Report on recommendations to address the ethical and societal challenges of FDP

Dr Gabrielle Samuel – King's College London, UK Professor

Barbara Prainsack - King's College London, UK

Dr Gabrielle Samuel

Department of Global Health and Social
Medicine King's College London
Bush House, North East (NE)
Wing 30 Aldwych
London WC2B 4BG

Professor Barbara Prainsack
Department of Global Health and Social
Medicine King's College London
Bush House, North East (NE)
Wing 30 Aldwych
London WC2B 4BG

Please cite as:
Samuel, G, Prainsack, B (2020). Report on recommendations to address the ethical and societal challenges of FDP. VISAGE

The work leading to the results in this report has received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement number 740580 (VISAGE).

May 2020

Table of contents

Executive summary

The purpose of this Report is to act as a Deliverable for the VISAGE ('Visible Attributes through genomics') project. The project's aim is to develop, validate and implement in the relevant environment of routine forensic DNA service, a set of new prototype DNA-based tools to allow for appearance, age, and biogeographical ancestry (BGA) prediction of unknown perpetrators from various crime scene traces in an efficient way. In this Report, and in line with the nomenclature within the VISAGE consortium, appearance as well as BGA and age are referred to as externally visible characteristics (EVCs), and the (probabilistic) predictive testing of EVCs are referred to with the umbrella term "forensic DNA phenotyping" (FDP). This label for FDP is consistent across our Deliverables and other Reports emerging from the VISAGE project.[1]

This Report recommends measures to be undertaken that enable the implementation of FDP in an ethically and societally responsible manner, including suggestions as to how privacy-enhancing measures can be built into the technology, as well as into legal and regulatory infrastructures and processes (*Privacy-by-Design*). The VISAGE project has implemented the principle of Privacy-by-Design in its own work, e.g. in the development of the VISAGE prototype software tool to obtain appearance and BGA probabilities as well as age estimates. This Report goes further than this by including recommendations beyond VISAGE's remit, to be implemented by legislators, scientists, policymakers, and those providing professional training to actors within the criminal justice system. Whenever recommendations are directed at specific (groups of) actors we identify these explicitly.

The Report builds on work conducted in VISAGE Work Package 5 in years 1 and 2, which have been reported (and were published) as *Deliverable 5.1: a map of the legal and regulatory landscape related to FDP in EU member state countries*, and *Deliverable 5.2: a comprehensive analysis of the main challenges for the ethically and societally responsible implementation of FDP*. This work has highlighted deep-seated disagreements that academics as well as civil society stakeholders hold about the use of FDP in the criminal justice system. Much of this disagreement revolves around the use of predictive testing for BGA, and is particularly strong in some countries, such as Germany, where BGA has been excluded from the recent legalisation of FDP, which was restricted to particular appearance traits and age. Views range from those in favour of the technology, to those who consider it acceptable in a narrow range of cases and with strong safeguards in place, to those who are opposed to it in

[1] At the same time, we acknowledge - and indeed our previous research has identified - that nomenclatures vary across national, institutional, and disciplinary contexts, and some stakeholders prefer not to subsume age and/or BGA under the label of umbrella of EVCs or FDP.

principle because they feel it raises a number of human rights concerns related to privacy and non-discrimination, and could lead to racial profiling (Samuel and Prainsack 2019). In particular, a range of authors and civil society stakeholders are concerned that the burden of the technology will predominantly be placed on minority groups. These individuals argue that FDP is deemed more useful if it indicates an inclusion of a minority group (i.e., the presence of an infrequent trait) than if it indicates an exclusion (i.e., the absence of an infrequent trait) (see e.g. (Ossorio 2006; Toom et al. 2016)). FDP developers respond to this by noting that there is no simple relationship between the presence/absence of a particular genetic trait, and minority group discrimination, because individuals who have infrequent traits are not always discriminated against (for example, individuals with red hair). In any event, the structural racism[2] and implicit bias against certain minorities that is present in many societies means that special care needs to be taken to ensure that new technologies are not used in a discriminatory fashion.

Against this backdrop, our recommendations include safeguards for FDP implementation and use addressing the concerns of privacy and non-discrimination, not only at the level of the technology, but also at the level of processes, evaluation and policy. We are aware that such safeguards cannot fully abolish the risk of discrimination against minorities in an already structurally racist society.

This Report does *not* provide recommendations on whether FDP should be implemented and used within the criminal justice system of any specific jurisdiction. If national policymakers do choose to permit the use of FDP, we advise they follow our recommendations for best practice. Our key findings and recommendations are highlighted in the table below, which describes the features and processes that need to be implemented to ensure the ethically and socially responsible use of FDP in the criminal justice system. The recommendations are discussed in more detail in the remainder of the Report.

Feature or process related to FDP testing	**Our recommendation**
VISAGE prototype software – privacy	• The VISAGE prototype software should only be run on computers that are not connected to the internet.

[2] Structural racism does not require intent; it is embodied by, and inscribed in, our societal and political institutions and shared practices. Thus, to say that FDP bears the risk of being used in a racially discriminatory manner does not imply that it is used with the intent of being racist, or that the person or institution doing so has a conscious dislike of certain groups of people. Instead, it is the context within which a discriminatory act takes place that makes it racist (or not) (Samuel and Prainsack 2019).

and data protection considerations	• The VISAGE prototype software should only store information that is essential for quality assurance, for documentation, and for prediction during the FDP analysis (principle of data minimisation (Privacy-by-Design). • The VISAGE prototype software should not allow the selection of priors.[3]
Infrastructures and processes - privacy and data protection considerations	• The same safeguards for transferring DNA information should apply as in other forensic DNA analyses. • When the laboratory that will be conducting FDP receives the DNA sample, sufficient measures should be taken to ensure the laboratory confirms that the correct sample has been delivered. • The VISAGE prototype software output report and associated files should only be viewed by the reporting officer (i.e., a specially trained forensic DNA scientist experienced in reporting forensic DNA outcomes to law enforcement staff). The reporting officer should be trained in the use of VISAGE software and the interpretation of FDP findings. The reporting officer should use the VISAGE prototype software output report to prepare a written report to be shared with the requesting law enforcement agency. • The reporting officer's written report summarising the VISAGE prototype software output report should only be shared with the requesting law enforcement agency and not with any other organisations or individuals. There should be restrictions on who the requesting organisations can share these findings with, which should not normally include individuals outside the judicial system and law enforcement. Moreover, unless the FDP findings can inform other cases where a link is suspected, the law enforcement agency should use the reporting officer's written report summarising the VISAGE prototype software output only for the case for which the FDP analysis was originally intended (i.e. no speculative searching). • Law enforcement officers should only request FDP analyses from laboratories that follow current standards and best practices in terms of handling and analysing DNA samples for forensic purposes i.e., that use FDP tests that have been scientifically, technically and forensically validated, and published in peer-reviewed scientific journals (see section 3.2.2 recommendation 3). • The DNA sequence information, the DNA sample from which the DNA sequence was determined, and the VISAGE prototype software output file should be stored in the forensic laboratory in line with national legal requirements for DNA sample and sequence information storage (where such storage is legal), as they pertain to criminal cases which remain open, as well as those which have been closed. There is no requirement to have additional procedures implemented for FDP in terms of data usage and storage.
VISAGE prototype software - general recommendations	• Each page of the VISAGE prototype software output report should be marked "for internal use only". • For bi-parental BGA testing using autosomal single nucleotide polymorphisms (SNPs), the VISAGE prototype software output report

[3] Short for 'prior probability distribution'; they describe the pre-existing knowledge and beliefs about a variable before evidence is acquired.

for reporting FDP findings	should include the probabilities for all tested geographic regions. A separate data sheet should provide information on the reference population including the sample sizes used for testing and references to published scientific sources upon which the analysis was based. • For paternal BGA testing using Y-chromosomal SNPs and derived Y haplogroups, the VISAGE prototype software output report should include the name of the inferred Y-chromosomal haplogroup, as well as a table and map stating the published Y-chromosomal haplogroup frequencies by country and, where applicable, by ethnic group. Countries/regions for which no data exists should be clearly distinguished from countries/regions that have a zero-frequency estimate. A separate data sheet should provide information about the reference populations, including the sample sizes used for testing and references to published scientific sources upon which the analysis was based. • For age and appearance testing, the VISAGE prototype software output report should include the probabilities of all tested appearance trait categories, as well as the estimated age. Where appropriate, this can be represented in a table. For every predicted appearance trait or age estimation, a separate data sheet providing details about the predictive markers; prediction model method; prediction model accuracy; data used for prediction model building and validation; and recommendations for the interpretation of the software outcome should be provided, together with references to the published scientific sources for the provided information. The VISAGE prototype software output report should also include (a) for age predictions, a statement on the source tissue, and (b) for all predictions, a statement on the sex indicated by the analysis of the DNA sample (see section 3.2.1 recommendation 2). • When the reporting officer's written report, produced from the VISAGE prototype software output report, is provided to law enforcement officers, it should be accompanied by a verbal explanation of the findings, and the possibility for law enforcement officers to ask questions.
Recommendations related to the circumstances under which FDP should be used	• FDP regulation should not focus on "traditional" coding-non-coding boundaries (see section 3.2.2 recommendation 1). • Only phenotypic tests for inference of appearance, BGA, and age should be permitted for use in the criminal justice system. FDP testing for phenotypic traits that are linked to an 'internal' characteristic should not be permitted. • FDP should only be permitted for those phenotypic traits whose predictive marker and prediction models, and analytical FDP laboratory tools, have been scientifically validated (see section 3.2.2 recommendation 3). • FDP should only be permitted for those phenotypic DNA tests that have been technically and forensically validated (see section 3.2.2 recommendation 3). • FDP should only be used in criminal cases investigating serious crimes, as specified in a jurisdiction's criminal code. • FDP should only be used in 'exceptional circumstances'. Where possible within the institutional arrangements of a criminal justice system, FDP use should require an order of a Magistrate or another oversight body, and this should be underwritten in regulation.

