Forensic DNA phenotyping in Europe: views "on the ground" from those who have a professional stake in the technology

Gabrielle Samuel & Barbara Prainsack

Published online: 29 Nov 2018.

Submit your article to this journal

View Crossmark data

Routledge
Taylor & Francis Group

Check for updates

Forensic DNA phenotyping in Europe: views "on the ground" from those who have a professional stake in the technology

Gabrielle Samuel [a] and Barbara Prainsack[a,b]*

aDepartment of Global Health & Social Medicine, King's College London, London, UK;
bDepartment of Political Science, University of Vienna, Vienna, Austria

Forensic DNA phenotyping (FDP) is an emerging technology that seeks to make probabilistic inferences regarding a person's observable characteristics ("phenotype") from DNA. The aim is to aid criminal investigations by helping to identify unknown suspected perpetrators, or to help with non-criminal missing persons cases. Here we provide results from the analysis of 36 interviews with those who have a professional stake in FDP, including forensic scientists, police officers, lawyers, government agencies and social scientists. Located in eight EU countries, these individuals were asked for their views on the benefits and problems associated with the prospective use of FDP. While all interviewees distinguished between those phenotypic tests perceived to either raise ethical, social or political concerns from those tests viewed as less ethically and socially problematic, there was wide variation regarding the criteria they used to make this distinction. We discuss the implications of this in terms of responsible technology development.

Keywords: Forensic DNA phenotyping; ethics; forensics; genetic testing; European Union (EU); ancestry testing; regulation

Introduction

Forensic DNA phenotyping (FDP) is an emerging technology seeking to make probabilistic inferences about an individual's "phenotype" (their observable characteristics) from their anonymous DNA sample. The aim is to aid criminal investigations by helping identify unknown perpetrators on the basis of the analysis of DNA from crime scene traces presumed to be deposited by the perpetrator, and to help with open missing persons cases by analysing the DNA of unidentified human remains for identification purposes.

Genetic data from FDP tests cannot "predict" the external characteristics of a person with certainty. It can, however, be used to infer an individual's phenotypic

---

*Corresponding author. Email: barbara.prainsack@univie.ac.at, barbara.prainsack@kcl.ac.uk

characteristics within a certain degree of probabilistic likelihood. Phenotypic traits are multifactorial, meaning that they are not determined by one gene but by a complex interplay between many genetic markers as well as the environment. At present, some traits such as appearance (also referred to as externally visible characteristics [EVCs], for example, hair, skin and eye color), biological age (as predicted by epigenetic testing[1]) and bio-geographical ancestry (the estimation of the geographical origin of a person's genetic ancestors at the continental level – herein referred to as ancestry) can be inferred, in certain circumstances, with high enough probabilities to make them useful in the criminal justice system (Kayser 2015).

Across the EU, FDP is currently only explicitly regulated in the Netherlands and Slovakia. In the Netherlands, testing for sex, biogeographical ancestry ("race") and hair and eye color is permitted and practiced; in Slovakia, testing for "visible phenotypic traits" is allowed (Samuel and Prainsack 2018). In all other EU countries, regulatory frameworks for, and practices relating to, FDP are complicated by either implicit or absent legislation. An important reason for this is that most EU countries regulated forensic DNA technologies in the 1990s, when FDP was not yet known.[2] In brief, in some countries where explicit regulation is lacking, experts and practitioners interpret this as permission to practice FDP in the criminal justice system, and FDP is practiced in these countries to varying degrees (e.g. Spain, UK). In other countries, current legislation is interpreted by experts as implicitly forbidding FDP (e.g. Germany, Belgium, Austria). On-going policy discussions regarding the regulation of FDP are currently underway in Germany and Switzerland.[3]

FDP poses a range of social and ethical issues. We note three concerns here that are most prominently discussed in the literature. First, given that FDP tests are probabilistic, concerns have been raised about the nature of information FDP can provide, and the possibility of the predictive nature of the information being misunderstood (Cino 2017; Enserink 2011; Sankar 2012; Seo *et al*. 2017; Toom *et al*. 2016). These concerns are compounded by expectation-generating for-profit companies such as Parabon Nanolabs, which markets FDP as a technology capable of creating composite faces of individuals from a sample of DNA alone; often based on tests which are under developed, scientifically not documented, and un-validated (Gannon 2017; Wienroth 2018). Tests offered by these companies are already used by some EU law enforcement agencies to aid criminal investigation (personal communication). Second, a number of scholars are particularly concerned that the use of FDP may exacerbate stigmatization or discrimination already directed at minority groups within society (Koops and Schellekens 2008), particularly with relation to biogeographical ancestry inferences. Here, whilst many FDP scientists stress that biogeographical ancestry is not the same as race, and pertains to ancestral continental geographical origin rather than the specific physical appearance of individuals (Kayser and Schneider 2012, e18), others scholars emphasize that a test result stating a person "is likely of African

origin" may be translated by non-specialists into the social language of identity or race, such as "African American" or "black". In this way, ancestry information pulled from a DNA profile might lead to law enforcement making decisions based on predisposed expectations about the link between ancestry and racial/ social identity (Lipphardt *et al*. 2017; Matheson 2016; M'charek 2013; Ossorio 2006; Sankar 2012). Finally, concerns have been raised that FDP may infringe privacy protections in certain instances. This could occur in a number of situations. First, in instances in which an individual has chosen to change her "natural" appearance, such as via instances of hair dying, plastic surgery and/or the use of contact lenses (Ossorio 2006). Second, if a test suggests an individual has a specific ancestry, and this does not match their appearance or narrative identity. Finally, if a test provides information that can also be correlated with a medical condition (for example the link between red hair and increased risk of melanomas) (Haga 2006). In each of these instances, these people's privacy may be infringed if police counter the appearance or narrative of the person with genetic information that tells another story.

Given such concerns, there have been appeals for the development, implementation and governance of FDP to be approached in an ethically and socially responsible fashion (Murphy 2013) – much like the broader calls for responsible research and innovation more generally (Stilgoe, Owen, and Macnaghten 2013). The aim of this paper is to contribute to this debate by providing empirical data about how people with professional stakes in phenotypic testing perceive the ethical, social and regulatory issues related to this field. Our analysis focuses on phenotypic testing for age, appearance and ancestry: our paper stems from VISAGE ("Visible Attributes Through Genomics") – a large EU-funded academic consortium project which aims to develop, validate and implement a set of prototype tools (the "VISAGE" toolkit) to allow for probabilistic inferences about age, ancestry and appearance from DNA.[4] We, the authors of the paper, form the work package of VISAGE that explores the social, ethical, and regulatory aspects of FDP, and is tasked with developing recommendations as to where and when FDP for age, biogeographical ancestry and appearance could be implemented in a socially, ethically and politically responsible manner in eight VISAGE represented EU member states, if at all. This paper contributes to this work.

