity; hence additional information obtained in sampling can be scant” (Mendizabal 2012, Supplement, 1). How the teams overcome this problem, what other criteria they use instead, remains unclear.

Sampling issues: Family relationships, small samples, privacy and social pressure

Population genetic studies need to account for the risk of tapping into population substructure, particularly when studying isolated populations (Ehler and Vanek 2017). If samples are taken from the same family or neighborhood, the risk of sampling bias is considerable. This is relevant for considerations of representativity in our case, especially in the context of small samples that were collected in small, ghettoized Roma settlements. In some cases, what is taken to be a representative sample of Roma in a specific country, or even of “European Roma,” may in the worst case be based on a limited number of related community members, in a limited number of locations that have become exclusive sampling sites for genetic studies of Roma over the last few decades. Several locations in some East European countries have been long term sampling sites for genetic studies (e.g. Baranya county in Hungary; Kosice in Slovakia). Even more problematic, some of these samples have been used and shared for decades.

Some of the studies give information on how and where sampling took place, such as the names of villages or city quarters. Martínez-Cruz et al. (2015), for example, recruited 110 subjects from seven neighboring villages in Greece, which seems problematic if family relations need to be avoided. (Whether publishing this information, including the villages’ names, complies with privacy and anonymity obligations is yet another open question.)

In other cases, samples were collected in clinics, doctors’ offices or medical care institutions, or in health care schemes addressing Roma or people with a specific health problem. While healthcare systems seem to ease access to individuals who have previously been labeled “Roma” by that very system, it is in many cases unclear whether those people would self-identify as Roma, or under what circumstances they would (not). Several forensic genetic papers using DNA data from Roma list co-authors affiliated with police, investigative or military forces (for details see Lipphardt, Rappold and Surdu 2022, under review). Three forensic studies explicitly mention that their samples were collected by medical doctors (Nagy et al. 2007, 25; Saiz et al. 2014, re-using the data of Novokmet and Pavčec 2007). The ethnic categorization of samples by the collectors indicates that systematic ethnic labelling for Roma is in place in medical institutions in some countries.

When such data collected in healthcare settings is used to address population genetics questions, this can have problematic consequences for representativity. After all, specific healthcare programs attract communities and families nearby, as well as, plausibly, people with similar genetic dispositions to particular non-genetic conditions. In the case of healthcare schemes addressing genetic diseases, it is plausible to assume that relatives will show up at the same medical institution to get healthcare. Third degree cousins, for example, aware or not aware of their kinship, are genetically more similar to each other than unrelated individuals. Hence the necessity to account for the risk of sampling bias, especially since some geneticists describe Roma communities as “inbreeding,” “consanguineous,” “endogamous,” or as large, complex family networks. Test subjects are obviously often asked for family information, so that known relationships could be detected already in the doctor-patient conversation. But doing this without infringing privacy is a challenge if family members do not come to the care facility together. To put it differently, with

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

"inbreeding" and "isolation," the exact background that makes such families interesting for researchers is also the thing that undermines the validity of sampling, not just from a biological point of view, but in an ethical sense as well.

Technical controls for kinship are mentioned in some studies. The 2013 study from Spain focused on "27 self-declared Romani within the framework of ESIGEM," stating that "all these individuals had suffered from meningococcal disease" (Gomez-Carballa et al. 2013, 2). The authors checked for family relationships using identity-by-state analysis and found only one pair of individuals (among the twenty-seven) matching the criterion of "closely related." Another pair showed statistically significant evidence for second degree relatedness. That is four out of twenty-seven – and more sensitive methods used by genetic epidemiologists today would probably find even more family relationships.

This could particularly be the case if very small sample sizes were used as representative for a national minority population. For example, Mendizabal et al. (2012) use a sample of eight Lithuanian individuals for their study of European Romani. The complete sampling information reads: "The Lithuanian Romani were sampled in the 'Kirtimai' tabor (Roma settlement) in Vilnius. They belong to Verchnij tabor group and are mostly Polish speakers" (Mendizabal et al. 2012, supplement, 1). Kirtimai is a neighborhood only a few kilometers south of Vilnius' old city center. It is the only compact Roma settlement in Lithuania, with a population of 354 to 500 people, depending on the source of the estimate (Poviliunas 2011). Eight individuals from one single city neighborhood, which is a small compact Roma settlement, are probably easy to re-identify. It seems unlikely that there are no family relationships between them. Mendizabal et al (2012) state that they used "Tukey's outliers detection" to remove "individuals either showing a higher amount of inbreeding or larger than average identity-by-state distances in their sampling population" (Mendizabal et al. 2012, supplement, 1). This, however, is probably not sufficient to exclude kinship in this specific community.

A small sample size from a small, societally excluded community carries a high risk for so-called "cryptic relatedness."[19] This would make the inference of the history of a larger Roma population from that sample even more questionable, because the risk of capturing population substructure is larger in such a small local sample. The Kirtimai sample of eight is perhaps more representative for a locally specific population substructure than for Lithuanian Roma in general. Lithuania, after all, has a population of ca. 2500-3000 self-declared Roma, and they do not all live in Kirtimai.

Notably, it is unclear under what conditions these eight samples were collected: Mendizabal et al. (2012) do not give any reference. Gomez-Carballa et al. (2013) also use a Lithuanian sample and refer to Gresham et al. (2001). Gresham et al (2001) use twenty Lithuanian samples without giving a reference, implying that this is primary data. As Vaidutis Kucinskas from the Department of Human and Medical Genetics at the Faculty of Medicine of Vilnius University is co-author on all three studies, it seems plausible that he has contributed these samples, but without any information on the sampling, it is hard to tell what exactly these twenty or eight samples represent. Without any information on data attrition, it is also hard to tell why the sample has been reduced to eight. This is only one out of many examples. In order to learn more about such instances, one would need to contact co-authors in dozens of cases in which the relevant information is lacking.

[19] Considerable definitory lack of clarity exists for the three terms "endogamy," "inbreeding" and "incest," both between the fields of genetics and social sciences and within them. "Endogamy" and "incest" can be viewed as the two end points of a spectrum of in-group parenthood. In between the two, there is a vast range of parenthood between more or less closely related partners. In genetics, "inbreeding" is used interchangeably with the other two terms, and for covering phenomena all over that spectrum. Social scientists understand endogamy mainly as in-group marriage and parenthood between partners from unrelated families. "Inbreeding" would rather be understood as overlapping with "incest." Of course, "relatedness" and "incest" are culturally contingent concepts and differ between countries and societies.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Excluding "mixed" individuals: Removing data sets from the samples

While much of the sampling practices in the field seems to aim at focusing in on those individuals who represent the descendants of the "proto-Romani," (i.e. the group that departed India some 1,000 years ago), for some geneticists there still seems to be too much noise in the samples the recruiters bring to the lab. In particular, we found instances in which researchers attempted to avoid "mixed" individuals in their samples. Individual data sets might be excluded from a sample in the lab after the DNA analysis yielded a result that does not accord with what genetically was expected from a Roma.

For example, Melegh et al. (2017) use genome-wide SNP data from 179 "Roma samples." Twenty-seven of the samples had been documented in another study (Moorjani et al. 2013), which states that most individuals were "from Hungary"[20] (Moorjani et al. 2013, 8). Those twenty-seven participants had extensive interviews before giving written informed consent. About their self-assignment, the authors state: "Roma individuals self-reported as being descendants of the same tribe for at least three generations" (Moorjani et al. 2013, 8). These twenty-seven samples were then merged with the dataset of 152 samples from Mendizabal et al. (2012) discussed above – in which sampling criteria were not described in any detail – and the overall dataset was treated as one Roma dataset without any further subdivision.

The sampling rationale clearly favored isolated groups – "tribes" – and Indian origin. But in spite of their sampling strategy to include only "descendants of the same tribe for at least three generations," the authors state: "Our results showed that Roma have on average 81.08% +/- 0.53% West Eurasian related ancestry" (Melegh et al. 2017, 7). And yet, to arrive at this conclusion, the authors had to do much more than asking for tribal affiliations. As they explain: "Based on PCA and clustering methods, we removed Roma individuals from the merged Roma dataset, which showed significant admixture with non-Roma Europeans. The merged dataset contained 158 Roma samples featuring 599,472 autosomal SNPs" (Melegh et al. 2017, 2). This means that, even after excluding 11% of all participants on the basis of DNA results as "admixed" individuals (namely those who had too much European ancestry in the eyes of the authors), "West Eurasian" ancestry still dominated heavily.

Put differently, in order to demonstrate the Indian ancestry of Roma, the authors of some of these studies removed samples from those individuals that they deemed not "Indian enough." They were then left with, unsurprisingly, some Indian ancestry, but only as a minority subset within a bigger sample with a huge amount of admixed ancestry, mostly from Europe. And yet the minority subset is seen as the authentic, autochthonous part representing Roma while the larger sample with European ancestry is seen as the admixed interference. In other words, on top of the already restrictive sampling strategy, another layer of filtering is added to ensure that only a subset of individuals with some Indian ancestry would be retained for analysis.

Excluding "mixed" individuals is a concern that applies to many studies. In a series of publications, Hungarian authors documented concerns about the representativity of a large shared sample: Kosa et al. (2015) claim that their sample is not representative because, firstly, "assimilated" Roma had not been included, and secondly, some of the Hungarians sampled for the comparison group might have been Roma themselves (303). They conclude that this may have "slightly diluted the true difference between the populations" (ibid). No matter how confusing social realities proved to be, no matter how well integrated or "mixed" people in Hungary were, the authors remained concerned with the "true" difference, the most clear-cut difference, so to speak, which in their case meant the genetic difference.

[20] The text reads: "from Hungary (3 linguistically and culturally separated sub-groups: 7 samples from Olah (Vlah), 4 samples from Beas (Boyash) and 4 samples from Romungro), 4 samples from Romania, 4 samples from Spain and 4 samples from Slovakia."

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Similarly, Nagy et al. (2017), using Kosa's data, conclude that "the presence of participants with mixed Roma/non-Roma ancestry … may result in a slight underestimation of the differences between the populations" (Nagy et al. 2017, 455). Piko et al. (2017) and Fiatal et al. (2016) both use Kosa's data and state that "those Roma who have, to various degrees, assimilated with the Hungarian general population" have been excluded from the sample. Furthermore, both studies state that "because many people are reluctant to self-define their ethnicity as Roma, this constraint would be very difficult to overcome" (Piko et al. 2017, 124; Fiatal et al. 2016, 2265).