Recommendations related to law enforcement officers' interpretations of FDP findings	• FDP findings should only be used as an investigative tool. • FDP findings should never be placed in a national or international police DNA database, for which there is no investigative need. • Law enforcement officers should not treat FDP findings analogous to the testimony of an eyewitness (i.e., a 'biological eyewitness') or any other evidence, but rather view the technology as a tool in its own right, which needs to be assessed against independent scientific criteria. • The outcome/interpretation of an FDP analysis should only be communicated to a law enforcement officer who has been trained in understanding such findings. • When a forensic DNA scientist provides the written FDP report to a law enforcement officer, this should be accompanied by a verbal discussion between both individuals to ensure information is communicated accurately (see section 3.2.1 recommendation 4). • FDP findings should not normally be the sole basis for law enforcement decision-making. • Law enforcement officers should not communicate FDP findings to the public via public media unless there is substantial evidence to corroborate them, or when communicating the findings to the public is necessary for the measure to succeed (e.g. DNA dragnet). • During DNA dragnet searches, law enforcement officers should communicate FDP findings to those involved in the dragnet in an open and transparent way and invite dialogue rather than merely passing on "facts".
Recommendations related to interpretation of FDP findings – education and training	• Recognising the different training needs for different types of practitioners, context-specific education and training in interpreting FDP findings is required for: all law enforcement officers ordering an FDP test and/or using FDP data; forensic DNA scientists producing FDP data; and forensic DNA scientists (reporting officer) interpreting FDP outcomes. Education strategies should also apply to other actors in the criminal justice system, including prosecutors, defence lawyers and Magistrates.
Evaluation of FDP use in the criminal justice system	• All FDP research outcomes that provide the basis for FDP tools applied in practice must be scientifically, technically and forensically validated, and published in peer-reviewed journals (including open access to algorithms and workflows) such that they are transparent and open to scrutiny by experts. • The outcomes of individual instances of FDP use by law enforcement officers should be reported back to the forensic DNA scientists developing the technology so that they can monitor their prediction software and adjust as and when appropriate. • The use and outcome of FDP by law enforcement authorities should be recorded and made available to an oversight body if requested (see section 3.2.5). This oversight body should be democratically legitimised and regularly evaluate the effects and use of FDP. Evaluation should be retrospective and observational. Where possible, civil society representatives should participate in oversight.
Recommendations specific for	• Only phenotypic tests that can predict appearance, BGA and age from DNA should be permitted for FDP use. • FDP should only be permitted for those phenotypic traits whose predictive marker and prediction models have been scientifically validated.

unidentified person cases[4]	• FDP should only be permitted for those phenotypic DNA tests that have been technically and forensically validated. • Context-specific education and training in interpreting FDP findings should be given to all law enforcement officers and/or other officers who handle unidentified person cases • The outcome/interpretation of an FDP analysis should be communicated very carefully via verbal communication to the officer who requested the analysis (and not reported to anyone outside of law enforcement or of the requesting authority).

4 FDP for unidentified person cases is not an aspect of VISAGE. Nevertheless, for the sake of comprehensiveness, recommendations pertaining to this use of FDP are included in this Report.

1. Introduction

1.1 Forensic DNA phenotyping

The VISAGE project aims to develop, validate and give recommendations for the responsible implementation in the relevant environment of routine forensic DNA service, a set of new prototype DNA-based tools to allow for appearance, age, and biogeographical ancestry (BGA)[5] prediction of unknown perpetrators from various crime scene traces.[6] FDP findings are probabilistic, meaning that they can only infer a specific phenotypic feature to a certain degree of probability. Whenever the word 'prediction' is used in this Report, it is to be understood as *probabilistic* inference. Below we provide more information about FDP for appearance, age and BGA inference, as well as the purpose of FDP findings.

1.2 Biogeographical ancestry, age and appearance inference

1.2.1 Biogeographical ancestry inference (BGA)

A person's biogeographical ancestry describes the geographic region in the world where the person's genetic ancestors originate from (also referred to as genetic ancestry). Only about 10% of genetic variation between individuals is dependent on their population of origin (Rosenberg et al. 2002). Nevertheless, this minor proportion of genetic differences between people from different places can be used to derive ancestry-informative DNA markers and for developing DNA tests for ancestry prediction, which are useful in forensic applications.

The genetic markers used to infer BGA in the VISAGE project are mostly Single Nucleotide Polymorphisms (SNPs).[7] These genetic markers are found on all autosomal chromosomes, as well as the sex chromosomes and the extrachromosomal mitochondrial genome. BGA is mostly predicted at the continental level,[8] that is, the continental region that the person's genetic ancestors originate from, and cannot be used to provide information about a suspected perpetrator's nationality. Nationalities and borders have political and cultural, not biological roots (see also Karberg 2018), and similarly, ethnic and racial identities are articulated through personal and social practices. BGA is thus not a proxy for nationality, ethnicity, or race.

[5] Not to be confused with lineage ancestry which mainly seeks to give information about an individual's paternal or maternal lineage.
[6] This could include "composite sketches" in the way that we explain on the VISAGE website: http://www.visage-h2020.eu/#FAQ.
[7] SNP: a variation in a single nucleotide that occurs at a specific position in the genome.
[8] We note that some markers can make inferences about BGA at a sub-continental level.

1.2.2 Age inference

In this Report, age inference refers to the prediction of the age of an unknown perpetrator from DNA for forensic purposes. Whilst various genetic tests have been used to infer biological age, those based on the analysis of epigenetic markers, specifically DNA methylation patterns (the presence of a specific 'methyl' molecule on the 'cytosine' nucleotide), otherwise known as epigenetic markers, have been shown to be the most reliable in the sense that they perform best in 'predicting' chronological age (Hannum et al. 2013). DNA methylation regulates gene expression (whether a specific gene is active or not active), and an individual's DNA methylation pattern (i.e., which parts of their DNA are methylated and which are not, as well as the relative proportion of methylation at a single site) changes with age. Specifically, about 20% of variation in DNA methylation in the human genome is correlated with the change in age and this relates to the presence of specific 'methyl' molecules on cytosines in a specific context of a DNA sequence motif.

1.2.3 Appearance inference

In this Report, appearance inference refers to the prediction of the appearance from DNA for forensic purposes. Appearance traits can be distinguished from a person's 'invisible' or 'hidden' traits, such as those related to specific health conditions or diseases. Prediction tests for appearance normally involve SNP testing and examples of characteristics which have been used for FDP include eye, hair and skin colour.

1.3 Purpose of FDP findings

In current police and judicial practices, DNA profiling is used to either exclude a suspect or to ascertain the identity of a suspected perpetrator. It can normally also be used as evidence in court. FDP is different; rather than providing confirmatory evidence, the key utility of the technology is seen to be in an investigatory context i.e., as a probabilistic predictor of characteristics about an unknown person to aid as one part in a wider investigation to identify an individual. Moreover, among experts working within the criminal justice system in European Union (EU) member countries, there is wide agreement that whilst FDP findings in a criminal case need to remain in a case file as evidence, FDP findings should never be used as substantiation that a suspected perpetrator had committed a crime. Once arrested, the suspected perpetrator's DNA has to be confirmed against the original DNA sample found at a

crime scene using "traditional" STR-based DNA analysis[9] (Cino 2017, de Cerqueira et al. 2016, Kayser and Schneider 2009). When the obtained statistical power is strong enough, however, FDP can establish that a specific person was *not* the sample donor.

FDP is also used for age, appearance and BGA prediction testing in unidentified person cases. Most scholars and stakeholders view the use of FDP for these cases as less ethically problematic. Furthermore, regulation pertinent to unidentified person cases is different in each EU member state; only some jurisdictions govern these cases under their penal code. We address the use of FDP for this purpose in a separate section of the Report.[10]

1.4 Coding and non-coding regions of DNA

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. Similarly, it has been assumed 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 was less ethically problematic to conduct forensic DNA testing/profiling using non-coding genetic markers (Benecke 2002; Kayser and de Knijff 2011).[11] The coding/non-coding boundary is redundant for FDP, which uses genetic markers that are located in both the non-coding and coding regions of the genome (Samuel and Prainsack 2018a).