We conducted 36 interviews in eight European countries[5] with members of criminal justice organizations, members of the police and representatives of government agencies, as well as with scientists from STEM (science, technology, engineering, and maths) subjects, social scientists, and academic lawyers who either worked in the broader forensic genetic technologies arena or whose work touched specifically on phenotypic DNA-based testing in the criminal justice context (in terms of the scientific development, regulatory questions, etc.) Interviewees included members of the VISAGE team.

As we conducted our interviews, it became clear that our respondents did not view FDP as a technology with clear boundaries that "raised" ethical issues.

Instead, FDP was portrayed as a heterogenous net of practices and material technologies that were partly *shaped by* ethical considerations. By this we mean that our interviewees subsumed only those technologies and practices under the label of FDP that did not, in their view, conflict with or detract from what they considered responsible science and technology. Whilst the latter was a view that virtually all of our interviewees articulated in one way or another, our respondents differed in where they drew the boundary between what they considered ethically challenging and what seemed unproblematic to them. These differences, by and large, were unrelated to the interviewees' professional identities, though we observed a trend for police officers being more inclusive regarding what they saw as ethically acceptable than other professional groups.

Before we discuss these findings in greater detail, and explore how they can contribute to discussions relating to the responsible regulation of FDP, we will clarify how the relationship between "traditional" forensic DNA profiling and FDP is commonly described in the literature. We will also introduce the notion of ethical boundary work, which has been helpful in conceptualizing our findings. The use of ethical boundary work has previously been applied to other areas of forensic DNA technology analysis to understand how forensic practitioners set certain boundaries around their work (Machado and Granja 2018).[6]

Background

The novelty of FDP compared to traditional forensic DNA profiling

In the scientific literature, FDP is portrayed to sit in contrast to standard forensic DNA profiling techniques. "Traditional" forensic DNA is commonly seen to aim to identify an individual on the basis of DNA that says nothing about the personal characteristics of that individual. It does this by analysing the profile (pattern) of a specific set of an individual's "non-coding" genetic markers, typically so-called short tandem repeats (STRs). STRs are markers located in a region of (non-coding) DNA which is argued to provide no information about a person's personal traits. FDP, in contrast, aims to analyze a sample of an unknown individual's DNA to *purposely* find out information about that person's observable characteristics. Genetic markers for FDP are therefore often located in the coding region of DNA (i.e. in a region of DNA which "codes" for a protein, and which therefore provides information about an observable characteristic). Moreover, the main utility (actual or envisaged) of the two technologies lies elsewhere: whilst STR-based profiling, besides its use in excluding or including suspects from or in an investigation, can be used as evidence in the courtroom in an attempt to convict a suspected perpetrator, FDP's intended use is to help find an unknown suspect during police investigations. Once a suspect is found with the help of FDP, it is envisaged that a "traditional" STR-based profile derived from this suspect would be compared with a crime scene profile and thus enable investigators to ascertain whether the two profiles match. FDP is also viewed as

potentially useful to help identify unidentified human remains in missing persons cases (Kayser 2015).

Changing ethical boundaries in forensic DNA technologies

The regulation of forensic DNA technologies in the criminal justice system has historically relied on the idea that DNA can be divided into coding and non-coding regions, and on the assumption that because only the sections of DNA within the coding region can provide information about an individual's observable characteristics (with the non-coding region considered as "junk" DNA), it is less ethically problematic to conduct forensic DNA testing/profiling using non-coding genetic markers (Benecke 2002; Kayser and de Knijff 2011). Despite the many issues associated with the coding/non-coding distinction, it has served a useful ethical boundary for regulators and practitioners between what is considered ethically acceptable and not acceptable for forensic DNA testing and, in some jurisdictions, is even written into forensic DNA legislation (Samuel and Prainsack 2018).[7]

The coding/non-coding ethical boundary is redundant for FDP, which uses genetic markers typically located in the coding region. As regulators and practitioners in the criminal justice system try and identify new ethical boundaries for FDP, they also need to justify why analysing coding regions of DNA for forensic purposes, which has historically been viewed as ethically problematic, is now acceptable (see Addison 2017 for an analogous situation in a different field). Such discussions are already emerging in the literature. Here, some scholars state that FDP testing should only be allowed for those tests where there is a very high degree of probability with which the phenotypic trait can be predicted, and conclusive meaningful results drawn, meaning that associations between genetic markers and predicted traits would need to be strong, and markers would need to have been validated (Murphy 2013; Smith and Urbas 2012). These same scholars, alongside others, have suggested that visible traits that are apparent to the naked eye are less "sensitive" to test for so that the visibility criterion could also be used as an "ethical boundary" (Murphy 2013; Smith and Urbas 2012). Seo *et al.* (2017) states, for example, that FDP should be used "only for predicting features that are as perceptible as to human eyewitnesses by using only the markers that are completely free of any ethical disputes" (Seo *et al.* 2017, 29). Similarly, Kayser and colleagues refer to FDP as a "biological eye witness" (Kayser 2015). Proponents of this view have argued that this demarcation would bypass most, though not all, privacy issues. Even if we accepted that FDP should be considered equivalent to an eyewitness statement, however, this would not make FDP ethically unproblematic, because eyewitness evidence is never an isolated statement about an objective fact but has its own associated ethical issues (Toom *et al.* 2016). Human perception is inevitably shaped by people's previous experiences, moral, social and political reference points, and other conscious or unconscious biases. An eyewitness' memory of a light-skinned person is not an objective statement

on skin tone on a standardized scale, but a statement of where the skin tone of the other person stands in relation to the eyewitness's own perception (which is in turn dependent on her personal situation including her human environment). For an eyewitness who grew up in Sub-Saharan Africa, light-skinnedness is likely to have a different meaning and a different form than if the eyewitness who grew up in Northern Europe. Similarly, while variations in the facial features of people might be perceptible to some people, they are not perceptible to others. In sum, Seo *et al*. (2017) solution that FDP should only be used to predict features that "are as perceptible as to human eyewitnesses" is problematic in the sense that it is unclear against what criteria such "perceptibility" should be assessed.

Ethical boundary work

The notion of drawing "ethical boundaries" is not limited to the realms of forensic science. Expanding on Gieryn's concept of boundary-work (Gieryn 1983), "ethical boundary work" was first articulated as a concept by Wainwright and colleagues in their 2006 paper exploring biomedical scientists working in the ethically controversial area of human embryonic stem cells (Wainwright *et al*. 2006). Since then it has been applied by many other scholars in their ethical explorations of innovative biotechnologies (Ehrich, Williams, and Farsides 2008; Frith, Jacoby, and Gabbay 2011; Hobson-West 2012; Holmberg 2010; Machado and Granja 2018; Stephens 2013). Wainwright and colleagues observed scientists drawing boundaries between their ethically legitimate research activity and other less ethically scientific activity (Wainwright *et al*. 2006). As such, the authors argued that "ethics has become another line of demarcation" drawn by scientists around their scientific research (p745). The authors argue that this process of social demarcation "simultaneously serves to define and defend the work of scientists involved in ethically sensitive science" (p745). In terms of the latter, by defending their work as ethically uncontroversial, located in a "positive ethical space" (compared to less ethically acceptable research), the authors argue that scientists can maintain the positive image of science, which in turn, permits them to continue to conduct their scientific research. These boundaries are thus socially constructed.