With the sharing of data or DNA samples between studies, the pattern of thinking is shared, too. If the "representative sample" of the general Hungarian population included some people who were "Roma" in the eyes of the authors, this had been revealed by the DNA analysis; no matter how those individuals would self-identify, genetically they had to be considered Roma. The Hungarians, then, were assumed to be sampled for their "unmixed" Hungarian ancestry; and if the sampling was successful, according to the authors, genetically speaking, there should be no "Roma" in that sample. As STS scholar Star (1983) has demonstrated, filtering data in the laboratory sometimes is (but should not be) part of the scientific work of transforming "ill structured" problems into "well structured" problems, by ignoring complexities and making choices in all stages of the research process, often under conditions of scarce resources and pressure to deliver significant results.

Terminology and wording

Representation also happens through language. In many studies, social, cultural and political separation comes to be reinterpreted in genetic terminology. Martínez-Cruz et al. (2015) hold that Roma are "an excellent model to evaluate the consequences of recent, multiple, and widespread dispersals and founder events," which sounds like a self-confident statement of solid knowledge (1). Similarly, a medical genetic study has stated that "the Gypsies are a young founder population comprising multiple genetically differentiated sub-isolates with strong founder effect and limited genetic diversity" (Kaneva et al. 2008, 191). Their population, according to these researchers, has "a substructure that can greatly facilitate the mapping and identification of disease genes" (Kaneva et al. 2003, 105). "Endogamy and inbreeding," another publication states, "lead to the accumulation of hereditary disorders" (Tournev 2016, 95). And in another study: "The proportion of slightly deleterious genetic variants accumulates during bottleneck events as the efficiency of purifying selection is diminished in small populations" (Mendizabal et al. 2013, 198). The mapping and identification of disease genes is the puzzle the scientists aim to tackle by employing their supposedly well-established *model* of an isolated population.

These quotes may sound like purely technical terminology – except for the population label "Gypsy" – but in fact, such statements are as much about society as they are about biology. Each term stands for a specific interpretation of societal situations Roma have experienced. Furthermore, there seems to be a misfit between the positive appraisal of usefulness on the one hand ("greatly facilitate the mapping and identification of disease genes"; "Our special 'research tool' will be the unique genetic heritage of Gypsies," Jordanova n.d.) and the negative connotations of inbreeding, deleterious genetic variants and selection on the other. Wordings such as these seem to speak for an instrumentalizing approach rather than one driven by empathy with people in miserable health conditions. While it is arguably not a priority for genetic publications to demonstrate empathy, there are role models in the field who manage to convey empathy in their scientific publications.

Some genetic studies speak of "Gypsy disorders," "Gypsy mutations," and even of "Gypsy chromosomes" (e.g. Morar et al. 2004). To call the population under investigation "Gypsy" or "Roma" is obviously seen as a scientifically irrelevant decision by many scientists. In conversations, we are told that this is just a question of sensitivity; many geneticists seem to strive for using the label least

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

offensive to their test subjects.[21] However, this cautionary approach might apply to consent forms and personal contact, but whether or not it also applies to scientific publications is not so clear. Would geneticists expect the individuals in question to read these publications? Would this be different for a societally well-established minority, as compared to a poor and discriminated one with high levels of illiteracy? In any case, many authors of these studies find it unproblematic to call the population "Gypsy," though many Roma would find this offensive. In conversations, we are sometimes told that it seems justified to ignore "political correctness" because even some Roma call themselves "Gypsy."

Viewed from the perspective of representativity, the following groups are not congruent and of very different size: Individuals who are willing to identify as "Gypsies" in private and public; individuals who are willing to identify as "Roma" in private and public; individuals who are called "Gypsies" by others; individuals who are called "Roma" by others. If genetic studies do not detail their sampling strategies in this regard, it is unclear what they represent. What is pretty clear, however, is the fact that these authors, when using the term "Gypsy", risk offending many of those they wish to represent.

The main unit: "All European Roma"; or rather "those with ancestors from India"?

Before we continue with reporting on the genetic studies' sampling practices and representativity, we include here an intermediating thought to make it easier to follow the rest of this paper. For social scientists who are not familiar with the relevant STS literature, the observation that these genetic studies seem to ignore obviously problematic aspects of representativity can be puzzling, even disturbing. In agreement with relevant STS literature (e.g. Fujimura and Rajagopalan 2011; Fujimura et al. 2014; Nash 2013; Gannett 2014; Bliss 2015, 2018), we offer a differentiated explanation, one on the level of conceptual differences. Geneticists carrying out these studies seem to have a different understanding of "population," namely, primarily a genetic one. Moreover, the genetic boundaries are the ones to define the boundaries of a group – that is, an individual is to be considered a Roma if the individual has biological ancestors from medieval Romani groups, and/or if genetic findings make the case.

To be sure, there might also be groups that are genetically quite closely related, but share no common idea of belonging; they may even have been enemies for centuries. But if genetic boundaries seem to match to some extent with some widely known social boundaries, population geneticists would view this as a successfully identified population structure. Social division perceived in society and biological difference studied in science seem to explain one another. That way, they reinforce the conceptual framework in which they both have been produced.

What the geneticists involved in these studies aim to explain are genetic differences between populations of today – or, more precisely, genetic differences between groups they consider as populations. Their epistemic object is not the social reality of Roma, but a genetically bounded population that, for them, seems to overlap strongly with the social group of Roma. The Roma seem to them one of the examples where genetic and social cohesion go hand in hand. The focus is on the genetic group, and any of the social markers used for recruitment are seen as powerful proxies for that group. That the social groupness of "Roma" could pose problems for their demarcation of the genetic population might not be a concern.

A significant boundary in genetic population structure can be made plausible by a historical explanation, a story through which readers can understand how the group came to be genetically

[21] A leading forensic journal stated in 2010: "It is therefore of utmost importance to carefully describe the sampled population correctly and in detail with respect to geographic origin and demographic background applying termini from molecular anthropology and population genetics. This includes the use of a correct ethonym [sic] (e.g. 'Roma' instead of 'Gypsy', 'Europeans' instead of 'Caucasians', etc.), the definition of the linguistic, and (if applicable) cultural groups (e.g. casts) and subgroups" (Parson and Roewer 2010, 506).

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

different. The public will find a story plausible that fits their understanding of groups and groupness. It is more difficult to publish counter-intuitive population histories, in particular if these stories do not resonate with what human evolutionary genetics textbooks say (e.g. Jobling 2014, 448).

In addition, the Roma – understood as a genetic population – are perceived to have common ancestry components supposedly making them distinguishable from "Europeans."[22] Thus, the focus of most these genetic studies is on the Indian origin. The authors aim at demonstrating genetic continuity between the group that migrated from India to Europe in medieval times and today's Roma. Some authors call the group that departed from India "proto-Romani." Romani who have been living in Europe ever since that original departure are all viewed as descendants of these "proto-Romani." Social integration, or any other social situation leading to "genetic mixing," makes the task of the geneticists more difficult. However, if one assumes that "mixing" has been negligible and that it has not eroded the group coherence as such – or that at least the core part of that group has remained intact and "unmixed" – then "mixing" can presumably be controlled for in a *model*. As one author states: "The basic common model considers a proto-Romani population that splits from a given population of the Indian subcontinent (Pakistan and India) and can admix with a hypothetical (unsampled) Central Asian, or Near or Middle Eastern population, as well as with non-Romani Europeans after arriving in Europe" (Mendizabal 2012, 2345).

When the authors of these population genetic studies look at DNA data collected from living Roma individuals over the past thirty years, what they understand themselves to be looking at is a proxy towards a hypothetical, ancestral "proto-Romani" population (see the quote above). Studying Roma migration routes over Europe, they are interested in dispersal and subfounder effects in single national contexts. In each case, they look for individuals who would most closely resemble the medieval Indian-Romani arrivals in that country, coming from Eastern Europe or the Near East. Of course, the most revealing markers for this would again be "Indian signatures" (Gomez-Carballa et al. 2013), marking those who are descendants of the first Romani arrivals in the respective country.

However, finding "Indian signatures" does not mean that large or significant parts of the genome are resembling the genomes of people from India, rather than those of others. Neither do readers learn about the results from the most powerful markers. In fact, in studies considering "mixture," an overwhelmingly large proportion of the genome of an average person sampled as Roma does not resemble "Indian signatures," but "European" or other "signatures." Yet, in the research questions, research designs and interpretation of the results, in reports on their findings in the conclusions or in press releases, the authors emphasize and focus on "Indian signatures" above all others (Mendizabal et al.2012; Melegh et al. 2017), whereas "signatures" from other regions are mentioned only briefly and marginally. For further research, what seems most promising to them is to explore the Indian ancestry further.

Mendizabal et al. (2012), for example, speak of admixture along the way from India to Europe and within Europe, but this seems secondary and negligible for the authors. Because their priority is to find a more specified regional origin of the Roma population within India, the authors match their Roma DNA data with DNA data from different areas in India. The applied computer program suggests that the "parental population" came from Kashmir or north/north-west India. As the authors admit, they are struggling with a lack of samples precisely from that region. Hence, "future dedicated sampling across linguistic and social strata in this Indian subregion is

[22] Or "Caucasians" (a term mostly used in biomedical studies), "Whites" (e.g. Castella et al. 2011; Varszegi et al. 2014), "indigenous" (e.g. Pamjav et al. 2011), or "non-Roma." "Non-Roma" and "Caucasian" are sometimes used interchangeably (e.g. Mašindová et al. 2015 and Molnar et al. 2012). This interchangeable use of categories suggests that in some cases "non-Roma" is used as a way of coding racial division.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

needed to identify the actual parental population of the European Romani from that Indian subregion" (Mendizabal et al. 2012, 2347).

To be sure, we do not maintain that there was no migration from India to Europe in medieval times. We also do not simply reject the claim that medieval migrants from India are among the ancestors of some of the Roma living in Europe today. However, this is only one source of their ancestry, and not even the dominant one. Like other populations, Roma have multiple sources of ancestry. Their genetic ancestry is manifold and complex and does not allow for a single historical narrative. Furthermore, in India and Europe, in between these regions and around the world, many people might have comparable ancestry but are not considered Roma.

The example of the Bulgarian Romani

Mendizabal et al. (2012) aimed at determining the temporal sequence of arrival in various countries. Therefore, they "attempted to identify the current Romani population that is genetically the most similar to the putative founder population of all European Romani groups," that is, the "actual parental population" in India (Mendizabal et al. 2012, 2345-6). As a result of this attempt, the authors conclude that Bulgarian Romani seem to be most similar.