As discussed in Deliverable 5.1 (Samuel and Prainsack 2018b), standard DNA profiling technologies using non-coding regions identify people based on an individual's unique set of genetic markers. The genetic markers used for this purpose do not provide any other information than the DNA profile used for individual identification via the principle of DNA profile comparisons. Identifying individuals in this way is seen to avoid privacy issues that relate to an individual's observable or internal characteristics, because the markers do not provide any information about such characteristics. However, research carried out during the last decade or so has clearly demonstrated that non-coding regions of DNA can contain

[9] A Short Tandem Repeat (STR) is a short sequence of DNA, normally 3-5 base pairs in length, that are tandemly repeated numerous times. STR-based analysis typically involves comparing a specific set of STR markers from an unidentified crime scene sample with the markers from a DNA sample of a suspected perpetrator to determine whether they "match". Two profiles match if the tested markers in the same locations look exactly the same.

[10] See footnote 4.

[11] Non-coding STR markers were not selected because they were ethically less problematic, but because they were highly polymorphic (which means that by nature they are not associated with phenotypes). Nonetheless, the non-coding-coding boundary offered a useful ethical boundary.

regulatory information relevant for coding regions. Such DNA regulators interact with genes to switch genes off or on and therefore are of similar importance for the expression of a gene to coding variants within a gene. Scientific evidence has demonstrated that a considerable proportion of gene function variation is provided by regulatory DNA elements, which can be within genes, nearby genes or even distant to genes, and not only by protein-coding DNA variants within genes as had been assumed in the past.

In instances in which a specific "non-coding" genetic marker is positioned close to a specific coding region, analysis of that specific genetic marker can inform us about the coding region. 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 SNP variant can be predictive of a specific observable trait. Indeed, a number of regulatory SNPs and SNPs in linkage disequilibrium with (currently unknown) coding SNPs are used to predict EVCs. For BGA inference from DNA, mostly non-coding SNPs, but sometimes coding SNPs are used. As such, the coding/non-coding distinction as it applies to what it can tell us about a person's observable characteristics, as in the case of FDP, is not binary.

1.5 Review of challenges associated with the ethically and societally responsible implementation of FDP

We reviewed the ethical and social challenges associated with FDP implementation and use in the criminal justice system in VISAGE Deliverable 5.2 (Samuel and Prainsack 2019). Below we provide a short summary of our findings to provide context to our recommendations.

1.5.1 Overview

Supporters of FDP hold that FDP is uniquely placed to provide leads in cases where no other leads are available, and that FDP use can be understood in analogy to a "biological eyewitness". As long as it is paired with suitable safety and accountability measures (as we describe in this Report), supporters argue, we have an ethical obligation to honour our commitment to try to solve crimes and enhance security. FDP could indeed have a positive impact on society if it turns out to be an adequate tool to increase effective criminal justice without leading to any undue harm to individuals and groups (e.g., by better targeting investigations, so that fewer innocent people are implicated). Societies - and particularly

those members of our societies whose friends and family were affected by serious crimes that may previously not have been solved, will benefit from progress in (or even solution of) the investigation of these crimes. They may also benefit from cases being solved faster, and/or investigations being more efficient. It would also be a clear benefit for society as a whole if perpetrators who otherwise might commit additional crimes could be apprehended. Finally, the use of FDP could help to strengthen trust in the criminal justice system and could have a positive effect on community life and shared values.

At the same time, key concerns relating to FDP include (a) the risks of FDP infringing privacy rights and supporting discrimination and stigmatisation along racial, ethnic and/or religious lines, (b) the effectiveness of FDP as an approach to helping identify a suspected perpetrator being uncertain, and (c) other forms of addressing the threat of crime may receive fewer resources and attention with increasing resources being invested into FDP. These arguments could also be made on the basis of the responsibility to use public funds prudently; it may be more prudent to invest in measures and technologies that are known to be effective in solving or reducing crime and can be used in a wide range of cases. For some, these latter concerns carry so much weight that they oppose the use of FDP entirely. Others argue for rigorous oversight, such that in a particular investigation the rights of suspects, as well as the population of interest, are handled with care and awareness of the potential risks associated with the technology. These challenges are described in more detail below.

1.5.2 Specific challenges

Privacy. When used in actual casework, FDP could interfere with people's privacy by including people in the investigation (as suspects, witnesses, or family members or friends of suspects) who would otherwise (without FDP) not have been included. While the inclusion of innocent people into criminal investigations is a routine occurrence in the criminal justice system that cannot be abolished entirely, if FDP were to increase the number or scope of people who are included in investigations, then this may have a negative impact not only on privacy but also on people's family lives. Similarly, if it increased the proportion of members of stigmatised or marginalised groups and minorities among those innocently implicated in investigations, this would have a negative effect on security as a whole. Issues also arise if an FDP finding can also provide information about a health condition, which - in contrast to appearance, age or BGA[12] - is likely not to be known to the person if the disease has not (yet) caused specific symptoms. For example, in the case of age prediction, a disparity between biological and chronological age can be an indication of an underlying health condition.

[12] Though, we note that in some instances BGA inference may reveal a narrative of BGA which the suspected perpetrator was unaware of, or which would conflict with aspects of their identity that they hold dear.

Data protection. In the EU context, personal data is regulated under the General Data Protection Regulation (GDPR). Under its remit, genetic information and other information about identified or identifiable persons are considered personal information and thus protected.[13] It is not clear whether FDP information obtained from an unknown crime scene sample falls under the remit of personal information. It is arguable that it is not, because it is not at that time relatable to an identified or identifiable person. Even if FDP information were treated as personal information, its processing by "competent authorities"[14] for law enforcement purposes does not fall under the remit of the GDPR because it is governed by a special law; the "Police Directive" (Directive 2016/680 on the protection of natural persons with regard to the processing of personal data by competent authorities for the purpose of law enforcement).[15] The Police Directive limits the rights of subjects regarding notification, consent, erasure, etc compared to the rights granted within the GDPR. Either way, an ethical imperative remains to safeguard any FDP data or findings against misuse or abuse.

Discrimination. If no adequate safeguards are in place, FDP could be used in ways that discriminate against, and stigmatise, persons and groups. This would affect people's civil rights and liberties as well as their culture and community. One example of this is related to ethnic minority populations. For instance, if FDP were used in such a way that emphasised dividing lines around ethnic, religious or 'cultural' lines, such that certain people were set apart from others in a public-facing manner (e.g. people with genetic ancestry from certain world regions) then this is likely to have a destructive effect on social cohesion and could have the effect of introducing new, or reinforcing existing, divisions in our society - especially if these groups were associated with a higher prevalence of crime. These risks relate to the context of structural racism, which does not require intent but is rather embodied by, and inscribed in, our societal and political institutions and shared practices. Discrimination could happen at a number of levels, including police misinterpretations of FDP findings which could lead to racial profiling; if FDP findings are released to the public it could upset community and social relations; bias inherent in the algorithms and data sets used

[13] GDPR (Article 4) states: "'personal data' means any information relating to an identified or identifiable natural person ('data subject'); an identifiable natural person is one who can be identified, directly or indirectly, in particular by reference to an identifier such as a name, an identification number, location data, an online identifier or to one or more factors specific to the physical, physiological, genetic, mental, economic, cultural or social identity of that natural person' ".

[14] "Competent authority" means: "any public authority competent for the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security; or (b) any other body or entity entrusted by Member State law to exercise public authority and public powers for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, including the safeguarding against and the prevention of threats to public security". https://eur-lex.europa.eu/eli/dir/2016/680/oj

[15] https://eur-lex.europa.eu/eli/dir/2016/680/oj.

in FDP; and in leading to reification of the mistaken belief of a biological basis of race, which might, in turn, deepen the social divide between different groups or individuals, and lead to stigmatisation.

Misinterpretation of FDP findings. There are concerns that the misinterpretation of FDP findings (because of difficulties interpreting probabilistic findings, the presence of stereotypes in European societies that could influence this interpretation, or the trust that people often place in DNA findings)[16] could lead law enforcement to follow unnecessary false leads. FDP interpretation education for law enforcement officers and other actors in the criminal justice system has been called for to address these issues, though these calls have lacked specificities about how this could be achieved. Moreover, these approaches are problematic for a number of reasons, including issues of capacity, and issues relating to ineffective strategies which cannot overcome deep-seated bias.

We note that many of the challenges we have described above are not specific to FDP, but they can also apply to other forensic DNA technologies (STR DNA profiling, genetic genealogy searching etc) and non-DNA technologies (biometric technologies, facial recognition etc). Therefore, the Recommendations we lay out in this Report, which have been developed to address the above concerns, fit into broader work in the forensic technology arena. This work involves balancing the principles and norms of protecting human rights with the rights and public good that forensic technologies seek to protect and foster (Williams and Wienroth 2017, Guillen et al. 2000, Toom 2018). In terms of FDP, this can result in undue restrictions on people's freedoms when people are included in an investigation on the basis of FDP without a justifiable reason. But a lot hinges on what would be considered a justifiable reason; the involvement of people in investigations that are known to be innocent (e.g. as witnesses or family members of suspects), or later turn out to be innocent (as suspects, or as part of intelligence-led mass screening), is not unjustifiable *as such* - it is a feature of criminal investigation that cannot be abolished entirely. Our recommendations seek to provide clarity in terms of when the use of FDP is justified.