   As we show below, our findings suggest that it was not just scientists who drew ethical boundaries around FDP, but also non-scientist FDP experts. Whilst little relationship emerged between interviewees' professional identities and their views on where ethical boundaries should be drawn, there was limited evidence that professional context meant that at least some police officers have "wider" ethical boundaries around FDP than other experts. In the discussion we note the implications of these findings in terms of responsible research and technology development. We note here that whilst the ethical boundaries drawn by our interviewees are socially constructed and contextualized, they still have a role to play in terms of contributing to responsible research and

technology development, since this should consider and weigh up all stake-holders' views – not independently – but with an understanding of their social and professional context.

Methods

Recruitment

Members of the VISAGE consortium were invited, via email, to speak to us about their views on the scientific, ethical, social and regulatory issues related to FDP. Using a snowballing method, experts from within the criminal justice system, science and academia from the UK, Austria, France, Germany, Poland, Spain, Sweden and the Netherlands were also contacted via email requesting their partici-pation. Our aim was to interview at least one scientist, one police representative and one governmental agency representative from each of the countries represented. We were unable to interview a police representative in Germany, Poland, Austria or Spain (in Germany and Austria FDP police do not use FDP). Table 1 displays the number of participants we interviewed from each profession. To protect partici-pants' privacy, we are unable to break numbers down further into the respective countries.

Interviews

Thirty-two interviews were conducted in 2017 by first author, GS, via face-to-face, skype or telephone. Interviews were conducted in English, digitally-recorded and lasted between 30 min and 1 h 20 min. Four interviewees requested that the inter-view be conducted in written format due to language barriers. The interview sche-dule included questions relating to interviewees' definitions of FDP; their knowledge and opinions about FDP regulation in their respective countries, includ-ing perceptions about benefits and drawbacks of different forms of regulation; and interviewees' concerns about the FDP technology, with particular focus on ethical and societal issues. A thematic analysis of interview data was carried out using an inductive methodology. Initially, both first and second author read the transcripts and developed broad categories from the data. These findings were discussed to ensure similar ideas and categories of interest had been identified, which they

Table 1.   Number of participants interviewed from different professions.

Profession	Number of participants
STEM scientist (geneticist, statistician etc)	14
Social scientist	5
Government agency representative	8
Police representative	6
Academic lawyer	3

had. The first author, GS, then conducted detailed coding of the interviews line-by-line using NVivo software to develop more intricate themes.

All interviewees were given pseudonyms so as to anonymize the transcripts. Due to the small field of FDP, and the need to protect confidentiality, we do not provide information about individual interviewees' country of residence or their job description. Where appropriate, in the body of the findings we have noted any relevant aggregative information about themes discussed and their relationship to specific job descriptions.

Findings

The heterogeneity of FDP

Our interviewees saw phenotypic testing as a heterogeneous cluster of practices and technologies. They included a range of different tests under the label, as well as various different research methods, and technologies at various stages of scientific and technological development. Many of our STEM scientist interviewees in particular discussed how we are still in the early stages of the development of phenotypic testing ($n = 10/14$). One scientist explained that whilst those in the criminal justice system may wish to use phenotypic testing to aid with identifying suspected perpetrators of serious crimes, at present, the science is not yet reliable enough and there is no one "toolbox of method" that can be applied and defined as FDP:

> People talk about phenotyping but … there is simply not the reliable technique currently. And most of the scientists think that techniques need to be developed before they can be applied … so there is a state of uncertainty, I think, concerning this methodology and I think that is true for other countries as well. I think there is simply not a toolbox of method available that can be applied directly at this moment with a very high confidence (interviewee 6)

Another STEM scientist pointed to other issues with the technology, explaining how the results of tests that use statistical methods to compare the DNA sample with large DNA datasets will be dependent upon on whose DNA sequences are contained in the reference databases. This is because datasets containing only certain subsets of the population's DNA can produce false or inconclusive results – a well-established issue with genetic analysis for health-related purposes (Aicardi *et al*. 2016; Nanda *et al*. 2005; Need and Goldstein 2009). This, the interviewee explained, suggests that there is still much work to be done in terms of creating and developing suitable datasets so that clear evidence on the predictive power of FDP tests for specific traits can be obtained. Until this was achieved – and this interviewee explained that it was still very much an ongoing process – it would be difficult to standardize testing, or to demarcate a field that could reliably be called "FDP":

> Although there are publications out there on the association between DNA and … ancestry, and even a discussion that one can read on DNA and phenotypic

information … we still need to do more research to better describe the systems … .For [example] … when you read that one is able to determine or estimate the chronological age [using epigenetic testing] with the certainty of plus or minus 3 to 4 to 5 years. But that is not across the entire age scale and not in every different issue, we are mostly talking about blood and nothing else … this is very similar for eye colour, hair colour, skin colour and also for others … […] … [and if] police officers, the judges, the stakeholders in the law enforcement, and if they would ask us a question as researchers we would all have different answers because this research is not finished … (interviewee 27)

From heterogeneity towards the boundaries of validity and reliability

With FDP comprising a cluster of different tests for different traits, some using DNA markers and others using epigenetic ones, for many interviewees (all interviewees who explicitly discussed the issue, $n = 18$), the question of whether a predictive test for a specific phenotypic trait had been scientifically validated was a key factor for deciding whether that test should be considered under the label of FDP. By including only those tests that were validated and reliable under the label of FDP, our interviewees distanced FDP from tests in the wider field of phenotypic testing that they perceived as under-developed, not appropriately validated, and producing unreliable results (e.g. such as creating a visual "phantom image" of an unknown perpetrator as some commercial companies currently claim to be able to do). In other words, the concept of FDP, in the view of our interviewees, did not merely denote an attempt to infer observable traits from DNA, but it signified the technical ability to do so at a certain level of validity and reliability. This is exemplified as interviewee 8 talks about the effect of private companies using unvalidated phenotypic tests, highlighting the important relationship between reliability and validity, and the need for science to be viewed as a trustworthy endeavor: "overseas companies offer already 'complete solutions' namely DNA-based photofit pictures to police … making our discipline, the forensic genetics, untrustworthy at long sight. [This is] not good for science."