In this study, the estimated 750,000 Bulgarian Roma (roughly 10% of the Bulgarian population) are represented by 18 individuals. The sampling information in the supplement is more detailed for the Bulgarians than for all other Roma populations. "The Bulgarian Romani samples were collected from the two major groups around the country: Wallachian and Yerli, and some of their subgroups (Dassikane, Horohane, Kaldarashi, Kopanari and Reshetari)" (Mendizabal et al. 2012, supplement, 1). No reference is given for the DNA data from Bulgaria, but in the acknowledgements, Ivailov Tournev is thanked for the "recruitment of Romani samples from Bulgaria."

In 2016, Ivailov Tournev published an account of his two decades long sampling endeavors in Bulgaria. The text's subtitle reads "Neuromuscular disorders in Roma (Gypsies) – collaborative studies, epidemiology, community-based carrier testing program and social activities" (Tournev, 2016). Starting in 1994, Tournev and his team collected the most detailed information on Bulgarian Roma, in cooperation with ethnographers:

> The main sources for collecting the epidemiological information were the field work studies. A neurological screening of hereditary neuromuscular disorders using the method "door to door" was performed in 2500 towns and villages (having predominantly Roma population) in the country. Those towns and villages where pedigrees with hereditary neuromuscular disorders resided were visited from 2 to 10 times with the aim of collecting pedigree information, blood samples for genetic studies and neurological examination of the patients. The field work studies covered a period of 20 years (1994–2014). 97% of the Roma population living in compact Gypsy quarters was encompassed. An ethnographical and linguistic examination was performed in every quarter using a semi-standard interview for identification of various Roma groups and subgroups. In those towns and villages where Roma people live in several quarters or more than one Roma group resides, the ethnographical and linguistic examinations were performed in every quarter and in every separate group. The field studies were performed with the support of the local ‚Roma' foundations and Roma health mediators from different parts of the country. (Tournev 2016, 99)

If 97% of the Roma population living in "compact quarters" were encompassed, as Tournev acknowledges in this quote, one could perhaps call this a "genetic census." Tournev gives a detailed account of the terrible living conditions of most Bulgarian Roma, including their segregation from the major society. More so than in other countries, Bulgarian Roma live in isolated settlements, and Tournev is explicit about the majority's role in creating genetic isolates by forced exclusion.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Hence, data on genetic isolation is rich in Bulgaria, to an extent that allows for a more differentiated view on single small groups, some of which have been more isolated than others, for different reasons. The results speak for a complex substructure, more complex and more variegated perhaps than elsewhere. How the 18 individuals from Mendizabal et al.'s study were recruited, or how their data was selected from many thousands, is an open question.

Tournev's text is rich in information, particularly about neuromuscular genetic diseases among Bulgarian Roma. A large number of studies were published from the collected data, in the fields of both medical and population genetics. The most productive main and senior author of these publications was Luba Kalaydijeva. Kalaydijeva involved two ethnologists in the research, and one of their tasks was to work out a classification for the different Roma groups.

The instances of data sharing with other teams across Europe and the world are numerous. For example, there is hardly any population genetic study on Roma that does not include data from the Bulgarian large-scale collection published by David Gresham, a former PhD student of Kalaydijeva, in co-authorship with Tournev, Kalaydijeva and a larger team (Gresham et al. 2001). Many studies have built "European Roma" samples by adding a small number of samples from other countries to the already existing Bulgarian data.

Accordingly, DNA data from Bulgarian Romani provide the most detailed, variegated and rich data collection, and it has come to be interpreted and used in many different ways. It is hard to imagine that Kalaydijeva, who has time and again emphasized the great and complex genetic diversity of Bulgarian Romani in a number of publications, would agree to represent Bulgarian Roma with 18 individuals from a low number of subgroups. Also, if no rationale is given for the selection of those 18 individuals from a supposedly huge number of data sets, it is hard to tell what these individuals stand for.

After all, the samples from Bulgarian Romani of today, or of 1994, cannot stand for any other national, transnational or regional group. Neither can they represent the "ancestral" Bulgarian Roma population, the one that supposedly arrived in medieval times in the region that is today the nation state of Bulgaria. If Bulgarian Roma have been isolated in many small communities, in varying constellations over the centuries, their current genetic diversity can be shaped by complex drift processes and is not the result of neatly definable, linear historical processes. The underlying history is inextricably complex and locally contingent.

Merging datasets across countries

A number of genetic studies seek to extend their scope across all of Europe, for which purpose they merge, share, transfer and reduce data of different provenance, collected with different sampling strategies. What does such a merging strategy imply? And what can such a merged sample achieve?

In each medieval principality, in each Early Modern Times empire, and in each modern nation state, foreigners – including those whom geneticists regard the ancestors of today's Roma – were treated, named and registered differently. Over the centuries, with political upheavals and wars, state borders and registration procedures changed and shifted, including and excluding minorities in different ways. Yet assuming genetic continuity, population geneticists tend to ignore such complexities, extract samples from a number of countries and look for patterns on maps of Europe. Their aim is to investigate Roma migration routes along with subfounder effects.

From the ca. 75 population genetic studies published after 1990, two might suffice to demonstrate this attempt at representing European Roma with a merged data set. Mendizabal et al. (2012) include a map of Europe with the nation states of today. It shows how many Roma live in each country, how many individuals were sampled per country, and at what historical date Roma were first mentioned in that country. The overall data set of 152 individuals, collected from thirteen Roma groups in thirteen countries, includes small samples, such as seven individuals from

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Wales (but these were excluded from some of the analyses because they seemed too admixed), ten from Spain, nine from Portugal, eight from Lithuania or seven from Estonia. The largest sample contains the eighteen from Bulgaria, followed by fourteen from Romania. The sample sizes stand in no correlation to the size of the respective Roma population, nor to their proportion of the overall national population. Following up the numbers reveals that some individuals were excluded from each national sample due to familial relationships.

No sample, neither in this nor in any other population genetic study on Roma, comes from Germany, France, Italy, Belgium, the Netherlands, Poland, Austria or Switzerland. This is not due to the small size of their Roma population - some of these countries have much larger Roma populations than Lithuania, Portugal or Ukraine. No reason is given for this sampling decision in any of the studies.

In one study, titled "Mutation history of the Roma/Gypsies," samples from Germany, France and Italy are mentioned, yet not marked as such in the data analysis (Morar et al. 2004). In order to account for this, we need to first look into general sampling decisions. "In this study," the authors state in the abstract, "we have used five disease loci harboring private Gypsy mutations to examine some missing historical parameters and current structure. We analyzed the frequency distribution of the five mutations in 832–1,363 unrelated controls, representing fourteen Gypsy populations, and the diversification of chromosomal haplotypes in 501 members of affected families" (Morar et al. 2004, 596).

Representing a population by mutations presents further problems for representativity in a population history study ("private Gypsy mutations" will be discussed in more detail below). If a third of all recruited individuals carry one out of five mutations understood by geneticists as "private Gypsy mutations," their data has likely either been collected in large scale screening programs or in doctors' offices and clinics.

This study comes from Kalaydjieva's lab, and Bulgarian samples make up a large part of the dataset. "Self-reported identity" in terms of "historical and cultural-anthropological classifications" was used in order to sort the individuals into group categories (Morar et al. 2004, 598). In total, a table states, 1175 individuals from fourteen "Gypsy groups" were sampled and their data was assembled in three large "migrational/linguistic categories": 419 individuals labeled "Balkan"; 366 labeled "Vlax"; and 390 labeled "Western European" (ibid). While the former two categories are subdivided into groups with local or professional names ("Musicians," "Kalajdjii"), the latter, "Western European," is subdivided into national groups: Hungarian (283 individuals), Lithuanian (20), Spanish (87) – in total, as stated above, 390 individuals.[23] However, the reader also learns that "individuals from Hungary, Slovenia, the Czech Republic, Lithuania, Germany, France, Italy, Spain, and Portugal, for whom information on Gypsy group identity was unavailable, partial, or contradictory, were classified together as western European" (ibid).

This raises further questions: Were the individuals from Germany, France and Italy put in the Hungarian, the Lithuanian or the Spanish sample? Based on what significant criteria? How many were there? Were they approached as mutation carriers in healthcare facilities, and then asked to identify as "Gypsies"? Or were they approached as "Gypsies," and if so, under which sampling scheme? What does it say about "self-reported identity" if information on "Gypsy group identity" was "unavailable, partial or contradictory"? Why was this the case for all individuals from these three countries? In none of these countries does census data collection include ethnicity. Did this play a role? – Or, one could also ask: Would a German patient self-identify as "Gypsy"? Would a German ethics board be comfortable with approving of applications for projects involving "Gypsies"? And would a German self-reported Roma approve of being sorted into a Hungarian, Lithuanian or Spanish "Gypsy" sample? And where are these samples and datasets today?

[23] Interestingly, if unusually, Hungary, Slovenia and Lithuania are included in the category "Western European,"

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

As we have noted above, such studies merge, share, transfer and reduce data of different provenance and collected with different sampling strategies. What the recruited individuals probably all have in common is the fact that they are not well integrated into their nation state's societies. The sampling practices in many of these studies seem to have favored "societally deprived groups," serving as a proxy to "genetic isolates." If this rationale underlies sampling decisions, it is hardly surprising that most research results confirm genetic isolation. If research teams simply use data sampled by another team without questioning the sampling strategy, they will reproduce the first team's results and biases.

One important sampling rationale seems to have been maximizing the likelihood of genetic variants that have also been found in India. One can perhaps increase that probability by focusing on people who have a certain genetic disease, who look Indian, speak Romani, and identify as Roma. But the merged sample cannot represent "European Roma"; it represents isolated groups, families or neighborhoods that are labeled "Gypsy" or "Roma," and a certain proportion of the recruited individuals may even identify as Roma. How many, and under what circumstances, remains unknown.

"Private Gypsy mutations"

The Roma are described as an isolate in which various "private mutations" for all kinds of diseases, especially neuromuscular diseases, have "accumulated," more than in any other group. Brubaker argues that "rare variants" would "not be *definitive* of any socially defined racial category" (2015, 82).