[16] The phenomenon whereby greater trust is placed in the scientific accuracy of forensic DNA technologies compared to other types of information and evidence in forensics and criminal investigation is known as the 'CSI effect' (e.g. (Baskin and Sommers 2010, Cole 2015, Cole and Dioso-Villa 2009, Schweitzer and Saks 2007, Holmgren and Fordham 2011). The 'physical and ontological location of DNA' (Prainsack 2009) contributes to the image of forensic DNA evidence as being a 'truth machine' (Lynch 2008). Because it is in the centre of our cell, and because it has a history of being described as the 'book', 'blueprint', or 'code' of life, DNA has assumed a symbolic place as the bearer of (at least part of) the essence of individuals, and the essence of humanness, in collective imageries (Nelkin and Lindee 1995).

2. Methods / terms of reference

Our recommendations were developed in five key stages:

- Stage 1: Drawing on our previous two years' work reported in Deliverables 5.1 and 5.2.
- Stage 2: Follow-up questions sent to forensic DNA scientists and law enforcement experts, as well as to VISAGE members.
- Stage 3. Consultations with VISAGE Ethics Advisory Board members.
- Stage 4: Consultations with VISAGE members at the 2019 VISAGE Consortium meeting.
- Stage 5: Two additional VISAGE meetings focusing on how to report FDP findings.
- Stage 6: Refinement of the draft recommendations based on final comments by VISAGE members, VISAGE Ethics Advisory Board members, and VISAGE Scientific Advisory Board members.

We provide details of each of these stages of recommendation development below.

Stage 1. Previous two years work

Deliverable 5.2, which analysed the challenges associated with the ethically and societally responsible implementation and use of FDP, provided the main considerations to address when developing our recommendations for best practice use of FDP in the criminal justice system. Our work in this Report is also informed by Deliverable 5.1 which provided an overview of the legal and regulatory landscape regarding FDP in EU countries and beyond.

Stage 2. Questions sent to forensic DNA scientists, law enforcement experts, and VISAGE members

This Report includes examples of best practice pertaining to the topics covered by our Recommendations. In order to collect these best practice cases, we contacted all members of the VISAGE team, as well as six experts who work in law enforcement or as forensic scientists in the field of FDP (identified previously as part of the VISAGE project WP5.1) and requested information on their practices related to FDP. Our overarching question asked what they consider best practice, especially with regards to privacy and reducing the risk of discrimination when using FDP. We also included specific questions relating to: when FDP is requested and used; which phenotypic traits are tested, and based on which criteria; which types of laboratories can conduct FDP testing in each country; how do law enforcement use FDP findings used in case work; how are FDP findings reported to law enforcement by forensic DNA scientists; and questions around data protection practices and FDP evaluation.

Stages 3 and 4. Refinement and consultation with VISAGE Ethics Advisory Board and VISAGE Consortium members

A. We discussed our draft recommendations with members of the VISAGE Ethics Advisory Board, including a number of unresolved issues regarding when and how best to use FDP in the criminal justice system.

B. Following the incorporation of comments from VISAGE Ethics Advisory Board members, remaining open questions were discussed with the VISAGE consortium. Specifically, a draft of the Report was sent to all members of the Consortium who were asked to read the recommendations prior to the December 2019 VISAGE Consortium meeting. At the meeting outstanding questions relating to the draft recommendations document were discussed.

Stage 5. VISAGE meetings on how to report FDP findings

A key social and ethical challenge to FDP implementation is the reporting of FDP findings, and their communication to law enforcement officers. Specific VISAGE meetings were arranged in January and March 2020 to address this issue with relation to the reporting of BGA using SNPs on the Y chromosome (Y-chromosomal haplogroups), and appearance and age traits, respectively. The meetings brought together Consortium members from key Work Packages that have a stake and/or interest in this issue. The discussions were multi-disciplinary, the outcomes of which have provided the basis for these elements of our recommendations.

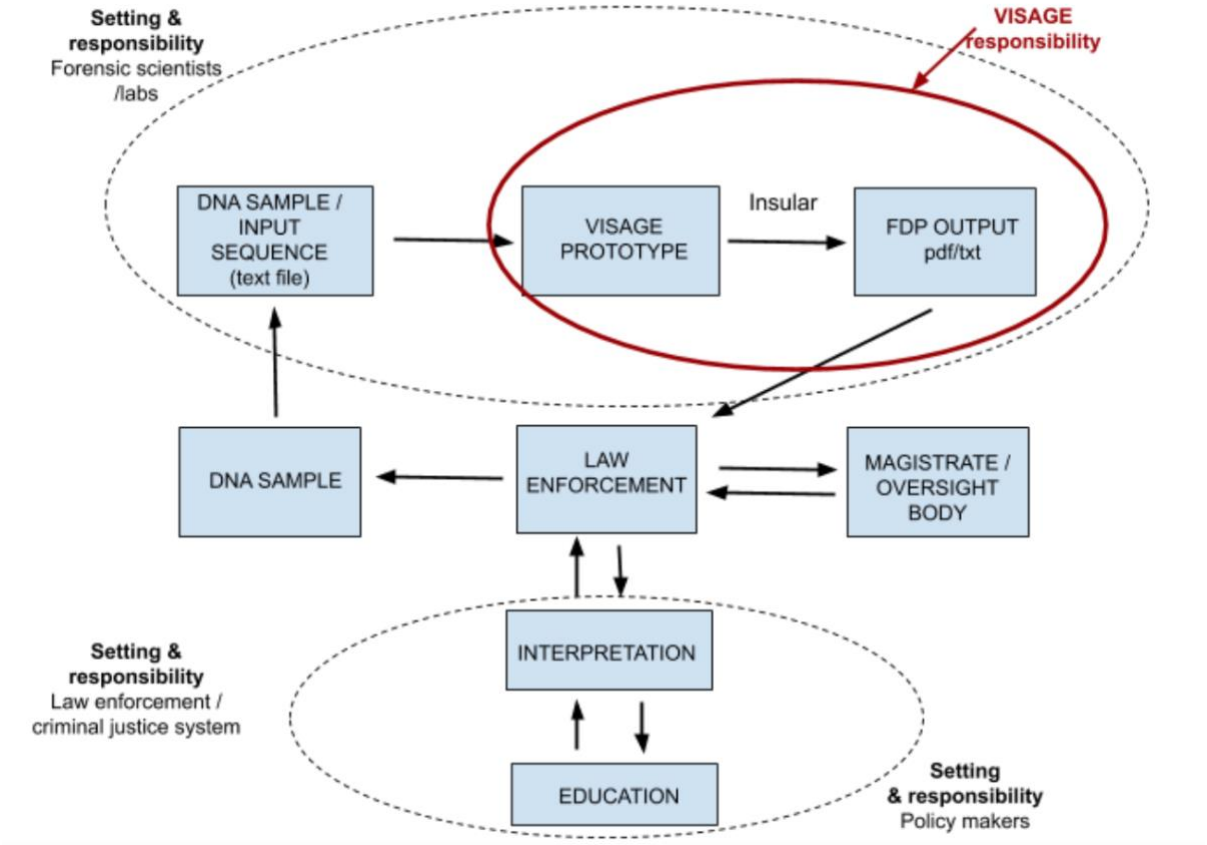
Stage 6: Refinement of document based on final comments by VISAGE members

The best practice recommendations document was emailed to VISAGE members, members of the VISAGE Ethics Advisory Board, and members of the Scientific Advisory Board, for final consideration and comment. On the basis of the feedback received, the document was refined to develop the final version.

3. Recommendations

Addressees of our recommendations include actors involved in the development, use and evaluation of FDP. We note that while our recommendations are broad-reaching - across all stages of FDP use and across software development as well as policy and practice - only some of these recommendations can and will apply to the development of the VISAGE prototype software itself (see Figure 1). It is the responsibility of other actors in the criminal justice system to ensure all our recommendations are integrated into best practice, to be applied to any use of FDP, whether using the VISAGE prototype software or otherwise.

Figure 1. Representation of the responsibilities of different actors in the criminal justice system with relation to the use of FDP



We divided our recommendations for the use of FDP for appearance, age, and BGA prediction of unknown perpetrators from various crime scene traces into three sections:

1. *Privacy-enhancing features*. This includes (a) features that can be built into the VISAGE technology in line with the principles of data minimisation, and (b) broader recommendations relating to privacy and data protection issues.
2. *General recommendations for FDP use in the criminal justice system.* These include recommendations related to (a) features that can be built into the VISAGE prototype software, (b) the appropriateness of using different phenotypic tests in the criminal justice system and the circumstances under which the use of FDP by law enforcement officers can be justified, (c) how law enforcement officers and other actors should interpret FDP findings, and (d) the need for transparent evaluation of any costs and benefits of FDP use in the criminal justice system.
3. *Country-specific recommendations.* This considers challenges and solutions that fit the particular regulatory, political, cultural, and economic configurations of the national setting for each of the VISAGE Consortium participant countries.