External validity – i.e. the ability of the test to accurately "predict" observable traits in a high proportion of cases in the real world rather than just the laboratory – was also seen as a necessary characteristic of FDP. For example, in the extract below, an interviewee distinguishes between a phenotypic predictive test that works in a scientific institution and one that works in practice "on the ground". Here, the interviewee is referring to fieldwork challenges such as DNA degradation and/or only having small, impure or incomplete samples of DNA to test. The interviewee notes that we cannot view research methods and tests created in a scientific laboratory as a technology (i.e. FDP) until they are validated and can function usefully in the criminal justice system in practice:

We need to do a lot more before we can use the technology, it's not only a legal thing it's also a scientific thing where we have some information from universities [about]

how things work. But that doesn't mean they work at the same practice and when working in certain environments. So what we have from the university research is that there is certain chances of prediction but this is isolated from the surrounding … . (interviewee 35)

Different conceptualisations of utility within the boundaries of validity and reliability

Whilst most STEM scientists ($n = 12/14$) explicitly pointed towards the boundary between scientifically reliable and validated tests deemed FDP on the one hand, and undeveloped probabilistic predictive tests for phenotypic traits on the other, among the police officers that we interviewed ($n = 4/6$) the line was more blurry. Whilst reliability and validity were important, they also considered it very valuable to have tests for as many phenotypic traits as possible included under the umbrella of FDP so that they could use them when trying to solve serious crimes, such as rape, murder or armed robbery:

> I would be very, very much interested … [to] … get more and more information from that specific crime stain to make the group of persons I am looking for as small as possible. So yes more I would like the age, I would like the skin colour, I would like the height of a person [(this currently can not be reliably predicted)], I would like the whole phenotype. (interviewee 11)

These findings suggest that for those who are the "front lines" of crime scene work such as police officers, and who may feel the pressure of having to solve a crime the most, it might be tempting to utilize tests even if they have not yet been validated or if they have low predictive value – just in case they could be helpful. These findings indicate a third category next to reliability and validity, namely utility (see Wienroth 2018), had bearing on where and how our interviewees drew the line around FDP. For most of the scientists in our sample, the utility of a phenotypic test was highest when the test was proven to be reliable and valid. For example, one scientist emphasized that the predictive test for male pattern baldness had no practical use because it could not predict baldness reliably enough to be valid:

> They do baldness, but, well they can't do that reliably, we do baldness but I wouldn't want to use those predictions or communicate those predictions in any way because they are not particularly useful (interviewee 20)

For those working at the coalface, expected utility in case work was the starting point. It included the ability to progress with crime scene work even if the information obtained from these tests had uncertain predictive value. These different takes on utility reflect different underlying professional views about what is "ethical" – for scientists, being "ethical" is only using validated and trustworthy tests; for police officers it is using all options available to try and catch a perpetrator. These views on utility therefore need to be situated in broader discussions about the ethically responsible implementation of FDP. It should be emphasized here that

those of our respondents who argued for an expansion of the types of tests to be used under the FDP label did *not* misunderstand or exaggerate the reliability or validity of early tests. Instead, despite their being fully aware of the low predictive value of the tests, they still thought that the utility of possibly making a breakthrough in the investigation outweighed the risk of wasting time and money on wrong leads.

Ethical boundary making

Whilst analytic validity and reliability clearly played an important role in defining FDP, interviewees commonly drew other boundaries around the phenotypic testing of traits. Most importantly, this related to whether the testing for a phenotypic trait was, or was perceived to be one that either (a) raised socially sensitive concerns, or (b) was associated with ethical issues. Phenotypic inference of traits not falling into either of these categories was more likely to be considered as FDP. This made FDP analogous to an "ethical safe space" for phenotypic testing in the criminal justice system, and this was contrasted with unethical practices that were either unvalidated or driven purely by commercial interests, such as the creation of "phantom images" from DNA. Despite the fact that all interviewees created such ethical boundaries around the FDP "ethical safe space", they came to different conclusions regarding what phenotypic tests to specifically locate inside and outside of this "ethical safe space". We discuss this below, paying particular attention to the detailed descriptions interviewees provided about their views on both socially sensitive testing, and the ethical issues associated with phenotypic testing. We note any relationships which emerged between interviewees' views and their professional background.

1.   *Socially sensitive issues: political debates and discrimination around ancestry*

A number of interviewees discussed issues related to the political and societal sensitivity of ancestry testing ($n = 11$). While a very small minority of these interviewees ($n = 2$) did not perceive this testing as particularly ethically, politically and socially sensitive: "(especially I don't worry about the problem of race and geographical ancestors … " [interviewee 13]), for most interviewees ($n = 9$) the reverse was the case (M'Charek, Toom, and Prainsack 2012; Toom 2012). This was especially apparent for interviewees located in Germany, where the historical context of the Nazi Regime heightens many people's sensitivity to risks of genetic discrimination (Sperling 2004, 2013), and where some political actors have recently linked the expanded use of DNA-testing in criminal investigation to the topic of migration in scientifically and politically problematic ways.[8] In the words of one respondent:

  I think this issue will be ancestry because we have in Germany based on our history some awareness when it comes to what ancestry – because people may think of race, and race was something that was very propagated in the thirties and forties by the

> Nazi regime. And everybody is very careful not to be compared to that area of time (interviewee 35)

Indeed, for most of these interviewees, the socially and politically sensitive nature of ancestry inference, and whether it could increase discrimination based on racial and ethnic stereotypes, was what differentiated it from other phenotypic traits, such as those for appearance ("I think ancestry has an extra political dimension which maybe appearance has in a less strong way" [interviewee 15]). This differentiation meant that they considered ancestry testing as excluded from FDP's "ethical safe space". It was rather something that needed to be "managed more carefully" than tests for other traits:

> I think ancestry for me would probably be the one that I would see would have more ethical issues. Because when you start profiling individuals I think that has to be carefully managed. [In this] current climate of operational crisis, I think everyone is on heightened alert, and I think that also you have got your historic stuff, black on black use, crime etc stop and search, so there is always that arena that is quite contentious. So I think that would for me, out of the three, be the one that needs to be managed more carefully (interviewee 34)

For other interviewees ($n = 2$), although they also saw testing for ancestry as sensitive, the relationship between probabilistic ancestry and appearance inference was inseparable. This was because these two sets of predictions were seen as tightly linked and complementing each other. For these interviewees, information about biogeographical ancestry could help to make the prediction of appearance (e.g. skin tone) more accurate, and vice versa. One interviewee explained, if appearance inference reveals that an individual is likely to have blue eyes then it is also likely saying that the individual is of European descent; and if an individual's skin color is predicted to be dark, then it also tells us something about their possible ancestry. Ancestry inference testing was therefore viewed as something that could be within the remit of FDP if it serves to improve appearance testing:

> Phenotypic traits and the ancestry … the nice thing is they help each other … it's something that supports each other in terms of the type of genetic information. And also if you have information that someone is maybe Saharan African origin and you find he has dark skin well that obviously fits together … you can't handle these traits separately (interviewee 28)

2. *Ethically problematic traits: observable versus "personal" and sensitive characteristics*

Echoing some of the literature, interviewees explained that FDP does not and should not include phenotypic tests for purely "internal" characteristics (e.g. those related to predispositions to disease, as well as other sensitive, personal information). Rather, FDP should only include the predictive testing of those phenotypic

characteristics which are deemed part of a person's appearance, for example, the EVCs of skin, hair and eye color – those which are the focus of the VISAGE project. Such traits were viewed as "not so problematic from the ethical point of view":

> I think the so-called appearance phenotype, so the ones which you can see if you are an eye witness, for example eye colour or hair colour so I think they are not so problematic from the ethical point of view because everyone can see them (interviewee 8)