The term "private mutation" is viewed as a purely technical term by human geneticists. It refers to any novel mutation that has been found in a narrow social group, for example, in a family, between relatives or in an isolated rural settlement. Speaking of "private Gypsy mutations," however, as many of the medical genetic studies on Roma do, gives the term a different resonance. "Private," in this context, takes on a more metaphorical meaning and also invites an interpretation of mutation carriers as *belonging* to that ethnic group, or at least as having ancestors from that group.

And indeed, in some studies, the ethnic attribute "Roma" is being assigned on the basis of disease mutations even though the patients have self-declared a different ethnicity. For example, the authors write about a patient with Hereditary Motor and Sensory Neuropathy (HMSN), a rare genetic disease attributed to Roma, noting that "the family was not aware of their Roma ancestor" (Brožková et al. 2016, 2). Neither do the authors consider that the mutation could also occur in non-Roma individuals. Similarly, Colomer et al. (2000) examine three Spanish patients with HMSN Lom disease and argue that the patients "belong to a non-consanguineous family with Gypsy background although they were unaware of the details of their ancestry" (578). Some deleterious genetic mutations are referred to as "Gypsy mutations" even though "mutation screening in 359 Eastern-European Gypsies failed to identify any carriers" (Barca-Tierno et al. 2011, 1218). Speaking of "private Gypsy mutations" also implies that the Roma are the population in which that mutation first emerged, or that they are the source population of a mutation brought over from India to Europe. From the 220 biomedical studies reviewed, only a handful mention that a rare disease could have been introduced into Roma communities from outside.[24]

This general depiction of Roma is, of course, misleading. The vast majority of mutations labeled "private Gypsy" or "private Roma" have not been shown to be more prevalent in India, or to be confined to Roma communities.[25] Seen from the perspective of human genetics and evolution, mutations can also first occur in a surrounding majority population; a subsequent ghettoization of "undesired" population groups – disrespectfully excluded as foreigners, poor, diseased, disabled,

[24] These are Angelicheva et al. 1997; Desviat, Perez and Ugarte 1997, 67; Morar and Kalaydjieva 2008; Kalanin et al. 1994.

[25] A good dozen of such "private" mutations are reported in the literature.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

deviant – can lead to the amplification of mutations in a societally isolated community. In spite of ethnic and social complexity, such a community may nevertheless be labeled “Gypsy” or “Roma” by the majority in a society. Writing the history of Roma migration routes by means of mutations, then, is a representational challenge.

A disagreement on representativity

As we have demonstrated, genetic studies that claim to have produced research results about “the Roma” or “the Gypsies” cannot formally represent the overall Roma population. The sampling schemes and practices do not satisfy the standards of representativity - not in the social sciences, nor in some branches of the life sciences. The authors of these studies would have to state precisely what the sampling decisions are aimed at and what the sample then represents: for example, people living in most isolated places, or people who might have Indian ancestry, or people who carry a mutation for one or more genetic diseases, or all of these, if applicable. Alternatively, if the authors stick to representing “all European Roma,” they would have to adopt a whole new sampling scheme, one that takes on the challenge of representing a relatively large and “superdiverse” group. This could not be done without extensive discussion with social sciences and humanities scholars, and, even more importantly, not without active involvement of Roma themselves. Representing Roma in such a participative and complex endeavor could lead to outcomes that are not easily predictable given the different perspectives and interests of the stakeholders. But with other vulnerable populations considered interesting for genomics, such participative options are already underway (e.g., Kowal and Radin 2015).

However, we believe that a closer look at the heart of the disagreement about representativity is warranted. The conflation of the population geneticists’ research objects – genetically bounded populations – and social or political population labels runs deeper and is more widespread than in the limited field of studies on isolated populations. The conflation of categories, markers and population labels in genetic research with notions of ethnicity and race has been addressed in multiple critical studies (Fujimura and Rajagopalan 2011; Fujimura et al. 2014; Nash 2013; Fortier 2012; Koenig, Lee and Richardson 2008; Schramm, Skinner and Rottenburg 2012). The four-grandparents-sampling approach is practiced widely and documented in recruitment guidelines and consent sheets of, for example, the 1000 Genomes Project. Nash describes how the “People of the British Isles Project,” by aiming at recruiting people with ancestry that fit the rationale, focused on rural, “rooted,” and white (Nash 2013, 201).

As Nash (2013) warns, one needs to watch carefully for the omissions that such sampling schemes entail. Of course, such a sample cannot represent any population, neither a historical nor a present one. It only represents a certain portion of a population that practices a specific social behavior. At many times, in many places, many people did not live close to where their grandparents lived, but migrated or were displaced, or practiced mobile or commuting life styles. Excluding these people from sampling schemes means excluding specific parts of a population.

Exclusions in the lab, if an individual DNA data set fails to meet the expectations, are also not restricted to the Roma studies. The famous Novembre et al. study of 2008, claiming the representation of “European population structure,” stated: “We applied various stringency criteria to avoid sampling individuals from outside of Europe, to create more even sample sizes across Europe, to exclude individuals with grandparental ancestry from more than [one] location, and to avoid potential complications of SNPs in high linkage disequilibrium,” adding that “these numbers exclude individuals who reported mixed grandparental ancestry, who are typically assigned to locations between those expected from their grandparental origins” (Novembre et al. 2008, 98).

If such data cleansing operations are implemented, the scope of the claim cannot extend to represent a living population, such as all Europeans, or European population structure in general. Such samples only represent people whose ancestors all come from the same group or region. As a

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

result, regions where marriage (or reproduction) between partners from two geographically distant regions is rare – as in rural regions, for example – come to be better represented than others. If, however, population geneticists stick to their goal of representing certain populations as groups genetically bounded over long time periods, then the small sample sizes they deem sufficient require further thought, if tapping into population substructure is to be avoided.

Contextualizing our case study in the critical interdisciplinary literature on human genetic variation research, we note that it provides an extreme case of what those specialists have warned against, from both an ethical-political and a conceptual-methodological perspective. It is an extreme case of problematic extrapolation, as Fujimura et al. have described:

> Human geneticists make decisions about which subset of individuals to use to "represent" a "race" or "national group" in their sampling procedures and in their cluster analysis. The subsets they use are obviously extremely small compared to the number of individuals who identify with that race or nationality label. They thus extrapolate their results from a small number of individuals to make inferences about a vastly larger number of individuals who self-identify with the same race or nationality label and whose genetics have not been studied. (Fujimura et al. 2014, 215)

It is a demonstrative case of what Catherine Nash (2013) has described as a problematic trend in which "continental and regional ancestries" are "genetically identified and described as bounded natural categories" (203).

Our case study on the "new genetics", focusing on a strand of research on Roma, adds to the growing STS literature criticizing essentialist and racialized versions of grouping humans through genetic accounts (Schramm, Skinner and Rottenburg 2012). Many STS scholars have addressed the particularities of the geneticization of minority identities and its interplay with social and political understandings of race and ethnicity (e.g. Egorova 2010; Kowal, Radin and Reardon 2013; Kyllingstad 2012; Tallbear 2013; Wade et al. 2014). Yet so far, Roma have been omitted from this critical examination, even though genetic studies on Roma show stunning conceptual continuities since their inception almost a century ago (Lipphardt 2016).

For our own perspective on groupness, we follow Hacking's ([1986]1999) "dynamic nominalism" which asserts that categorization and labeling are constitutive for a group's social formation and dynamics. To be sure, such a position is not one of a naïve constructivism denying groups as real entities; as groups are socially defined, historically contingent and changing, codified into legal systems, embedded in administrative and techno-scientific assemblages, self-internalized or rejected and ubiquitous objects of everyday politics, they are indeed real and consequential (ibid).

From an STS perspective, "populations" (the postwar conceptual replacement of "race" in human population genetics) are not natural kinds. Their genetic profiles follow from the models and technologies used to measure similarity and difference, as well as from the assumptions and decisions that have been made throughout the research process. Sampling strategies, genetic markers and reference groups chosen for comparisons, all these may shape the ethnic groups which genetic work purports to merely describe (M'Charek 2005). The alignment of genetic data to socio-political relevant racial and ethnic categories appears less an effect of data aggregation, but rather reflects practical, pragmatic, conceptual, methodological, theoretical and socio-politically relevant choices the researchers make (Bolnick 2008; Duster 2015; Fullwiley 2008; Gannett 2003; Lee et al. 2001; M'charek 2005). As STS scholars have demonstrated, genetic classification and social order are not separate endeavors, but they are co-produced in entangled projects of race and ethnicity (Reardon 2005; Tallbear 2013). It is worth noting that in some cases, genetic research targeting minority groups is carried out by geneticists who self-identify as members of these groups for whom they seek social justice, political and medical attention (Fullwiley 2008, Bliss 2015). However, while genetic research on minority groups can have empowering effects, the case of the Roma is just one of many cases in which it has no such effect,

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

but does have the potential to feed into socially divisive processes (Egorova 2015; Kent et al. 2015; Santos, Da Silva and Gibbon 2014; Wade et al. 2017).

The genetic studies we have examined could be the starting point for further questions and debates, ranging from the methodological and conceptual to the ethical and societal. For Roma, STS or humanities scholars could discuss a wealth of topics that have been raised with regard to other populations in the past few years. These include, to name but a few, the branding and commodification of unique populations (Reardon 2017; Tupasela 2016; Tarkkala and Tupasela 2018), or the creation of biovalue and biocapital around the many existing cell lines and biobanks with tissue samples from these groups (Birch and Tyfield 2013), or the "biomedicalisation" of a much discriminated minority, its history and corporeality (Clarke 2014), or how the genetic framing of Roma will transform the policies of national states or the European Union. One could also contextualize this case study within the work on other isolated populations, their specific ethical challenges (Mascalzoni et al. 2010), and how they have been met with more sensitivity in other cases (Floersch, Longhofer and Latta 1997; Lindee 2005). We believe that the case of the Roma offers new insights and perspectives for this strand of research; in turn, we also hope that STS has something to offer to the Roma.

Clearly, many of the problematic aspects we have pointed out here are not specific to the research in Roma communities, but of a more general scope in biomedical research. Questions like how to best collect, curate and share data, how to best protect the privacy of donors, how to gain and hold the trust of patients and donors, how to deal with issues of property, how to handle attrition and documentation, how to speak to and about patients or test subjects in sensitive and responsible ways, how to involve them and how to guarantee benefit sharing, how to meet the reproducibility crisis - these are issues widely discussed in research institutions and ethics committees as well. All large-scale research projects involving humans have to grapple with these problems.