We also provide an additional, fourth section, pertaining to recommendations for the use of FDP for appearance, age, and BGA prediction in unidentified person cases.

We note that in line with the research design presented in the project proposal, this Report *does not* make recommendations about whether FDP should be used in the criminal justice system in any specific country, since this is a political question that needs to be decided upon by policymakers as the result of a democratic process. Rather, if countries do make the decision to permit FDP (or have already), this Report offers recommendations on how the use of FDP should be implemented in policy and practice.

3.1 Privacy-enhancing features

This section makes recommendations on how privacy-enhancing features can be built into the VISAGE prototype software (Privacy-by-Design), as well as legal and regulatory infrastructures and processes. The recommendations reflect principles enshrined in both the GDPR and Police Directive.[17] These include (a) data minimisation and storage limitation, which ensures the processing of personal data is limited only to what is necessary with regard to the purpose for which they are processed, and that the data are kept in a form which permits identification of data subjects for no longer than is necessary; (b) integrity and confidentiality, which ensures that data is processed in a manner that protects the security of the personal data, including protection against unauthorised or unlawful processing and against accidental loss, destruction or damage, using appropriate measures; (c) purpose limitation, which ensures data is only collected for a particular purpose and is only collected for as long as necessary; and (d) transparency.[18]

3.1.1 The VISAGE prototype software

These recommendations reflect privacy-enhancing features that should be built into the VISAGE prototype software. We also recommend they are built into all FDP software developed in the foreseeable future.

1. We recommend that ***the VISAGE prototype software should only be run on computers that are not connected to the internet.*** Ensuring the software is insular, which means that it cannot access online data (or be accessed online), enhances confidentiality because it decreases both the chance that the data can be hacked, and also the chance that the DNA sequence can be linked to other data. Linkage to other data increases the chance of re-identification. Such steps to maintain privacy are commonplace for both national forensic STR-profile databases, as well as for a wide range of clinical datasets; solutions for FDP testing could be moulded after these examples.

2. To respect the principle of data minimisation to the fullest, we recommend that ***the VISAGE prototype software should only store information that is essential for quality assurance, for documentation, and for prediction during the FDP analysis***. This includes the FDP report

[17] Directive on the protection of natural persons with regard to the processing of personal data by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.
[18] http://www.privacy-regulation.eu/en/article-5-principles-relating-to-processing-of-personal-data-GDPR.htm.

containing the actual prediction, other intermediate information generated by the software, as well as other necessary files such as sequence, methylation and genotype information. By ensuring the above, data minimisation and storage limitation are ensured, and there is less potential for the data to be leaked, hacked or misused at a later stage.

3. We recommend that ***the VISAGE prototype software does not allow the selection of priors.*** Priors describe the pre-existing knowledge and beliefs about a variable before some evidence is acquired. The VISAGE prototype software should not permit their selection because 1) evidence shows a partial high susceptibility to mis-classification of prior values as there is a lack of spatial distributional information, and 2) there is too coarse a BGA inference with limited marker sets: if priors are meant to be spatial prevalence values (the original idea in combination with BGA inference), those BGA regions can be too large. For example, eye colour prevalence changes even within Europe as one ancestry region (or class), most likely resulting in a misspecification of priors.

3.1.2 Data protection and quality control: legal and regulatory infrastructures and processes

These recommendations are relevant to forensic DNA scientists who use the VISAGE prototype software, or conduct other FDP testing in the criminal justice system for forensic purposes.

1. Law enforcement officers may give information about the criminal case under investigation to the laboratory that is conducting the FDP analysis. In some jurisdictions, case details are routinely shared with forensic laboratories. We recommend that ***in such instances, the same safeguards for transferring this information should apply as in other forensic DNA analyses.*** We note that limiting the transfer of potentially identifying information about the criminal case between individuals in the laboratory will decrease the likelihood of data leakage and/or data misuse, maintaining the confidentiality of the data.

2. In many instances, law enforcement officers will request an FDP analysis from the same laboratories that conducted earlier STR DNA profiling. In these instances, the laboratory will already have the crime scene sample which they can use for the analysis. However, sometimes FDP analysis is conducted in a different laboratory to the earlier STR DNA profiling. In these instances, we recommend that ***sufficient measures should be taken to ensure the laboratory confirm that the correct sample has been delivered.***

3. The VISAGE prototype software output report is not meant for the law enforcement agency directly, as it requires interpretation by a trained forensic DNA scientist. We recommend that ***the VISAGE prototype software output report and associated files only be viewed by the reporting officer (i.e., a specially trained forensic DNA scientist experienced in reporting forensic DNA outcomes to law enforcement staff). The reporting officer should be a specialist forensic DNA data scientist who is trained in the use of VISAGE prototype software and the interpretation of FDP findings. The reporting officer should use the VISAGE prototype software output report to prepare a written report to be shared with the requesting law enforcement agency,*** and not to any other organisations or individuals. In addition, in line with the principle of data minimisation, we recommend that ***there should be restrictions on who the requesting organisations can share the FDP written report with, which should be in line with national regulations relating to disclosing and non-disclosing, though should not normally include individuals outside the judicial system and law enforcement*** (for exceptions, see section 3.2.3 recommendation 6). In this way, the transfer of data remains as a circular closed loop, decreasing the chance of data leakage and/or data misuse, and maintaining the confidentiality of the data. We envisage two instances in which it may be useful for the requesting officers to share the FDP written report with other law enforcement officers: First, when law enforcement officers have a DNA sample from a criminal case that has a confirmed identical STR-profile to the DNA sample tested via FDP, and second, when the crime in question is linked to another crime in terms of M.O. (*modus operandi;* i.e., the specific behaviour of a perpetrator). In these instances, the FDP findings could inform other suspected linked cases.

4. Commercial companies market FDP as a testing service capable of creating composite faces of individuals from a sample of DNA alone; often based on tests which are underdeveloped, not sufficiently specified and documented, and unvalidated as far as knowledge made available via peer-reviewed publications is concerned (Samuel and Prainsack 2019). The concern here is that they can lead to incorrect conclusions from the DNA data and FDP findings, which law enforcement officers could potentially act upon. There is also a concern that data protection standards are not maintained. In spite of this, tests offered by these companies are already used by some EU law enforcement agencies to aid criminal investigation (Samuel and Prainsack 2019). We recommend that the same safeguards for choosing laboratories to conduct an FDP analysis should apply as in other forensic analysis: ***Law enforcement officers should only request FDP analyses from laboratories that follow current standards and best practices in terms of handling and analysing DNA samples for forensic purposes i.e., that use FDP tests that have been scientifically, technically and forensically validated, and published in peer-reviewed scientific journals*** (see section 3.2.2 recommendation 3)**.** These laboratories should also have obtained the same level of

certifications that would be required for STR analysis (if FDP analysis is not carried out in the same lab as STR analysis). This will provide safeguards to ensure adherence with standards relating to data protection, minimisation, storage limitation and confidentiality are maintained. We also note that requirements for the storage of genetic data in laboratories due to quality management and assurance guidelines already exist for laboratories. We recommend that ***there is no requirement to have additional procedures implemented for FDP in terms of data usage and storage.***

5. We recommend that ***all DNA sequence information, the DNA sample from which the DNA sequence was determined, and the VISAGE prototype software output file, should be stored in the laboratory in line with national legal requirements for DNA sample and sequence information storage (where such storage is legal), as they pertain to criminal cases which remain open, as well as those which have been closed.*** This is in line with the principle of purpose limitation that exists within EU data protection law (including the Police Directive and the GDPR).

3.2 General recommendations for FDP use

3.2.1 Recommendations relevant to the VISAGE prototype software

These recommendations align with the human right principles of respect for human dignity and non-discrimination. Specifically, they aim to reduce the risk that FDP could be used in a way that has a negative effect on respect for human dignity. This could occur if the technology was used in unduly discriminatory ways (i.e. on the basis of racist or religious stereotypes), or when it unduly restricts people's freedom (when people are included in an investigation on the basis of FDP without a justifiable reason). We provided a detailed analysis of the specific risks to human dignity, presumption of innocence and discrimination in our Societal Impact Assessment of FDP in Deliverable 5.2 (Samuel and Prainsack 2019). FDP should be used in such a way that it reduces, rather than increases, the number and range of people who are unduly involved in an investigation. Care should be taken in terms of how FDP findings are communicated to law enforcement officers, and how they are interpreted.

1. We recommend ***that for bi-parental BGA testing using autosomal SNPs, the VISAGE prototype software output report should include the probabilities for all tested geographic regions. A separate data sheet should provide information on the reference population, including the sample sizes used for testing and references to published scientific sources upon which the analysis was based.*** No DNA sequence information should be included.