This ethical boundary between EVCs and internal phenotypic traits as they relate to FDP is in fact echoed in explicit FDP regulation in some countries, including the Netherlands and Slovakia (Samuel and Prainsack 2018). It was also evident in a recent court case in France, which ruled that FDP for "morphological characteristics" (interpreted as those characteristics which can be seen at the outside of the body)[9] was permitted, but predictive phenotyping for internal or private characteristics was not:

> Basically what they say is if it's a [unknown DNA] trace then you can do phenotyping if it's a criminal investigation. And what they say in the court decision is that you are allowed to predict any external physical appearance traits. So nothing internal, nothing private, but everything that could be seen on a photograph then technically you should be able to provide it … . (interviewee 33)

Some interviewees' narratives suggested a gray area between observable characteristics and internal, personal and health traits. This gray area came in a number of forms, of which we note three: considerations of age inference, personally sensitive traits, and the blurriness of health predispositions. We briefly discuss each of these in turn below to give more of a sense of what the gray area entails.

Considerations of age inference.   Interviewees had different views regarding whether DNA-based (e.g. epigenetic) testing for age inference could reveal sensitive information. For nearly all interviewees who discussed this in detail ($n = 11/14$), this type of testing was so unlikely to reveal sensitive information that it was viewed as ethically unproblematic ("age I would consider the least problematic in terms of social impact … I don't see a real [potential for the] misuse of the data" (interviewee 27)). Interviewees explained that this was because age represented a descriptor of what can be seen by looking at a person. Indeed, one of our interviewees explicitly said that s/he considered age an EVC:

> There are the rare examples where some people look way older or way younger than they are. But most of the people when it comes to such broad age estimation in terms of age categories it is visible. So for me on that broad level age is a visible trait (interviewee 36)

A minority of our interviewees, all STEM scientists ($n = 3$) held a different view. In particular, one of these interviewees perceived age prediction as especially sensitive because of its potential to also disclose information on disease

predispositions/status, for example with relation to cancer or other medical problems. For this interviewee, age prediction tests represented a different category of tests from those commonly subsumed under the FDP label:

> We understand forensic DNA phenotyping … [as the] … prediction of appearance traits, but there are two additional aspects … the first is biogeographic ancestry and the second one is, of course, prediction of age … I'm not sure that [the] forensic DNA phenotyping term covers all these three aspects – at least I'm hesitating about age prediction because it's something completely different. Age prediction can … give some information about mortality … so it's … very sensitive [and] it's a little bit different to FDP and what we mean by FDP … […] … From the biological age we can directly know … the medical state [of the individual]. For example, let's say [a person is] 40 years old, and this is his or her chronological age, but after analysis we know that the biological age is 50 years old … [This]. means that there is something wrong with his or her … medical status … – this can be an indication of cancer, for example, or other medical problems. So it can potentially be very sensitive (interviewee 30)

Sensitive personal traits.   Interviewees had different views regarding the intimately personal nature of ancestry, as well as some EVC inference findings. These views seemed independent of their professional background, though no police officer interviewees viewed ancestry testing or EVC testing as personal or sensitive.

In terms of the personal nature of some EVC inference findings, one particular interviewee highlighted a scenario with relation to early onset baldness inference testing. This interviewee explained that whilst "baldness" may be viewed as an EVC by some interviewees, this may not be the case if a person chooses to hide their baldness by wearing a hairpiece. For this interviewee, this raised questions about where to locate an early onset baldness inference test – within or outside of the remit of the ethical safe space of FDP. As this interviewee remarked, "[some] visible traits can be sensitive – like baldness … – some individuals may be really cautious about disclosing that they are bald" (interviewee 30).

An analogous scenario emerged in terms of ancestry inference. Some interviewees believed that "ancestry testing of course can discover personal data which are more sensitive in my point of view than physical traits" (interviewee 14). As one interviewee explained:

> Ancestry [testing] is not appearance phenotyping but it's something else … perhaps you do not want your ancestry to be disclosed. So this also has to be handled more sensitively than for example, appearance phenotypes. So yes, I think it would be a good idea also to differentiate between these different types of phenotyping (interviewee 8)

In contrast, other interviewees viewed information about a person's ancestry as less about revealing personal (internal) information and more in line with a person's EVCs:

If you are doing your ancestry statement on a continental region [as with VISAGE] then if a person is of pure continental ancestry and that means both the paternal and maternal ancestors come from that region, then the person will have appearance traits that are typical for those type of continental regions. And then you could argue ancestry is visible in the same way as appearance is … […] … Of course there are couples of different continental ancestry which have offspring but most people are still of uniform continental ancestry … indeed you can have cases … where the ancestry is non visible … [and] the individual doesn't know … [and] may not want to know … [and] … also doesn't want others to know. And then you would be with this typical application of right not to know privacy example. But … [this] … applies to a minority of the population (interviewee 36)

The blurriness of health predispositions.   Nearly all interviewees emphasised that any information relating to an individual's health/a medical predisposition was intrinsically sensitive personal information, and testing for such traits should not be permitted within the remit of FDP.

Of course what is very sensitive is disease related … th[ose] areas we are not going, that's not allowed, and I think that's a very good thing (interviewee 29)

At the same time, a few of our interviewees ($n = 5$) discussed how in the process of phenotypic testing, information could be produced about an individual's health/a medical predisposition as an "incidental finding" (Haga 2006; Scudder *et al*. 2018). This would be true if the marker for the tested trait is in close proximity to a marker for a specific health pre-disposition.[10] This is because portions of DNA in close proximity are often genetically linked, that is, inherited together and so the presence of a variant of one marker can be predictive of another. In the extract below, an interviewee explains these blurred boundaries, having arisen during the development of DNA-based tests for red hair, and the difficult questions which remain for the criminal justice field in terms of "how far you can go" when using such tests in a criminal case:

I think the whole notion about pigmentation comes from the medical domain because … red hair … to a degree was an indicator of certain medical conditions. And that really needs to be considered and recognised that if you have certain visible traits, you have to set a limit there as to how far you might be able to go in further investigating (interviewee 19)

This interviewee concluded: "I wouldn't necessarily agree with enabling someone to do a test for clinically relevant traits" (interviewee 19). However, not all interviewees held this view, and here differences emerged between the views of (some) police officers and other interviewees. One police officer stated that whilst we do need to think these issues through, we should not discount the benefits of also including those traits within FDP that may provide some clinically relevant information. For this interviewee, and most other police officers, it is important to weigh the value of finding a suspected

perpetrator of a serious crime against identifying some personal information about that individual:

> If these markers are powerful and really help to lead the investigation to be accurate is it not worth it? … We need to decide what to do, do we need to be as accurate as possible or do we need to not go at all in the medical part of the prediction? And even if the marker is really slightly attributed to pathology does it really mean something on the future of the health of the guy you're trying to predict? I don't really have an answer but it's just a question of what you gain versus what you ethically allow (interviewee 33)

In this way, whilst all interviewees defined FDP as including only those predictive tests for phenotypic traits which do not primarily aim at providing information about personal, internal and/or health characteristics, the boundaries of this ethical safe space were diffuse in terms of the classification of traits that could potentially also disclose health- and disease-relevant information without aiming to do so. Preliminary evidence suggests that some of these boundaries are related to professional backgrounds.