Yet for the genetic studies about Roma, or other such vulnerable groups, we note three specificities. First, in a regular national health study in, say, Germany, a participant is not, or at least not to the same extent, at risk of being stigmatized on the basis of their nationality or "origin" or "ancestry." Second, their privacy risk is generally a given – no data is perfectly safe – but not a heightened one; information on their home village will not be published along with their genetic data. Thirdly, in the studies on Roma, many of the problematic issues we have noted here appear at once, and sometimes in cumulative ways.

A population that is represented along these lines is at heightened risk, not only for stigmatization and privacy violation, but also for application errors in medical and forensic genetics. Representativity is hence a matter of concern - for the scientists who want to avoid flawed research results and the negative consequences thereof for the contributing volunteers; for the group, in this case the Roma, who have a right to benefit from scientific progress without being exposed to further harm; and for social sciences and humanities scholars writing about the risks of essentialization.

Outlook

What can be said about the usefulness of this strand of genetic research? What is its potential? Who should have an interest in genetic studies of vulnerable groups?

Apart from the obvious benefits for the scientists involved – interesting research questions, topics for PhD theses, data to draw upon – and the benefits for society – increased understanding of genetic diseases, development of new therapies – the Roma themselves could benefit from this research, if that was a built-in goal on the side of researchers, healthcare providers and public support programs. Providing Roma with access to all the data dealing with those disorders that are particularly prevalent in Roma groups would further the goal of "patient empowerment,"

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

allowing them to use all the new findings to their own benefit (e.g. within the scope of genetic diagnostics). This, of course, requires free access to health services and increased resources for supporting Roma families and communities. Furthermore, for an individual to seek the advice of a genetic counselor, trust is a crucial prerequisite. That individual must feel secure that their data will be handled with adequate care. However, though Roma may theoretically be among those who could benefit from medical therapies developed on the basis of their data, a geneticist who has studied DNA data from Roma acknowledges that, for the moment: "Most studies have remained in the realm of scientific exploration, away from the health needs of the Roma" (Kalaydjieva, Gresham and Calafell 2001, 2).

With regard to forensic genetics, Roma could benefit if the databases used for frequency assessment were not biased against them (Lipphardt, Rappold and Surdu 2022, under review). Though Roma could certainly have an interest in learning about their own history from population history genetics, the narrative currently dominating these studies may have an essentializing, stigmatizing and exclusionary effect, while other scenarios that could perhaps also be backed by the data are not considered.

Providing one's data in the hope of becoming a recipient of benefit rather than of harm is a matter of trust. Trust in genomics is obviously not easy for vulnerable and isolated groups: not just because they are vulnerable, but because their privacy and stigmatization risks are much higher than those of majority populations, and because geneticists might be so fascinated by a specific population that they lose sight of the community's own perspective and interests. We therefore do not share the optimism of Brubaker (2015) with regard to the de-essentializing power of genomics. As long as vulnerable groups are exposed to the kind of representation we have observed, there is still a great deal of (inter- and transdisciplinary) work to be done. For example, it takes extensive efforts to assure vulnerable groups that they will not be misrepresented.

One could ask a wealth of historical questions about overlaps and continuities between race science and population genetics. As the sampling schemes studies target an "unmixed" core proportion of an otherwise much more fuzzy population, the authors seem to assume and favor an ideal type of representative individual for that population. Whether the conceptual justification employs the term "race" or the term "population," whether the data consists of anthropometric measurements or DNA data - this is typological thinking in the paradigm of population genetics.

It would also be legitimate and necessary, in our view, to pose the question of how this research could relate to racist discrimination of Roma. More urgent, perhaps, is a discussion of genetic essentialism, or genetic determinism, and the many possible implications and consequences for those who are labeled Roma. Applications of genetic technologies are to be expected – if they are not already in place – in citizenship issues, in law enforcement, and on the biomedical market. There is a vast abundance of laudatory statements in these studies about how valuable a tool, how unique a resource the Roma are for the geneticists, for the investigation of rare and complex diseases, for forensic technologies - and, one might add, also for patient organizations, pharma companies, and biomedical investment.

For the sake of focus, we have deliberately not raised these questions here, but we think they need to be discussed. Though we have very selectively focused on questions of representativity and sampling, we emphasize that many of those methodological themes are tightly connected to questions of ethics, privacy, justice, and benefits. We do not mean to suggest that the practices behind these studies per se have an unethical bias. There might indeed be a respectful and supportive engagement with individuals in these studies. As we said in the beginning, in spite of the manifold problems we have pointed out, we do note a growing awareness in the genetic studies over the past twenty years towards more regular ethical procedures, caution in terminology, and self-identification as a basic principle in recruitment. Perhaps, in some publications, the nebulous silence around sampling is in fact an expression of care or concern about how to protect individuals from harm, or from being exposed. However, we argue that silence, in this case, contributes to

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

misrepresentation. Not addressing the challenges does not reduce the harm, it rather adds to the risk of inflicting more harm.

While isolated groups are among the most promising ones for genetic research, these groups are often also the most vulnerable. Recognizing that such communities, families and neighborhoods exist does not mean accepting that "all European Roma" can be represented by samples from those places. Even more importantly, recognizing that such communities, families and neighborhoods exist under highly precarious conditions comes with incredible responsibilities. How are we to engage with these communities and individuals? to support and protect them? to represent them fairly and correctly? Which way – according to the Roma themselves – would be the best way to find out about their genetic "history"? Is it in their interest at all to know about it? And is it in their interest to have data on genetic predispositions to specific diseases possibly enriched in the genomes of some members of some families or communities? Whom else would they agree to share such knowledge with? What are the risks and downsides of knowing? Increased awareness and new institutional forms of recognition are needed, as the challenges of representing vulnerable populations have not yet gained the full attention they deserve.

Acknowledgements. We are grateful for a DFG grant on our former project *The Genetic Construction of Roma Groupness and its Interdisciplinary Entanglements* (2016-2019). It was during this project that we built up a large part of our database with genetic studies on Roma and developed contextual information relevant to our analysis. Many thanks to Anja Reuss and her colleagues from the Zentralrat der Deutschen Sinti und Roma, and to Frank Reuter from the Forschungsstelle für Antiziganismus. We thank Yves Moreau for the insightful conversations on ethical aspects related to genetic data collection. We thank Leon Kokkoliadis and Eric Llaveria Caselles, student assistants at *Max Planck Institute for the History of Science,* who contributed by collecting genetic papers focusing on Roma at the beginning of our project (2014-2015). We thank Cedric Bradbury and Sarah Weitz, student assistants at University College Freiburg, who helped with collecting, sorting and organizing the database of genetic studies of Roma (2016-2020). We are grateful to Silvia Stößer who helped us with various administrative tasks. We had inspiring discussions about genetic studies on Roma with Anna Lipphardt, Nicholas Buchanan, Amade M'charek, Huub van Baar, Ildikó Plájás, and many colleagues at University Freiburg and in other places. Our understanding of the genetic studies benefitted from FRIAS funding during 2018/2019, and particularly from very fruitful and insightful encounters with our colleagues from the interdisciplinary project group at Freiburg Institute for Advanced Studies (FRIAS): Anna Köttgen, Anne-Christine Mupepele, Peter Pfaffelhuber and Fabian Staubach. We are grateful for numerous conversations with Matthias Wienroth, Sabine Lutz-Bonengel, Ulrike Schmidt, Till Andlauer, Tino Plümecke and Thomas Schulze on this topic. We have received helpful comments on this text from Denise Syndercomb Court, Nils Ellebrecht, Tino Plümecke, Nicholas Buchanan, Anna Lipphardt and our students of the seminar "Genetic studies of vulnerable populations." We are grateful for advice from the DFG-Senatskommission für Grundsatzfragen in der Genforschung, in particular from Brigitte Schlegelberger.

Bibliography

About, Ilsen. 2012. "Underclass Gypsies. An Historical Approach on Categorisation and Exclusion in France, in the Nineteenth and the Twentieth Centuries." In *The Gypsy 'Menace': Populism and the New Anti-Gypsy Politics*, edited by Michael Stewart, 95–117. London: Hurst & Company.

Achim, Viorel. 1998. *The Roma in Romanian History*. Budapest, New York: Central European University Press.

Acton, Thomas A. 2015. "Scientific Racism, Popular Racism and the Discourse of the Gypsy Lore Society." *Ethnic and Racial Studies* **39** (7):1187–1204.

Angelicheva, Dora, Francesc Calafell, Alexey Savov, Albena Jordanova, Annie Kufardjieva, Vania Nedkova, Tanya Ivanova, et al. 1997. "Cystic Fibrosis Mutations and Associated Haplotypes in Bulgaria: A Comparative Population Genetic Study." *Human Genetics* **99** (4):513–20.

Barca-Tierno, Verónica, Miriam Aza-Carmona, Eva Barroso, Damia Heine-Suner, Dimitar Azmanov, Jordi Rosell, Begoña Ezquieta, et al. 2011. "Identification of a Gypsy SHOX Mutation (P.A170P) in Léri-Weill Dyschondrosteosis and Langer Mesomelic Dysplasia." *European Journal of Human Genetics* **19** (12):1218–25.

Benjamin, Ruha. 2009. "A Lab of Their Own: Genomic Sovereignty as Postcolonial Science Policy." *Policy and Society* **28** (4):341–55.

Berescu, Cătălin. 2019. "How Many Ghettos Can We Count? Identifying Roma Neighbourhoods in Romanian Municipalities." In *Racialized Labour in Romania*, edited by Vincze Enikö, Norbert Petrovici, Cristina Raţ, and Giovanni Picker, 179–205. Cham Springer International Publishing.

Birch, Kean and David Tyfield. 2013. "Theorizing the Bioeconomy." *Science, Technology & Human Values* **38** (3): 299–327.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Bliss, Catherine. 2015. "Science and Struggle: Emerging Forms of Race and Activism in the Genomic Era." *The ANNALS of the American Academy of Political and Social Science* **661** (1): 86–108.

Bliss, Catherine. 2018. *Social by Nature. The Promise and Peril of Sociogenomics*. Stanford: Stanford University Press.

Bolnick, Deborah. 2008. "Individual Ancestry Inference and the Reification of Race as a Biological Phenomenon." In *Revisiting Race in a Genomic Age*, edited by Koenig Barbara, Sandra Soo-Jin Lee, Sarah Richardson, 70–85. New Brunswick, New Jersey and London: Rutgers University Press.

Bogdal, Klaus-Michael. 2011. *Europa erfindet die Zigeuner: eine Geschichte von Faszination und Verachtung*. Berlin: Suhrkamp.