   We recommend ***that for paternal BGA testing using Y-chromosomal SNPs and inferred Y haplogroups, the VISAGE prototype software output report should include the name of the inferred Y-chromosomal haplogroup; and a table and map stating the published Y-chromosomal haplogroup frequencies by country and, where applicable, by ethnic group.*** The map should be coloured or shaded in neutral colours (i.e., not traffic light colours) to mitigate the chances of misinterpreting the data. Neither the table nor the map should include DNA sequence information. Y-chromosomal haplogroup testing is prone to small sample sizes and missing data which can increase the risk of misinterpretation of BGA predictions. As such, we also recommend that the ***countries/regions for which no data exists should be clearly distinguished from countries/regions that have a zero-frequency estimate. A separate data sheet should provide information about the reference populations, including the sample sizes used for testing and references to published scientific sources upon which the analysis was based.***

2. We recommend ***that for age and appearance testing, the VISAGE prototype software output report should include the probabilities of all tested appearance trait categories, as well as the estimated age. Where appropriate, this can be represented in a table.*** For every predicted appearance trait or age estimation, a separate data sheet providing details about the the predictive markers; prediction model method; prediction model accuracy, expressed via typically used parameters such as sensitivity, specificity, AUC,[19] negative predictive value, positive predictive value, and for age prediction the prediction error; data used for prediction model building and validation; and recommendations for the interpretation of the software outcome should be provided, together with references to the published scientific sources for the provided information. We recommend that ***the VISAGE prototype software output report should also include (a) for age predictions, a statement on the source tissue, and (b) for all predictions, a statement on the sex indicated by the unknown DNA sample.***

3. We recommend ***that each page of the VISAGE prototype software output report should be marked "for internal use only"***. As described above, depending on the nature of the analysis, the VISAGE prototype software output report will contain information about predictive markers; probabilities of BGA for all tested geographic regions; and/or inferred Y-chromosomal haplogroups. This information is vital for the reporting officer who needs to accurately interpret the VISAGE prototype software output report when she/he develops their written report for law enforcement officers. The information could, however, be misinterpreted or misused by those who have limited or no training in the use of VISAGE software and the interpretation of FDP findings.

4. Our public engagement processes, reported in Deliverable 5.2, identified concerns about: the utility of findings with low probability values; law enforcement officers placing too much weight on such findings; and if a prediction reflected a law enforcement officer's unconscious bias, the fact that a law enforcement officer might be quicker to accept the finding than if the opposite was the case (i.e., the law enforcement officer will start to direct his/her investigation according to his/her more or less unconscious bias and therefore "find" new findings). As such, for many of our interviewees, there needed to be a high predictive value for FDP findings to be communicated with law enforcement (Samuel and Prainsack 2019). There is little indication in the literature regarding what such a threshold could look like i.e., how much risk is deemed appropriate to communicate to law enforcement officers. This is in spite of the fact that a wealth of literature exists on communicating uncertainty/risk in a range of fields (Fischhoff and Davis 2014, van der Bles et al. 2019). This is because this literature focuses on *how* to report risk rather than *when* (and to whom) to report it.

[19] Area under the ROC curve.

Exceptions include, for instance, predictive genomic testing, where questions about when to disclose a risk for a specific health condition are routinely asked (i.e., scientific vs clinical utility) (Sainani 2012, Cassa et al. 2012, Goddard et al. 2013, Chambrone and Armitage 2016). Though these discussions apply less to the forensic field. Given this, applying thresholds to the reporting of FDP findings will be arbitrary, and as such, we do not recommend them. Rather, we recommend that the FDP written report signed by the reporting officer, and given to law enforcement officers, should include an easily accessible (for a lay person) description of the findings, in addition to the probabilistic likelihoods as described above. We recommend that ***when a forensic DNA scientist provides the written FDP report to a law enforcement officer, this should be accompanied by a verbal explanation of the findings, and the possibility for law enforcement officers to ask questions***. This verbal explanation (phone call, meeting) between both individuals will give the opportunity for law enforcement officers to ask questions. This will ensure information is communicated accurately, that questions that may arise can be answered immediately, and that the forensic scientist can ascertain in real time the understanding of the law enforcement officer of what the findings mean. This will help mitigate the chance of misunderstanding. Previous research exploring how forensic science is communicated to law enforcement suggests discussion is one mode of valuable and effective communication (Howes 2017).

3.2.2 Recommendations related to the circumstances under which FDP should be used in the criminal justice system

As we noted earlier, this Report does not make recommendations about *whether* FDP should be used in the criminal justice system. Rather, for countries which have made the decision to permit FDP, this Report offers best practice recommendations on FDP policy and practice.

Potential exists for misuse of FDP. Misuse can manifest in terms of excessive surveillance, whereby law enforcement use FDP in circumstances that may not be illegal but may be ethically or politically problematic. Moreover, if FDP is used in actual casework, it could interfere with people's privacy by drawing suspicion upon somebody (because they fit an FDP profile). This may lead to them being included in the investigation (as suspects, witnesses, or family members or friends of suspects) when they would otherwise not have been included.[20] This, as such, is not necessarily disproportionate or otherwise undue. We are also aware that if used *appropriately*, FDP has the potential of decreasing privacy infringements by correctly narrowing a suspect population pool. But given the potential for

[20] This would be assessed analogous to the assessment of proportionality in connection with other forensic techniques or technologies that leads to the inclusion of people into investigations.

harm which can result from the use of FDP, to minimise the risk of privacy infringements, to ensure respect for human dignity and non-discrimination, and to maximise transparency, there is a duty for legislators to explicitly regulate its use. Existing tools such as the precautionary principle, or proportionality, tell us how to weigh different stakes, but while they are useful principles, they are not prescriptive, and cannot provide concrete guidance. Regulation is needed to inform law enforcement officers about what circumstances it is appropriate to use the technology, and for which phenotypic traits.

1. To address the potential for FDP misuse, which can manifest in terms of excessive surveillance if law enforcement or civil society use FDP in circumstances that may not be illegal but ethically or politically problematic**,** we recommend that ***the use of FDP is explicitly regulated***. As discussed in more detail in Section 1.4, the coding/non-coding distinction, as it applies to what it can tell us about a person's observable characteristics, is problematic and unfit for regulatory purposes in the context of FDP: the primary purpose of FDP is to provide information about observable characteristics and the technology draws on DNA sequences in both the coding and non-coding region to achieve this aim. As such, we recommend that ***FDP regulation should not focus on the "traditional" coding/non-coding boundaries***.

2. Phenotypic testing for traits related to non-visible, 'internal' characteristics - health, psychological, and/or mental disposition - excessively and unnecessarily infringe on the suspected perpetrator's privacy and should not be allowed. We recommend that ***only phenotypic tests that can predict appearance, biogeographical ancestry (BGA), and age (i.e., EVCs) be permitted for use in the criminal justice system.*** Though we note that while some aspects of appearance come close to being "objectively" determinable in terms of externally visible features, this does not apply to the age of a person and her or his BGA: Both have an influence on externally visible features, but cannot be objectively determined from the outside. As such, some EVC prediction testing could provide information about non-visible traits that may be otherwise unknown to an individual. This can happen, for example, if the testing reveals an unknown health condition.[21] It can also happen if the BGA testing reveals a narrative of BGA that the suspected perpetrator was unaware of, or which would conflict with aspects of their identity that they hold dear. In the unlikely circumstance that an association is identified between an EVC and an internal characteristic at the population level, we recommend that ***FDP testing for the phenotypic trait which is linked to an 'internal' characteristic should not be permitted.*** In the circumstance that an association is identified between either an age prediction test and an underlying health condition at an

[21] We note that current markers used for FDP testing do not currently reveal such information.

individual level, or a BGA test and an individual's seeming biogeographical ancestry narrative, we recommend that regulators have systems in place for handling such information (disclosure versus nondisclosure). Similarly, regulators need to consider how to address some EVCs sitting at the boundary of what is deemed by society as a health condition, for example height and dwarfism.

3. FDP technology uses algorithms to predict specific appearance traits. These algorithms have been 'trained'[22] using specific DNA datasets. Currently there is no single "reference" database of genetic sequences used for FDP model training or analysis (Liu et al. 2009). Concerns arise if care is not taken by the scientists who develop such statistical prediction models such that all forms of the predictable appearance traits are well represented in the underlying reference dataset used for model building. Regarding BGA, prediction works by using a reference dataset directly. Concerns for BGA arise when, for example, certain geographic regions are underrepresented in the reference dataset meaning that specific populations may respectively be overrepresented, leading to systematic bias in the findings produced from the analysis (Samuel and Prainsack 2019). To introduce standards of reproducibility of results, we recommend that ***FDP should only be permitted for those phenotypic traits whose predictive marker and prediction models, and analytical FDP tools, have been scientifically validated, and for those phenotypic DNA tests that have been technically and forensically validated***. Specifically, validation should be made public via publication in peer-review scientific journals. Validation includes the following steps: (a) the validation of DNA markers and the genetic model, (b) the markers and model being applied to a second set of samples for independent verification, and (c) the successful use of the analytical FDP laboratory tool in the forensic context (typically referred to as forensic developmental validation), i.e. on a limited amount of DNA, mixed DNA, DNA which has been degraded, DNA that has been contaminated with inhibitors, analysis of mock casework samples etc.