Conclusion

We have shown that understandings of what FDP is, and what tests and practices should be subsumed under its remit, is heterogeneous and complex among practitioners and experts in the field. In particular, we have shown how whilst interviewees created boundaries between those phenotypic tests relating to age, ancestry and appearance that were perceived to be/not to be under the remit of FDP in terms of reliability and validity, and ethical and social issues/sensitivity, interviewees had different views on where to draw these boundaries (we note that the remit of FDP could be fluid and subject to change and interference).

Many previous studies have noted how forensic scientists have closed discussions about forensic DNA testing by solely relating them to the importance of reliability and validity. For our interviewees, a test for a specific phenotypic trait had to meet two criteria – be valid and reliable, and ethically unproblematic– to be contained in an "ethical safe space"; traits sitting outside the safe space were distanced from FDP. Around this ethical safe space was a gray area containing tests using markers that could, albeit aimed at testing for traits that were not considered "private", inadvertently also disclose disease-relevant or personally sensitive information. Different interviewees had different views regarding which of these tests could be encapsulated within the ethical safe space – in particular, police officer interviewees had broader notions of this, and we discuss this in more detail below.

The fact that our FDP experts consider it important to only use tests which are deemed ethically "safe" resonates with studies on practices of scientists and practitioners in other areas. Scientific researchers in other fields have also used ethical boundaries to distinguish between "settled", "no-ethics", "positive ethical space" practice and more contested practice (Frith, Jacoby, and Gabbay 2011;

Wainwright *et al.* 2006).[11] Wainwright argues that researchers distance themselves from the latter in order to maintain the image of science, and to give scientific researchers permission to continue their work (Wainwright *et al.* 2006). Perhaps this is what is happening here as our interviewees defined and defended an ethical boundary in their practices. Though our findings suggest that something else is occurring too. It was not just our scientist interviewees who drew ethical boundaries, but also other practitioners and experts in the field of FDP. And whilst ethical boundaries were used to justify the ethical permissibility of the technology and separate FDP from a contested space, the boundary-making worked the other way too: interviewees concerned about phenotypic testing used FDP as a way to distance those tests they deemed ethically problematic/ socially sensitive away from scientists' "grasp". Further research is now warranted to determine how these different modes of boundary making are emerging in the "adoption space" of FDP (Wienroth 2018); how different expectations are attached to these boundaries; and how these expectations play a role in developing communities of FDP practice (see Wienroth 2018).

On the basis of our findings, we note two main ways in which these can inform responsible regulation strategies. First, our findings corroborate the importance of the "reliability" and "validity" boundary, which has been discussed in the literature (Murphy 2013; Smith and Urbas 2012). This gives empirical strength to the need to incorporate this as a key underpinning of any governance oversight system for FDP. A regulatory framework which only permits those tests which are validated and reliable protects itself from instances in which there may be a "push" from the criminal justice system, possibly via (as some of our interviewees explained to us) the public and/or mass media, as they attempt to solve a serious, possibly highly emotive, horrific crime. This is important since our findings suggest at least some evidence that (some) police are more likely to wish to conduct less reliable or less valid tests in exchange for a perceived increased utility in terms of apprehending a perpetrator. For these police officers being ethically responsible is related to using all technologies available in the hope of identifying a lead. In these instances, we can imagine a scenario in which a person has murdered a number of people. Police officers, understandably, want as much help as possible when trying to identify the perpetrator (for example, in the form of information about what the perpetrator looks like), especially if they have no other leads. However, we argue that un-validated and/or unreliable tests may offer little in the way of actual (rather than perceived) utility in police officers' attempts to apprehend the suspected perpetrator; and possibly have the potential to provide false leads. In these instances, we would argue it is more ethically responsible to put more effort into those technologies and approaches with a better chance of yielding promising results for the police. We do not view the technology as capable of "finding answers" before it is has been sufficiently and responsibly developed to do so.

Having said all of this, questions still remain about what constitutes a "valid" and "reliable" phenotypic test, and what threshold of probabilistic accuracy must be reached to classify a test as FDP. Our findings indicate differences in views about this, at least some of which can be attributed to professional background, with police officers more willing to accept lower standards of reliability and validity in exchange for an increased perceived utility. Due to the nature of the tests, FDP will never predict traits with certainty, and we still need to explore where to draw the line with respect to these different professional views, and whether this should depend on the type of crime being investigated. Our further work continues to explore this.

Second, whilst the heterogeneity of our interviewees' opinions about which phenotypic inference tests should or should not be considered under the remit of FDP makes it difficult to prescribe recommendations about a responsible regulatory framework with relation to this, our findings are instructive in that they highlight that attempts to draw boundaries around which FDP tests are permitted in terms of ethically disputed and non-disputed tests lack nuance (see also Seo *et al*. 2017). This is not only because all tests, to a certain degree, can be perceived as ethically problematic, but also because, as we have shown, stakeholders prioritize different ethical issues as more or less valuable, and moreover, (some) different professional stakeholders have different views on the ethical utility of such tests. Such different understandings of what tests are deemed ethically problematic and/or valuable, whilst socially constructed and related to each individual's or profession's own identity, still need consideration. Indeed, the heterogeneity of views highlighted in our findings, which resonates with findings from the ethical boundary literature previously (Hobson-West 2012; Wainwright *et al*. 2006), illustrate the pluralistic and contested nature of ethical values "on the ground" between different individuals and different professional stakeholders in the criminal justice system. Similar findings have been described in other areas of science where different professional or individual stakeholders adopt different ethical "role positions" (Cribb *et al*. 2008; Samuel *et al*. 2016), and findings by Williams and Johnson (2004) show how various actors in the criminal justice system represent forensic DNA technologies differently. This raises questions about how best to weigh up such role positions/representations given their social, cultural and political contexts. We argue that all positions are important and need to be considered as having merit within their socio-cultural context. Our findings therefore provide important empirical evidence for policymakers as they develop frameworks for FDP. It reminds them to consider and incorporate this pluralism, and to carefully balance the nuances of the different values of interviewees (and professions) against each other, whilst at the same time concomitantly taking into account their different social contexts. This is the approach we took (i.e. considering ethical pluralism as well as considering different social context) when developing our recommendations to only consider valid and reliable tests under FDP even though our

interviewees varied in their beliefs about this (though we note that we are conducting further research to define how "valid" and "reliable" the tests need to be).

Finally, whilst, we conceptualized our findings within ethical boundary making, instead of using concepts of "social relationships", "rhetoric" or "discourses of power" (Kruse 2015; Vailly 2017; Williams and Johnson 2004) – the latter of which would have paid more attention to the social, political and cultural context of different discourses which have led to the emergence of FDP (Jong and M'charek 2017; Vailly 2017; Wienroth 2018) – we remind policymakers to remain vigilant of distributions, practices and institutions of power in their respective societies when making decisions about where to draw regulatory boundaries. What is required is a more considered approach with relation to this.