Brožková, Šafka, Dana, Jaroslava Paulasová Schwabová, Jana Neupauerová, Jana Sabová, Marcela Krůtová, Vladimír Peřina, Marie Trková, Petra Laššuthová, and Pavel Seeman. 2016. "HMSN Lom in 12 Czech Patients, with One Unusual Case due to Uniparental Isodisomy of Chromosome 8." *Journal of Human Genetics* **62** (3): 431–35.

Brubaker, Rogers. 2015. *Grounds for Difference*. Cambridge, Mass.: Harvard University Press.

Brubaker, Rogers. 2004. *Ethnicity without Groups*. Cambridge: Harvard University Press.

Castella, Maria, Roser Pujol, Elsa Callén, Juan P. Trujillo, José A. Casado, Hans Gille, Francis P. Lach, et al. 2011. "Origin, Functional Role, and Clinical Impact of Fanconi Anemia FANCA Mutations." *Blood* **117** (14): 3759–69.

Cazacu, Catalin, Cristina Chinole, Monica Hancianu, and Vasile Astarastoae. "Personalized Medicine for Whom? The Situation of Romani People." 2013. *Revista Romana de Bioetica* **11** (3): 84–91.

Cell Press. 2012. "European Romani Exodus Began 1,500 Years Ago, DNA Evidence Shows." *EurekAlert*, 6 December.

Clarke, Victor Alan.1973. "Genetic Factors in Some British Gypsies." In *Genetic Variation in Britain*, edited by **Derek Frank Roberts and Eric Sunderland**, 181–195. London: Taylor & Francis.

Clarke Adele. 2014. "Biomedicalization." In *The Wiley-Blackwell Encyclopedia of Health, Illness, Behavior, and Society*, edited by William C. Cockerham, Robert Dingwall, and Stella R Quah, 137-142.

Colomer, Jaume, Cristina Iturriaga, Luba Kalaydjieva, Dora Angelicheva, R. H. M. King, and P. K. Thomas. 2000. "Hereditary Motor and Sensory Neuropathy-Lom (HMSNL) in a Spanish Family: Clinical, Electrophysiological, Pathological and Genetic Studies." *Neuromuscular Disorders* **10** (8): 578–583.

Council of Europe (CoE). 2012. "Descriptive Glossary of Terms Relating to Roma Issues." Ver. 18 May 2012. http://a.cs.coe.int/team20/cahrom/documents/Glossary%20Roma%20EN%20version%2018%20May%202012.pdf (last accessed March 14, 2022).

Desviat, Lourdes R., Belén Pérez, and Magdalena Ugarte. 1997. "Phenylketonuria in Spanish Gypsies: Prevalence of the IVS10nt546 Mutation on Haplotype 34." *Human Mutation* **9** (1): 66–68.

Donert, Celia. 2008. "'The Struggle for the Soul of the Gypsy': Marginality and Mass Mobilization in Stalinist Czechoslovakia." *Social History* **33** (2): 123–44.

Duster, Troy. 2015. "A Post-Genomic Surprise. The Molecular Reinscription of Race in Science, Law and Medicine." *The British Journal of Sociology* **66** (1): 1–27.

Egorova, Yulia. 2015. "Theorizing 'Jewish Genetics': DNA, Culture, and Historical Narrative." In *The Routledge Handbook of Contemporary Jewish Cultures*, edited by Roth, Laurence, and Nadia Valman, 353–364. London and New York: Routledge.

Egorova, Yulia. 2010. "De/Geneticizing Caste: Population Genetic Research in South Asia." *Science as Culture* **18** (4): 417–34.

Ehler, Edvard, and Daniel Vanek. 2017. "Forensic Genetic Analyses in Isolated Populations with Examples of Central European Valachs and Roma." *Journal of Forensic and Legal Medicine* **48** (May): 46–52.

Epstein, Steven. 2007. *Inclusion: The Politics of Difference in Medical Research*. Chicago: University of Chicago Press.

Fiatal, Szilvia, Réka Tóth, Ágota Moravcsik-Kornyicki, Zsigmond Kósa, János Sándor, Martin McKee, and Róza Ádány. 2016. "High Prevalence of Smoking in the Roma Population Seems to Have No Genetic Background." *Nicotine & Tobacco Research* **18** (12): 2260–67.

Filhol, Emmanuel. 2013. *Le Contrôle Des Tsiganes En France, (1912-1969)*. Paris: É;d. Karthala.

Floersch, Jerry, Jeffrey Longhofer, and Kristine Latta. 1997. "Writing Amish Culture into Genes: Biological Reductionism in a Study of Manic Depression." *Culture, Medicine and Psychiatry* **21** (2): 137–159.

Fortier, Anne-Marie. 2012. "Genetic Indigenisation in 'the People of the British Isles.'" *Science as Culture* **21** (2): 153–75.

Fraser, Angus. 1992. *The Gypsies*. Oxford, UK: Blackwell.

Fujimura, Joan H., Deborah A. Bolnick, Ramya Rajagopalan, Jay S. Kaufman, Richard C. Lewontin, Troy Duster, Pilar Ossorio, and Jonathan Marks. 2014. "Clines without Classes: How to Make Sense of Human Variation." *Sociological Theory* **32** (3): 208–27.

Fujimura, Joan H., and Ramya Rajagopalan. 2011. "Different Differences: The Use of 'Genetic Ancestry' versus Race in Biomedical Human Genetic Research." *Social Studies of Science* **41** (1): 5–30.

Fullwiley, Duana. 2015. "Race, Genes, Power." *The British Journal of Sociology* **66** (1): 36–45.

Fullwiley, Duana. 2008. "The Biologistical Construction of Race." *Social Studies of Science* **38** (5): 695–735.

Gannett, Lisa. 2014. "Biogeographical Ancestry and Race." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **47** (September): 173–84.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Gannett, Lisa. 2003. "Making Populations: Bounding Genes in Space and in Time." *Philosophy of Science* **70** (5): 989–1001.

Gómez-Carballa, Alberto, Jacobo Pardo-Seco, Laura Fachal, Ana Vega, Miriam Cebey, Nazareth Martinón-Torres, Federico Martinón-Torres, and Antonio Salas. 2013. "Indian Signatures in the Westernmost Edge of the European Romani Diaspora: New Insight from Mitogenomes." *PLoS ONE* **8** (10): e75397.

Gresham, David, Bharti Morar, Peter A. Underhill, Giuseppe Passarino, Alice A. Lin, Cheryl Wise, Dora Angelicheva, et al. 2001. "Origins and Divergence of the Roma (Gypsies)." *The American Journal of Human Genetics* **69** (6): 1314–31.

Hacking, Ian. [1986]1999. "Making Up People." In *Reconstructing Individualism: Autonomy, Individuality, and the Self in Western Thought*, edited by Heller, Thomas, Morton Sosna and David Wellbery, 222-36. Stanford: Stanford University Press.

Hinterberger, Amy. 2012. "Investing in Life, Investing in Difference: Nations, Populations and Genomes." *Theory, Culture & Society* **29** (3): 72–93.

Jobling, Mark A. 2014. *Human Evolutionary Genetics*. New York: Garland Science.

Jonuz, Elizabeta. 2009. *Stigma Ethnizität: Wie Zugewanderte Romafamilien Der Ethnisierungsfalle Begegnen*. Opladen: Budrich Unipress.

Kalanin, Jan, Yutaka Takarada, Shohei Kagawa, Keiko Yamashita, Norimitsu Ohtsuka, and Akira Matsuoka. 1994. "Gypsy Phenylketonuria: A Point Mutation of the Phenylalanine Hydroxylase Gene in Gypsy Families from Slovakia." *American Journal of Medical Genetics* **49** (2): 235–39.

Kalaydjieva, Luba, Bharti Morar, Raphaelle Chaix, and Hua Tang. 2005. "A Newly Discovered Founder Population: The Roma/Gypsies." *Bioessays* **27** (10): 1084–94.

Kalaydjieva, Luba, David Gresham, and Francesc Calafell. 2001. "Genetic Studies of the Roma (Gypsies): A Review." *BMC Medical Genetics* **2** (1): 2–5.

Kaneva, Radka. et al. 2003. "A Genome-wide Linkage Scan of Bipolar Disorder in Three Extended Gypsy Families (P164)." In *Abstracts for the XIth World Congress of Psychiatric Genetics Quebec City Convention Centre October 4-8*, edited by Nicholas Barden, Marleine Cote´, Lynn DeLisi, Michael Gill, John Kelsoe, and Martin Schalling. *American Journal of Medical Genetics Part B* 112 (1): 105.

Kaneva, Radka, Vihra Milanova, Dora Angelicheva, Stuart MacGregor, Christian Kostov, Rositza Vladimirova, Spiridon Aleksiev, et al. 2008. "Bipolar Disorder in the Bulgarian Gypsies: Genetic Heterogeneity in a Young Founder Population." *American Journal of Medical Genetics Part B: Neuropsychiatric Genetics* **150B** (2): 191–201.

Kent, Michael, Vivette García-Deister, Carlos López-Beltrán, Ricardo Ventura Santos, Ernesto Schwartz-Marín, and Peter Wade. 2015. "Building the Genomic Nation: 'Homo Brasilis' and the 'Genoma Mexicano' in Comparative Cultural Perspective." *Social Studies of Science* **45** (6): 839–61.

Koenig, Barbara A, Sandra Soo-Jin Lee, and Sarah S. Richardson. 2008. *Revisiting Race in a Genomic Age*. New Brunswick, N.J.: Rutgers University Press.

Kósa, Zsigmond, Ágota Moravcsik-Kornyicki, Judit Diószegi, Bayard Roberts, Zoltán Szabó, János Sándor, and Róza Ádány. 2015. "Prevalence of Metabolic Syndrome among Roma: A Comparative Health Examination Survey in Hungary." *European Journal of Public Health* **25** (2): 299–304.

Kowal, Emma, and Joanna Radin. 2015. "Indigenous Biospecimen Collections and the Cryopolitics of Frozen Life." *Journal of Sociology* **51** (1): 63–80.

Kowal, Emma, Joanna Radin, and Jenny Reardon. 2013. "Indigenous Body Parts, Mutating Temporalities, and the Half-Lives of Postcolonial Technoscience." *Social Studies of Science* **43** (4): 465–83.

Kovats, Martin. 2013. "Integration and the Politicisation of Roma Identity." In From Victimhood to Citizenship: The Path of Roma Integration, edited by Will Guy, 260–342. Budapest: Kossuth Kiadó.