4. FDP is an intrusive technology insofar as it involves the sequencing and analysis of the suspected perpetrator's DNA, and therefore can potentially invade their privacy. Any infringement of privacy should be minimised, and justified in the sense that the technology is only used when the potential benefit of apprehending the suspect outweighs the potential harms which may ensue from using the technology (infringement on privacy, but also potential for subpopulation discrimination because the creation of suspect populations may be based on the predictive technology). We recommend that ***FDP should only be used in***

[22] For appearance trait prediction algorithms are trained and tested using datasets which contain both the genotypic information as well as the phenotypic information of a large number of subjects that allows to test if a prediction outcome is correct or incorrect and within the latter if it is false positive or false negative.

criminal cases investigating serious crimes (as specified in a jurisdiction's criminal code) which have caused serious physical harm to an individual, including homicide, sexual assault and terrorist acts, and that this be underwritten in regulation.

5. FDP should only be used if law enforcement officers can show that other, less intrusive means (for privacy) are not effective in a specific case; and only when the use of FDP could be of significant help in apprehending the suspected perpetrator. We define this as 'exceptional circumstances'. We recommend ***that FDP is only used in such 'exceptional circumstances'.*** Where possible within the institutional arrangements of a criminal justice system, ***FDP use in the criminal justice system should require an order of a Magistrate or another oversight body***.

3.2.3 Recommendations related to law enforcement officers' interpretation of FDP findings

These recommendations relate to operational issues within law enforcement. In particular, they address issues relating to the uncertain and predictive nature of FDP findings, and concerns related to their possible misinterpretation. These recommendations, if put in place, will strengthen the control mechanisms around the use of FDP in law enforcement practice.

1. We recommend that ***FDP findings should only be used as an investigative tool.*** The key utility of FDP is in an investigatory context i.e., as a probabilistic predictor of characteristics about an unknown suspected perpetrator to aid as one part in a wider investigation to identify the individual. And while FDP findings in a criminal case need to remain in a case file as evidence, FDP findings need never be used as confirmation that a suspected perpetrator had committed a crime. This is because if a suspected perpetrator is arrested on the basis of FDP findings and corroborated with other evidence, his/her DNA has to be confirmed against the original DNA sample found at a crime scene using "traditional" STR-(or SNP) based DNA profiling analysis (i.e., if the suspected perpetrator's STR profile matches that obtained from the crime scene it shows that the suspect is the sample donor, if it does not match it demonstrates that the suspect is not the sample donor).

2. There is no value in storing FDP findings in a central national or international database, given that the key utility of FDP is as an investigatory tool in specific cases. While sharing FDP findings between law enforcement forces could be beneficial, such intelligence sharing does not require phenotype markers to be included in any criminal database (Scudder et al. 2018). We recommend that ***FDP outcomes should never be placed in a national or international***

police DNA database, for which there is no investigative need. By choosing not to store the genetic data in a centralised database, many concerns ordinarily related to storing genetic data, such as confidentiality and data protection, are minimised (Samuel and Prainsack 2019).

3. One particular argument in favour of the use of FDP in the criminal justice system draws analogies between the technology, and traditional eyewitness statements. FDP findings - just like an eyewitness, so the argument goes - provide important information about what the perpetrator of a crime looks like. This analogy fails to accommodate potentially relevant differences between the two: it neglects the trust people place in DNA information and the effect this may have on how they interpret FDP outcomes even if they are told that such outcomes are only predictions. This could especially be the case if the FDP predictions support any possible deep-seated biases.[23] In addition, whilst FDP provides de-contextualised statistical information about a person's most likely appearance, eyewitness accounts often provide important context i.e. dense information about clothing, hairstyle, size, behaviour, etc. that DNA does not provide (for a more detailed review see (Samuel and Prainsack 2019)). Because of this, we recommend that ***law enforcement officers do not treat FDP findings analogous to the testimony of an eyewitness (i.e., a 'biological eyewitness') or any other kind of evidence,*** but rather view the technology as a tool in its own right, which needs to be assessed against independent scientific criteria as described in section 3.2.2; recommendation 3.

4. Given the predictive nature of FDP, alongside the concern by some scholars that forensic DNA evidence is imaged as being a 'truth machine' in the sense that it enjoys a very high level of public trust (Lynch 2008), there is a concern that FDP findings could be misinterpreted in such a way that too much faith is placed in the predictions supplied by the FDP output. This is particularly the case if the FDP findings support deep-seated biases. If FDP findings are not correctly interpreted and appropriately used in a criminal investigation, the potential for following unnecessary false leads and/or aiding discriminatory practices (particularly for BGA prediction testing) increases. The concern is that if a prediction reflects a law enforcement officer's unconscious bias the law enforcement officer may be quicker to accept the finding than if the opposite was the case. Such (even unintentional) racial profiling is a problem broader than FDP, but it is important that FDP does not add to the issue and that FDP is used in a way which minimises harm, respects human dignity, and is socially and culturally sensitive. Training for law enforcement officers, prosecutors and Magistrates (or

[23] While DNA evidence is more likely to be given weight if the findings support deep seated biases, the same cannot be said for the opposite, i.e., DNA evidence may be rejected as untrue if it does not support deep seated bias. This happened in the Maria Vaatstra case in the Netherlands. In this case BGA prediction testing revealed the suspected perpetrator's likely European ancestry, but this finding was not believed by many individuals who believed that the suspected perpetrator was an Asylum seeker.

holders of analogous functions) is specifically required for them to understand DNA evidence, and the statistics involved in the analysis. For example, previous research has explored how the format of DNA findings can influence interpretation (for example, see (Lindsey, Hertwig, and Gigerenzer 2003, Nance and Morris 2005, Cashman and Henning 2012)). We recommend ***context-specific education and training in interpreting FDP findings for forensic DNA scientists producing FDP data, forensic DNA scientists (reporting officer) interpreting FDP outcomes, and law enforcement officers ordering FDP tests***, recognising the different training needs for different types of practitioners. Education can reduce the risk of negative effects on human dignity when the technology is used in unduly discriminatory ways (i.e. on the basis of racist or religious stereotypes), or when it unduly restricts people's freedom.

We further recommend that ***the FDP written report only be shared with a law enforcement officer who has been trained in understanding such findings.*** Training efforts directed at forensic DNA scientists need to include guidelines on the interpretation of VISAGE prototype software output data (and FDP findings in general) and the best format for communicating the results to law enforcement. Scudder and colleagues' (2018) suggest to '*establish...a group of senior police officers and forensic [DNA] scientists, similar to an ethics board, to authorise release of less reliable or privacy intrusive predictive information on a case by case basis'* (Scudder et al. 2018: 227). Such a trained individual or group of individuals should also be consulted by a Magistrate or prosecutor if and when they are required. It is vital that Magistrates as well as law enforcement officers are able to understand FDP findings, and be able to put them in the correct context.

5. Given the uncertainty attached to the predictive nature of FDP findings, we recommend that ***FDP findings should not normally be the sole basis for law enforcement decision-making,*** but rather, other corroborating evidence should be required to identify or exclude suspects or groups. For example, such other evidence could narrow down the broadness of the FDP finding (e.g. Eastern European descent) to the specific (e.g. suspected perpetrator is likely Turkish because the crime was committed in a Turkish neighbourhood). In instances where no other evidence is available, FDP findings should only be used as the sole basis for law enforcement decision-making after careful consideration of the uncertainty of FDP findings, which should be viewed as predictive inferences only.

   In certain jurisdictions, for example in the Netherlands, one particular application of FDP findings and their corroborating evidence is to implement DNA dragnets. During dragnets (also known as intelligence-led mass screenings), a specific group of the population - defined according to criteria that are believed to apply to the perpetrator, such as sex, age, place of

residence, or ethnicity - is asked to "voluntarily" submit a DNA sample for testing to rule them out as a suspect (Lipphardt et al. 2017, M'Charek, Toom, and Prainsack 2012). In these instances, in which law enforcement officers are interacting very closely with the community, we recommend that ***law enforcement officers should communicate FDP findings to those involved in the dragnet in an open and transparent way and invite dialogue rather than merely passing on "facts". For example, the meaning of an FDP test result in a specific community could be different from how law enforcement actors envisage it, and it could affect communities in unintended ways.***

6. The utility of FDP lies in the fact that it can provide key intelligence to law enforcement officers as they search for a suspected criminal perpetrator. In spite of this potential utility, FDP can only disclose probabilities, and, in addition, it provides information on some phenotypic traits that are particularly socially sensitive. Because of this, to respect the rights of non-discrimination and human dignity, and to minimise potential harms from the technology, we recommend that ***law enforcement officers should not communicate FDP findings to the public via public media unless (a) there is substantial evidence to corroborate them, or (b) when communicating the findings to the public is necessary for the measure to succeed (e.g. voluntary DNA dragnet).*** In the event that FDP findings are communicated to the public, we recommend law enforcement officers state publicly that the findings are based on probabilistic information from DNA and carefully consider how they phrase the findings to minimise the risk of stigma to minorities and other groups. In no instance should images of the likely appearance of perpetrators based on FDP evidence be disseminated to the public.