To finish, any responsible research and innovation program must engage both professional and public stakeholder views with regards to future policy decisions. We believe that our findings offer a useful base to now explore expert, stakeholder, and potentially also public views on FDP in more detail i.e. now that we empirically understand the nature of the different boundaries our interviewees created around phenotypic testing in the criminal justice system; as well as the questions remaining regarding how to define such boundaries (for example in terms of "reliable" and "valid"), we can use these findings to frame our discussions with the public and civil society. In essence, interpreting our findings in terms of ethical boundary making gives us the conceptual tool for this further research. Combining this further research with the findings reported in this paper will then allow us to move forwards more concretely towards proposing responsible recommendations for the regulation of FDP for age, ancestry and appearance.

Disclosure statement

No potential conflict of interest was reported by the authors.

Ethical approval

This work has received ethical approval from King's College London ethics review committee. Reference MR/16/17-551.

Notes

1. Epigenetic testing is based on the analysis of DNA methylation patterns.
2. For an in-depth discussion of the complexity of these frameworks and practices in EU countries, and any on-going policy discussions, see Samuel and Prainsack (2018).
3. http://www.visage-h2020.eu/.
4. http://www.visage-h2020.eu/.
5. The eight countries were chosen because they are represented in the VISAGE consortium.
6. See (Vailly 2017; Williams and Johnson 2004) for examples of other conceptualizations describing different views associated with forensic DNA technologies (e.g. "different representations"; "different rhetorical (power) discourses"). See (Jong and M'charek 2017; Vailly 2017; Wienroth 2018) for a deeper discussion the social, political and cultural context of different discourses which have led to the emergence of FDP. See (Kruse 2015) for a deeper discussion on the social context of the use of forensic evidence.
7. This is in spite of continuing concerns about ethical issues – in particular, related to privacy and undue discrimination – which continue to exist; and also despite of an awareness that the non-coding/coding DNA boundary is blurry, with instances existing of specific "non-coding" genetic markers being used to predict information about phenotypic traits (Kayser 2015). There are two reasons why this could be the case. First, the marker is within an area of the coding region's regulatory function (typically defined as non-coding). Second, because the marker and the respective coding region are in close proximity. Portions of DNA in close proximity are often genetically linked, that is, inherited together, and so the presence of a specific STR variant (haplotype) can be predictive of a specific observable trait (because specific sequence of DNA are inherited together). Indeed, a number of non-coding STR haplotype markers are used to predict EVCs.
8. https://www.nature.com/news/forensics-germany-considers-wider-use-of-dna-evidence-in-criminal-cases-1.21713.
9. The Court case was ruling on the permissibility of using FDP in a serial rapist case in Lyon. The investigating Magistrate had requested FDP testing for hair, eye and skin color not knowing that it was forbidden to perform such tests in France. When he tried to erase the order the status of the legality of the test was examined.
10. This is different for age markers. Here, the methylation pattern of DNA which gives information about the biological age of a person, can also be indicative for a disease-related status.
11. Limitations of using the ethical-boundary work perspective exist, including the notion that by referring to a safe space as "ethical" we are also taking part in categorizing – in this case with relation to what we as authors deem ethical and socially sensitive (Hobson-West 2012).

ORCID

Gabrielle Samuel ⓘ http://orcid.org/0000-0001-8111-2730

References

Addison, C. 2017. "Spliced: Boundary-Work and the Establishment of Human Gene Therapy." *Biosocieties* 12 (2): 257–281.

Aicardi, C., M. Damjanovicova, L. Del Savio, F. Lucivero, M. Mormina, M. Niezen, and B. Prainsack. 2016. "Could DTC Genome Testing Exacerbate Research Inequities." In *Bioethics Forum of The Hastings Center*. https://www.thehastingscenter.org/response-to-call-for-essays-could-dtc-genome-testing-exacerbate-research-inequities/.

Benecke, M. 2002. "Coding or non-Coding, That is the Question: Having Solved the Last Technical Hurdles to Extract DNA Information from Virtually any Biological Material, Forensic Biologists Now Have to Ponder the Ethical and Social Questions of Using Information from Exonic DNA." *EMBO Reports* 3 (6): 498–502.

Cino, J. G. 2017. "Tackling Technical Debt: Managing Advances in DNA Technology That Outpace the Evolution of law." *American Criminal Law Review* 54: 373–421.

Cribb, A., S. Wainwright, C. Williams, B. Farsides, and M. Michael. 2008. "Towards the Applied: The Construction of Ethical Positions in Stem Cell Translational Research." *Medicine, Health Care and Philosophy* 11 (3): 351–361.

Ehrich, K., C. Williams, and B. Farsides. 2008. "The Embryo as Moral Work Object: PGD/IVF Staff Views and Experiences." *Sociology of Health & Illness* 30 (5): 772–787.

Enserink, M. 2011. "Can This DNA Sleuth Help Catch Criminals?" *Science* 331 (6019): 838–840. doi:10.1126/science.331.6019.838.

Frith, L., A. Jacoby, and M. Gabbay. 2011. "Ethical Boundary-Work in the Infertility Clinic." *Sociology of Health & Illness* 33 (4): 570–585. doi:10.1111/j.1467-9566.2010.01308.x.

Gannon, M. 2017. "Amazing DNA Tool Gives Cops a New Way to Crack Cold Cases: DNA Phenotyping Can Produce a Sketch of the Suspect. But is It Ready for Primetime?" *NBCnews*. https://http://www.nbcnews.com/mach/science/amazing-dna-tool-gives-cops-new-way-crack-cold-cases-ncna781946.

Gieryn, T. F. 1983. "Boundary-Work and the Demarcation of Science from Non-Science: Strains and Interests in Professional Ideologies of Scientists." *American Sociological Review* 48 (6): 781–795.

Haga, S. B. 2006. "Policy Implications of Defining Race and More by Genome Profiling." *Genomics, Society and Policy* 2 (1): 57–71.

Hobson-West, P. 2012. "Ethical Boundary-Work in the Animal Research Laboratory." *Sociology* 46 (4): 649–663.

Holmberg, T. 2010. "Tail Tales: How Researchers Handle Transgenic Dilemmas." *New Genetics and Society* 29 (1): 37–54.

Jong, L., and A. M'charek. 2017. "The High-Profile Case as 'Fire Object': Following Hte Marianne Vaatstra Murder Case through the Media." *Crime Media Culture* 4: 347–363. doi:10.1177/1741659017718036.

Kayser, M. 2015. "Forensic DNA Phenotyping: Predicting Human Appearance from Crime Scene Material for Investigative Purposes." *Forensic Science International: Genetics* 18: 33–48. doi:10.1016/j.fsigen.2015.02.003.