Kóczé, Angéla. 2018. "Race, Migration and Neoliberalism: Distorted Notions of Romani Migration in European Public Discourses." *Social Identities* **24** (4): 459–73.

Kyllingstad, Jon Røyne. 2012. "Norwegian Physical Anthropology and the Idea of a Nordic Master Race." *Current Anthropology* **53** (S5): S46–56.

Ladányi, János, and Iván Szelényi. 2001. "The Social Construction of Roma Ethnicity in Bulgaria, Romania and Hungary during Market Transition." *Review of Sociology* **7** (2): 79–89.

Law, Ian, and Martin Kovats. 2018. *Rethinking Roma: Identities, Politicisation and New Agendas*. Basingstoke, Hampshire: Palgrave Macmillan.

Lee, Sandra Soo-Jin, Joanna Mountain, and Barbara A. Koenig. 2001. "The Meanings of Race in the New Genomics: Implications for Health Disparities Research." *Yale Journal of Health, Policy, Law & Ethics* **1** (1): 33–75.

Lindee, Susan. 2005. *Moments of Truth in Genetic Medicine*. Baltimore: Johns Hopkins University Press.

Lipphardt, Veronika, Gudrun A. Rappold and Mihai Surdu. under review, 2022. "Ethical Standards in Forensic Genetic Research - a Critical Appraisal of Roma Studies." Forensic Science International: Genetics.

Lipphardt, Veronika. 2019. "Über den allzu sorglosen Umgang mit population labels und sampling schemes." *NTM Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin* **27** (2): 167–77.

Lipphardt, Veronika. 2016. " The Body as a Substrate of Differentiation: Shifting the Focus from Race Science to Life Scientists' Research on Human Variation." *Varia Historia* **33** (61): 109–33.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Lipphardt, Veronika. 2014. "'Geographical Distribution Patterns of Various Genes': Genetic Studies of Human Variation after 1945." *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* **47**: 50–61.

Lipphardt, Veronika, and Jörg Niewöhner. 2007. "Producing Difference in an Age of Biosociality. Biohistorical Narratives, Standardisation and Resistance as Translations." *Science, Technology & Innovation Studies* **3** (1): 45–65.

Lipphardt, Veronika, Mihai Surdu, Nils Ellebrecht, Peter Pfaffelhuber, Matthias Wienroth, and Gudrun A. Rappold. 2021. "Europe's Roma People are Vulnerable to Poor Practice in Genetics." *Nature* **599**: 368–371.

Lucassen, Leo. 1991. "The Power of Definition: Stigmatisation, Minoritisation and Ethnicity Illustrated by the History of Gypsies in the Netherlands." *Netherlands Journal of Social Sciences* **27**: 80–91.

Lucassen, Leo. 1997. "'Harmful Tramps': Police Professionalization and Gypsies in Germany, 1700-19451." *Crime, Histoire & Sociétés* **1** (1): 29–50.

Lucassen, Leo, Wim Willems, and Anne-Marie Cottaar.1998. *Gypsies and Other Itinerant Groups: A Socio-Historical Approach*. London: Palgrave Macmillan.

Magyari, Lili, Dalma Varszegi, Patricia Sarlos, Luca Jaromi, Bela I Melegh, Balazs Duga, Peter Kisfali, et al. 2014. "Marked Differences of Haplotype Tagging SNP Distribution, Linkage, and Haplotype Profile of IL23 Receptor Gene in Roma and Hungarian Population Samples." *Cytokine* **65** (2): 148–52.

Martínez-Cruz, Begoña, Isabel Mendizabal, Christine Harmant, Rosario de Pablo, Mihai Ioana, Dora Angelicheva, Anastasia Kouvatsi, et al. 2015. "Origins, Admixture and Founder Lineages in European Roma." *European Journal of Human Genetics* **24** (6): 937–43.

Mascalzoni, Deborah, A Cecile JW Janssens, Alison Stewart, Peter Pramstaller, Ulf Gyllensten, Igor Rudan, Cornelia M van Duijn, James F Wilson, Harry Campbell, and Ruth Mc Quillan. 2010. "Comparison of Participant Information and Informed Consent Forms of Five European Studies in Genetic Isolated Populations." *European Journal of Human Genetics* **18** (3): 296–302.

Mašindová, Ivica, Andrea Šoltýsová, Lukáš Varga, Petra Mátyás, Andrej Ficek, Miloslava Hučková, Martina Sůrová, et al. 2015. "MARVELD2 (DFNB49) Mutations in the Hearing Impaired Central European Roma Population - Prevalence, Clinical Impact and the Common Origin." *PLOS ONE* **10** (4): e0124232.

Mayall, David. 2004. *Gypsy Identities 1500-2000: From Egipcyans and Moon-Men to the Ethnic Romany*. New York: Routledge.

M'charek, Amade. 2005. *The Human Genome Diversity Project: An Ethnography of Scientific Practice*. Cambridge University Press.

M'charek, Amade, Katharina Schramm, and David Skinner. 2014. "Topologies of Race: Doing Territory, Population and Identity in Europe." *Science, Technology, & Human Values* **39** (4): 468–87.

Melegh, Bela I., Zsolt Banfai, Kinga Hadzsiev, Attila Miseta, and Bela Melegh. 2017. "Refining the South Asian Origin of the Romani People." *BMC Genetics* **18** (1): 1–13.

Mendizabal, Isabel, Oscar Lao, Urko M. Marigorta, Manfred Kayser, and David Comas. 2013. "Implications of Population History of European Romani on Genetic Susceptibility to Disease." *Human Heredity* **76** (3-4): 194–200.

Mendizabal, Isabel, Oscar Lao, Urko M. Marigorta, Andreas Wollstein, Leonor Gusmão, Vladimir Ferak, Mihai Ioana, et al. 2012. "Reconstructing the Population History of European Romani from Genome-Wide Data." *Current Biology* **22** (24): 2342–49.

Molnar, Miklos Z., Robert M. Langer, Adam Remport, Maria E. Czira, Katalin Rajczy, Kamyar Kalantar-Zadeh, Csaba P. Kovesdy, Marta Novak, Istvan Mucsi, and Laszlo Rosivall. 2012. "Roma Ethnicity and Clinical Outcomes in Kidney Transplant Recipients." *International Urology and Nephrology* **44** (3): 945–54.

Morar, Bharti, and Luba Kalaydjieva. 2008. "Roma/Gypsies: Footprints in the Genome." In *Population Genetic Research Progress*, edited by Viktor T. Koven, 229–244. Portland: Nova Biomedical Books.

Morar, Bharti, David Gresham, Dora Angelicheva, Ivailo Tournev, Rebecca Gooding, Velina Guergueltcheva, Carolin Schmidt, et al. 2004. "Mutation History of the Roma/Gypsies." *The American Journal of Human Genetics* **75** (4): 596–609.

Moreau, Yves. 2019. "Crack down on Genomic Surveillance." *Nature* **576** (7785): 36–38.

Moorjani, Priya, Nick Patterson, Po-Ru Loh, Mark Lipson, Péter Kisfali, Bela I. Melegh, Michael Bonin, et al. 2013. "Reconstructing Roma History from Genome-Wide Data." *PLoS ONE* **8** (3): e58633.

Munsterhjelm, Mark. 2014. "Beyond the Line: Violence and the Objectification of the Karitiana Indigenous People as Extreme Other in Forensic Genetics." *International Journal for the Semiotics of Law - Revue Internationale de Sémiotique Juridique* **28** (2): 289–316.

Myers, Martin. 2019. "An Inheritance of Exclusion: Roma Education, Genetics and the Turn to Biosocial Solutions." *Research in Education* **107** (1): 55–71.

Nagy, Károly, Szilvia Fiatal, János Sándor, and Róza Ádány. 2017. "Distinct Penetrance of Obesity-Associated Susceptibility Alleles in the Hungarian General and Roma Populations." *Obesity Facts* **10** (5): 444–57.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Nagy, Melinda, Lotte Henke, Jürgen Henke, Prasanta K. Chatthopadhyay, Antónia Völgyi, Andrea Zalán, Orsolya Peterman, Jarmila Bernasovská, and Horolma Pamjav. 2007. "Searching for the Origin of Romanies: Slovakian Romani, Jats of Haryana and Jat Sikhs Y-STR Data in Comparison with Different Romani Populations." *Forensic Science International* **169** (1): 19–26.

NaKo Gesundheitsstudie. n.d. https://nako.de/allgemeines/glossar/ (last accessed March 14, 2022).

Nash, Catherine. 2013. "Genome Geographies: Mapping National Ancestry and Diversity in Human Population Genetics." *Transactions of the Institute of British Geographers* **38** (2): 193–206.

Novokmet, Natalija, and Zlatko Pavčec. 2007. "Genetic Polymorphisms of 15 AmpFlSTR Identifiler Loci in Romani Population from Northwestern Croatia." *Forensic Science International* **168** (2-3): e43–46.

Novembre, John, Toby Johnson, Katarzyna Bryc, Zoltán Kutalik, Adam R. Boyko, Adam Auton, Amit Indap, et al. 2008. "Genes Mirror Geography within Europe." *Nature* **456** (7218): 98–101.

Okely, Judith. 1983. *The Traveller-Gypsies*. Cambridge University Press.

Ong, Aihwa. 2016. *Fungible Life: Experiment in the Asian City of Life*. Durham: Duke University Press.

Pablo, Rosario, Carlos Vilches, Maria E. Moreno, M. Carmen Rementería, Rosario Solís, and Miguel Kreisler. 1992. "Distribution of HLA Antigens in Spanish Gypsies: A Comparative Study." *Tissue Antigens* **40** (4):187–96.

Pálsson, Gísli. 2008. "The Rise and Fall of a Biobank: The Case of Iceland." In *Biobanks. Governance in Comparative Perspective*, edited by Herbert Gottweis and Alan Petersen, 41–56. London and New York: Routledge.

Pamjav, Horolma, Andrea, Zalán Judit, Béres, Melinda Nagy, and Yuet Meng Chang. 2011. "Genetic Structure of the Paternal Lineage of the Roma People." *American Journal of Physical Anthropology* **145** (1): 21–9.

Parson, Walther, and Lutz Roewer. 2010. "Publication of Population Data of Linearly Inherited DNA Markers in the International Journal of Legal Medicine." *International Journal of Legal Medicine* **124** (5): 505–9.

Picker, Giovanni. 2017. *Racial Cities: Governance and the Segregation of Romani People in Urban Europe*. London and New York: Routledge.

Pikó, Péter, Szilvia Fiatal, Zsigmond Kósa, János Sándor, and Róza Ádány. 2017. "Genetic Factors Exist behind the High Prevalence of Reduced High-Density Lipoprotein Cholesterol Levels in the Roma Population." *Atherosclerosis* **263** (August): 119–26.

Plášilová, Martina, Ivaylo Stoilov, Mansoor Sarfarazi, Ludovít Kádasi, Eva Feráková, and Vladimír Ferák. 1999. "Identification of a Single Ancestral CYP1B1 Mutation in Slovak Gypsies (Roms) Affected with Primary Congenital Glaucoma." *Journal of Medical Genetics* **36** (4): 290–294.

Plájás, Ildikó Z., Amade M'charek, and Huub van Baar. 2019. "Knowing 'the Roma': Visual Technologies of Sorting Populations and the Policing of Mobility in Europe." *Environment and Planning D: Society and Space* **37** (4): 589–605.

Poviliunas, Arunas. 2011. *Lithuania. Promoting Social Inclusion of Roma: A Study of National Policies*. On behalf of the European Commission DG Employment, Social Affairs and Inclusion.

Radin, Joanna and Emma Kowal. 2015. "Indigenous Blood and Ethical Regimes in the United States and Australia since the 1960s." *American Ethnologist* **42** (4): 749–65.

Rajagopalan, Ramya M., Alondra Nelson, and Joan H. Fujimura. 2017. "Race and Science in the Twenty-First Century." In *The Handbook of Science and Technology Studies. Fourth Edition*, edited by Ulrike Felt, Rayvon Fouché, Clark A. Miller and Laurel Smith-Doerr, 349–378. Cambridge: MIT Press.

Ramal, L.M., R. De Pablo, M.J. Guadix, J. Sánchez, A. Garrido, F. Garrido, J. Jiménez-Alonso, and M.A. López-Nevot. 2001. "HLA Class II Allele Distribution in the Gypsy Community of Andalusia, Southern Spain." *Tissue Antigens* **57** (2): 138–43.

Reardon, Jenny. 2017. *The Postgenomic Condition: Ethics, Justice, and Knowledge after the Genome*. Chicago: The University Of Chicago Press.

Reardon, Jenny. 2005. *Race to the Finish: Identity and Governance in an Age of Genomics*. Princeton: Princeton University Press.

Regueiro, Maria, Aleksandar Stanojevic, Shilpa Chennakrishnaiah, Luis Rivera, Tatjana Varljen, Djordje Alempijevic, Oliver Stojkovic, Tanya Simms, Tenzin Gayden, and Rene J. Herrera. 2011. "Divergent Patrilineal Signals in Three Roma Populations." *American Journal of Physical Anthropology* **144** (1): 80–91.

Rex-Kiss, Bela, Laszlo Szabó and Sandor Szabó S. 1972. "Blood Group Investigations among the Gypsy Population of Hungary. I. Examination of ABO, MN and Rh Blood Groups." *Annales immunologiae Hungaricae* **16**: 355–370.

Saiz, Maria, Maria J. Alvarez-Cubero, Juan C. Alvarez, Luis J. Martinez-Gonzalez, and Jose A. Lorente. 2014. "Action Protocols in DNA Identification of Isolated Populations." *Journal of Forensic Research* **5** (218): 2.

Salihović, Marijana Peričić, Ana Barešić, Irena Martinović Klarić, Slavena Cukrov, Lovorka Barać Lauc, and Branka Janićijević. 2011. "The Role of the Vlax Roma in Shaping the European Romani Maternal Genetic History." *American Journal of Physical Anthropology* **146** (2): 262–70.

Santos, Ricardo Ventura, Gláucia Oliveira da Silva, and Sahra Gibbon. 2014. "Pharmacogenomics, Human Genetic Diversity and the Incorporation and Rejection of Color/Race in Brazil." *BioSocieties* **10** (1): 48–69.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Schramm, Katharina, David Skinner, and Richard Rottenburg. 2012. *Identity Politics and the New Genetics: Re/Creating Categories of Difference and Belonging*. New York, Oxford: Berghahn Books.

Schwartz-Marín, Ernesto, Peter Wade, Arely Cruz-Santiago, and Roosbelinda Cárdenas. 2015. "Colombian Forensic Genetics as a Form of Public Science: The Role of Race, Nation and Common Sense in the Stabilization of DNA Populations." *Social Studies of Science* **45** (6): 862–85.

Surdu, Mihai. 2016. *Those Who Count: Expert Practices of Roma Classification*. Budapest, New York: Central European University Press.

Surdu, Mihai and Martin Kovats. 2015. "Roma Identity as an Expert-Political Construction." *Social Inclusion* **3** (5): 5.

Surdu, Mihai. 2019. "Why the 'Real' Numbers on Roma Are Fictitious: Revisiting Practices of Ethnic Quantification." *Ethnicities* **19** (3): 486–502.

Star, Susan Leigh. 1983. "Simplification in Scientific Work: An Example from Neuroscience Research." *Social Studies of Science* **13** (2): 205–28.

Stewart, Michael. 1997. *The Time of the Gypsies*. Boulder, Colo: Westview Press.

Stewart, Michael. 2013. "Roma and Gypsy 'Ethnicity' as a Subject of Anthropological Inquiry." *Annual Review of Anthropology* **42** (1): 415–32.

Tallbear, Kimberly. 2013. *Native American DNA: Tribal Belonging and the False Promise of Genetic Science*. Minneapolis: University Of Minnesota Press.

Tarkkala, Heta and Aaro Tupasela. 2018. "Shortcut to Success? Negotiating Genetic Uniqueness in Global Biomedicine." *Social Studies of Science* **48** (5): 740–61.

Tournev, Ivailo. 2016. "The Meryon Lecture at the 18th Annual Meeting of the Meryon Society Wolfson College, Oxford, UK, 12th September 2014: Neuromuscular Disorders in Roma (Gypsies) – Collaborative Studies, Epidemiology, Community-Based Carrier Testing Program and Social Activities." *Neuromuscular Disorders* 26 (1): 94–103.

Tremlett, Annabel. 2014. "Making a Difference without Creating a Difference: Super-Diversity as a New Direction for Research on Roma Minorities." *Ethnicities* **14** (6): 830–48.

Tsai, Yu-yueh. 2010. "Geneticizing Ethnicity: A Study on the 'Taiwan Bio-Bank.'" *East Asian Science, Technology and Society: An International Journal* **4** (3): 433–55.

Tupasela, Aaro. 2016. "Populations as Brands in Medical Research: Placing Genes on the Global Genetic Atlas." *BioSocieties* **12** (1): 47–65.

van Baar, Huub. 2018. "Contained Mobility and the Racialization of Poverty in Europe: The Roma at the Development–Security Nexus." *Social Identities* **24** (4): 442–58.

van Baar, Huub. 2015 "The Perpetual Mobile Machine of Forced Mobility: Europe's Roma and the Institutionalization of Rootlessness." In *Irregularization of Migration in Contemporary Europe: Deportation, Detention, Drowning*, edited by Joost De Bloois, Robin Celikates, and Yolande Jansen, 71–86. London: Rowman & Littlefield International.

Varszegi, Dalma, Balazs Duga, Bela I. Melegh, Katalin Sumegi, Peter Kisfali, Anita Maasz, and Bela Melegh. 2014. "Hodgkin Disease Therapy Induced Second Malignancy Susceptibility 6q21 Functional Variants in Roma and Hungarian Population Samples." *Pathology & Oncology Research* **20** (3): 529–33.

Vincze, Enikő. 2019. "Ghettoization: The Production of Marginal Spaces of Housing and the Reproduction of Racialized Labour." In *Racialized Labour in Romania*, edited by Vincze, Enikő, Norbert Petrovici, Cristina Raţ, and Giovanni Picker, 63–95. Cham Springer International Publishing.

Vermeersch, Peter. 2005. "Marginality, Advocacy, and the Ambiguities of Multiculturalism: Notes on Romani Activism in Central Europe." *Identities: Global Studies in Culture and Power* **12**: 451–478.

Wade, Peter. 2017. "Liberalism and Its Contradictions: Democracy and Hierarchy in Mestizaje and Genomics in Latin America." *Latin American Research Review* **52** (4): 623–38.

Wade, Peter, Carlos Lopez-Beltran, Eduardo Restrepo, and Ricardo Ventura Santos. 2014 *Mestizo Genomics: Race Mixture, Nation, and Science in Latin America*. Durham and London: Duke University Press.

Willems, Wim. 1997. *In Search of the True Gypsy: From Enlightenment to Final Solution*. London: Frank Cass.

Yıldız, Can, and Nicholas De Genova. 2017. "Un/Free Mobility: Roma Migrants in the European Union." *Social Identities* **24** (4): 425–41.

Veronika Lipphardt is professor for Science and Technology Studies at University College Freiburg. She is a trained biologist and historian and holds a PhD in History of Science. Her research centers on the history and the social life of DNA, especially in population genetics and forensic genetics. Together with colleagues from a wide range of disciplines, she has founded the research initiative "WIE-DNA" and has published interdisciplinary co-authored texts in journals from the life sciences as well as the social sciences and the humanities.

Gudrun A. Rappold is professor of human genetics and director of the Department of Molecular Human genetics at the University of Heidelberg. She has a joint affiliation with the faculty of medicine and biosciences. Her research has identified and functionally characterized more than 30 novel genes underlying genetic developmental disorders. She has published 300 original articles, reviews, book chapters, and one book as lead author.

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press

Dr. Mihai Surdu is a sociologist, visiting researcher at Freiburg University. His recent research focuses on genetics and society. Previously, he critically addressed the politics of knowledge production about Roma with a focus on data collection procedures, processes of stigmatization and minoritization, and social consequences resulting from ethnic categorization in various fields. His research has been supported by various organizations: Deutsche Forschungsgemeinschaft (DFG), Freiburg Institute of Advanced Studies (FRIAS), the Institute of Advanced Study at Central European University, the Max Planck Institute for the History of Science and Open Society Foundations.

Cite this article: Lipphardt, Veronika, Gudrun A. Rappold, and Mihai Surdu. 2021. “Representing Vulnerable Populations in Genetic Studies: The Case of the Roma,” *Science in Context* **34**:69–100. doi:10.1017/S0269889722000023

https://doi.org/10.1017/S0269889722000023 Published online by Cambridge University Press