Kayser, M., and P. de Knijff. 2011. "Improving Human Forensics through Advances in Genetics, Genomics and Molecular Biology." *Nature Reviews Genetics* 12 (3): 179–192. doi:10.1038/nrg2952.

Kayser, M., and P. M. Schneider. 2012. "Reply to 'Bracketing off Population Does Not Advance Ethical Reflection on EVCs: A Reply to Kayser and Schneider' by A. M'charek, V. Toom, and B. Prainsack." *Forensic Science International: Genetics* 6: E18–E19.

Koops, B.-J., and M. Schellekens. 2008. "Forensic DNA Phenotyping: Regulatory Issues." *Science and Technology Law Review* IX: 158–202.

Kruse, C. 2015. *The Social Life of Forensic Evidence*. Oakland, CA: University of California Press.

Lipphardt, V., V. Toom, A.-C. Mupepele, and T. Lemke. 2017. "Open Letter on Critical Approaches to Forensic DNA Phenotyping and Biogeographical Ancestry." http://www.fb03.uni-frankfurt.de/67048840/Lipphardt-et-al-2017—OPEN-LETTER-ON-CRITICAL-APPROACHES-TO-FORENSIC-DNA-PHENOTYPING-AND-BIOGEOGRAPHICAL-ANCESTRY.pdf.

Machado, H., and R. Granja. 2018. "Ethics in Transnational Forensic DNA Data Exchange in the EU: Constructing Boundaries and Managing Controversies." *Science as Culture* 27: 242–264. doi: https://doi.org/10.1080/09505431.2018.1425385.

Matheson, S. 2016. "DNA Phenotyping: Snapshot of a Criminal." *Cell* 166 (5): 1061–1064. doi:10.1016/j.cell.2016.08.016.

M'charek, A. 2013. "Beyond Fact or Fiction: On the Materiality of Race in Practice." *Cultural Anthropology* 28 (3): 420–442.

M'Charek, A., V. Toom, and B. Prainsack. 2012. "Bracketing off Population Does Not Advance Ethical Reflection on EVCs: a Reply to Kayser and Schneider." *Forensic Science International: Genetics* 6 (1): e16–e17; author reply e18-9. doi:10.1016/j.fsigen.2010.12.012.

Murphy, E. 2013. "Legal and Ethical Issues in Forensic DNA Phenotyping." *New York University Public Law and Legal Theory Working Papers* Paper 415. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2288204.

Nanda, R., L. P. Schumm, S. Cummings, J. D. Fackenthal, L. Sveen, F. Ademuyiwa, M. Cobleigh, et al. 2005. "Genetic Testing in an Ethnically Diverse Cohort of High-Risk Women: a Comparative Analysis of BRCA1 and BRCA2 Mutations in American Families of European and African Ancestry." *JAMA* 294 (15): 1925–1933.

Need, A. C., and D. B. Goldstein. 2009. "Next Generation Disparities in Human Genomics: Concerns and Remedies." *Trends in Genetics* 25 (11): 489–494.

Ossorio, P. N. 2006. "About Face: Forensic Genetic Testing for Race and Visible Traits." *The Journal of Law, Medicine & Ethics* 34 (2): 277–292. doi:10.1111/j.1748-720X.2006.00033.x.

Samuel, G., A. Cribb, J. Owens, and C. Williams. 2016. "Relative Values: Perspectives on a Neuroimaging Technology From Above and Within the Ethical Landscape." *Journal of Bioethical Inquiry* 13 (3): 407–418.

Samuel, G., and B. Prainsack. 2018. "The Regulatory Landscape of Forensic DNA Phenotyping in Europe." *VISAGE. Forthcoming – will be.* http://www.visage-h2020.eu/.

Sankar, P. 2012. "Forensic DNA Phenotyping: Continuity and Change in the History of Race, Genetics, and Policiing." In *Genetics and the Unsettled Past: The Collision of DNA*, edited by K. Wailoo, A. Nelson, and C. Lee, 104–113. Piscataway: Rutgers University Press.

Scudder, N., D. McNevin, S. F. Kelty, S. J. Walsh, and J. Robertson. 2018. "Forensic DNA Phenotyping: Developing a Model Privacy Impact Assessment." *Forensic Science International: Genetics* 34: 222–230. doi:10.1016/j.fsigen.2018.03.005.

Seo, H. J., S. Cho, J. H. Lee, S. H. Lyoo, M.-Y. Kim, and S. D. Lee. 2017. "Forensic DNA Phenotyping: A Review in Korean Perspective." *Korean Journal of Legal Medicine* 41: 23–31.

Smith, M., and G. F. Urbas. 2012. "Regulating New Forms of Forensic DNA Profiling Under Australian Legislation: Familial Matching and DNA Phenotyping." *Australian Journal of Forensic Sciences* 44 (1): 63–81.

Sperling, S. 2004. "Managing Potential Selves: Stem Cells, Immigrants, and German Identity." *Science and Public Policy* 31 (2): 139–149.

Sperling, S. 2013. *Reasons of Conscience: The Bioethics Debate in Germany.* Chicago: University of Chicago Press.

Stephens, N. 2013. "Growing Meat in Laboratories: The Promise, Ontology, and Ethical Boundary-Work of Using Muscle Cells to Make Food." *Configurations* 21 (2): 159–181.

Stilgoe, J., R. Owen, and P. Macnaghten. 2013. "Developing a Framework for Responsible Innovation." *Research Policy* 42: 1568–1580.

Toom, V. 2012. "Bodies of Science and Law: Forensic DNA Profiling, Biological Bodies, and Biopower." *Journal of Law and Society* 39 (1): 150–166.

Toom, V., M. Wienroth, A. M'Charek, B. Prainsack, R. Williams, T. Duster, T. Heinemann, C. Kruse, H. Machado, and E. Murphy. 2016. "Approaching Ethical, Legal and Social Issues of Emerging Forensic DNA Phenotyping (FDP) Technologies Comprehensively: Reply to 'Forensic DNA Phenotyping: Predicting Human Appearance From Crime Scene Material for Investigative Purposes' by Manfred Kayser." *Forensic Science International: Genetics* 22: e1–e4. doi:10.1016/j.fsigen.2016.01.010.

Vailly, J. 2017. "The Politics of Suspects' geo-Genetic Origin in France: The Conditions, Expression, and Effects of Problematisation." *BioSocieties* 12 (1): 66–88.

Wainwright, S. P., C. Williams, M. Michael, B. Farsides, and A. Cribb. 2006. "Ethical Boundary-Work in the Embryonic Stem Cell Laboratory." *Sociology of Health & Illness* 28 (6): 732–748. doi:10.1111/j.1467-9566.2006.00539.x.

Wienroth, M. 2018. "Governing Anticipatory Technology Practices. Forensic DNA Phenotyping and the Forensic Genetics Community in Europe." *New Genetics and Society* 37 (2): 137–152.

Williams, R., and P. Johnson. 2004. "'Wonderment and Dread': Representations of DNA in Ethical Disputes About Forensic DNA Databases'." *New Genetics and Society* 23 (2): 205–223.