3.2.4 Recommendations related to other actors' interpretation of FDP findings

These recommendations relate to the need to educate actors in the criminal justice system about FDP. They aim to strengthen due process in the legal system, ensure FDP analysis is only requested, conducted and used when reasonable, and in a way which guarantees proportionality of the means (in terms of balancing invasions of privacy/discrimination), and ensures the state does not profit from the violation of the due process in order to achieve a conviction.

1. Magistrates and prosecutors need to be educated about how to understand and interpret FDP findings to ensure FDP analysis is only requested in situations in which it is deemed reasonable (in terms of proportionality) and lawful, and such that any FDP findings are interpreted correctly. We therefore recommend that ***the education strategies set out in***

recommendation 3.2.3 (recommendation 4) should also apply to prosecutors and Magistrates.

2. During the proceedings of a criminal case it is vital to determine whether evidence is permissible in court. If an (early) FDP analysis has been conducted illegally (misuse, without respecting, for example the proportionality of the mean, or if the findings have been misinterpreted, such as motivated by racial bias), this may have implications in terms of admissible evidence in Court. Defence lawyers should be empowered to seek for an exclusion of FDP evidence if it was obtained without due process; even if the suspected perpetrator has been apprehended and their identity verified. This will also have the effect of ensuring law enforcement and prosecutors maintain due process regarding the handling of FDP. We therefore recommend that ***the education strategies set out in recommendation 3.2.3 (recommendation 4) should also apply to defence lawyers.***

3.2.5 Transparent evaluation of FDP use in the criminal justice system

As we have stated previously, it is of key importance to track the effects of FDP, both positively and negatively, because while criminal justice, as a whole, protects and fosters some public rights (e.g. security), it can also conflict with other public rights (e.g. privacy); and the contribution of FDP in protecting public rights is unknown. Robust evaluation research must show that the effect of FDP in case work is beneficial - in the sense that it increases detection rates, reduces cost and/or investigation time, saves public resources, and does not lead to undue harms for people or groups such as those related to discrimination or bias. Any evaluation process must be explained in a way that can be understood by stakeholders and the public (Samuel and Prainsack 2019). Where possible, civil society representatives should be involved in oversight.

Evaluation of FDP should be integrated throughout the innovation lifecycle, helping to strengthen quality through cycles of insight building and iteration, as well as informing difficult decisions about disinvestment.[24] The innovation lifecycle for FDP includes the research process (does the technology work, is it validated?); FDP use by law enforcement (how is the technology used in practice?); and societal impact (what is the impact of FDP i.e., has it achieved what it was designed to achieve, does it have unintended consequences?). Evaluation is critical to minimise the risk that FDP unduly interferes with the right to privacy,

[24] Caludia Pagliari stated this in an interview at the WHO conference on the future of digital health (2019) https://www.ictandhealth.com/news/the-risks-of-basing-digital-health-strategy-on-industry-hype-and-alluring-prototypes/.

non-discrimination and respect for dignity, and that, given these risks, it has a positive benefit on society. With regard to the use of FDP, it will be of utmost importance and we recommend fully that:

1. The development and, where legally and ethically permissible, the deployment of FDP is embedded in robust and transparent research, so that legal provisions and protocols in practice can be adjusted to maximise benefits and eliminate harms. If FDP is to protect and foster public rights then it is of great importance that research on *all* aspects of FDP is submitted to expert scrutiny via publication. We recommend that ***all FDP research outcomes that provide the basis for FDP tools applied in practice are scientifically, technically and forensically validated, and published in peer-reviewed scientific journals (including open access to algorithms and workflows) such that it is transparent and open to scrutiny by experts*** (section 3.2.2; recommendation 3).

2. To minimise the risk that FDP unduly interferes with the right to privacy, public authorities need to be transparent about its use in casework and evaluate this use. We recommend that the outcomes of ***individual instances of FDP use by law enforcement officers should be fed back to the forensic scientists developing the technology*** so that they can monitor their prediction software and adjust as and when appropriate. We also recommend that ***the use and outcome of FDP by law enforcement authorities should be recorded and made available to an oversight body for evaluation, if requested.*** This record should include information on the context for which FDP was used (when during the case, which tests etc), the effects of FDP, both in terms of contribution to the solution of serious crimes, as well as possible human rights infringements or other forms of discrimination (see Table 1). Given that FDP should only be used for serious crimes, and will be appropriate in only some of these cases, such record-keeping seems feasible.

Record-keeping
What is the context of the criminal case (what type of crime)?
At what point during the case was FDP testing requested and why was it considered appropriate?
How, if at all, were the FDP findings incorporated into the case? (including whether the FDP finding resulted in surveillance/screening of suspects, and if so to what extent and to what effect?)
Is the case still ongoing or closed?
Was the information provided by FDP helpful and/or successful in solving the particular case? If so, which particular information and why? If not, why not?

Table 1. Example of the types of data which should be collected during FDP record-keeping.

3. To ensure the use of FDP is effective, and to promote accountability, we recommend that ***legislators and regulators regularly evaluate the effects and use of FDP.*** We recognise that evaluation (choosing what to evaluate, selecting design and targets, to dissemination) will always be a valuing process in terms of whose values are operationalised (Oliver, Lorenc, and Tinkler 2019). A controlled trial is not a feasible tool for FDP evaluation given the ethical issues regarding who could be considered a control group, the practical issues that FDP will only be indicated for a small number of criminal cases, and the political issues in terms of who would administer the trial. We therefore recommend that ***documentation on FDP use, which has been recorded by law enforcement authorities (see recommendation 2 and table 1 above), should be evaluated via a democratically legitimised national oversight body; and that the evaluation should be retrospective and observational***. To improve the evaluation procedure, national evaluation findings should also be collated at a European level. This will involve some investment in terms of appointing an accountable body to conduct the evaluation. It is in the public interest to do this since it is vital that we, as a society, determine whether the funding invested into this technology has a positive impact in terms of increasing the number of crimes solved or decreasing the time to solving crimes, or a negative impact in terms of increasing discriminatory practices or unnecessarily leading to an invasion of privacy.

   We recommend that evaluation and oversight mechanisms should include civil society representatives where possible. We also recommend that the evaluation process is transparent via public annual reporting (similar to, for example, the annual reports of the United Kingdom (UK) National DNA Strategy Board which keeps a record of the number of familial searches which have been performed annually[25]). The evaluation outcome must be used to inform future use of the technology. Specifically, based on the evaluation, where necessary, adjustments should be made that address concerns regarding human rights infringements and other unintended consequences, and (only insofar as this does not conflict with the former point) enhance the utility of FDP for the protection of public rights. Similarly, if the use of FDP shows little public benefit compared to resources invested, we recommend that the possibility of disinvestment is considered (in this case, phenotypic tests for each trait will need to be assessed independently). Evidence of negative impacts of FDP could lead to strong emotional reactions amongst researchers and policymakers who are personally invested in the technology. This is important to consider when thinking about disinvestment, but it is imperative to ensure that no pressure exists to simply ignore inconvenient evaluation findings (Oliver, Lorenc, and Tinkler 2019). We note that whilst evaluation findings should inform the future use of FDP, a decision will need to be made about how to strike a balance

[25] https://www.gov.uk/government/groups/national-dna-database-strategy-board.

between any potential costs and benefits of the technology, and whether this balance should be crime dependent.

The above recommendations integrate evaluation into the innovation lifecycle, and can be represented as shown in Figure 2. This figure has been adapted from Claudia Pagliari's conceptualisation of the evaluation of digital health technologies.[26]

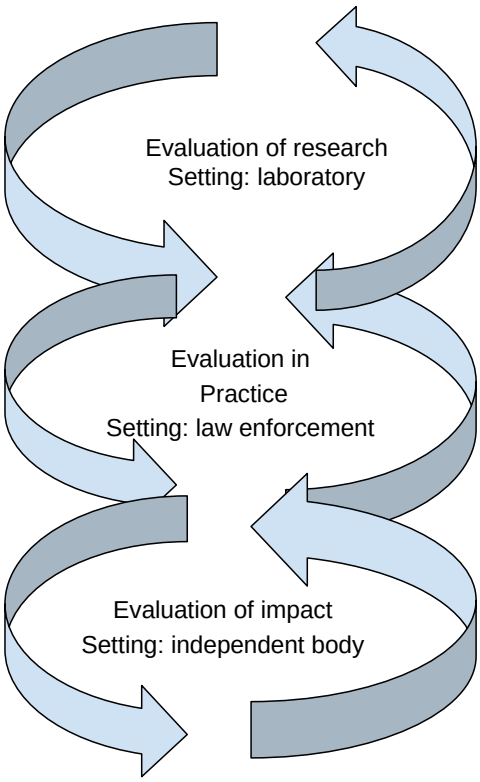


Figure 2. Diagrammatic representation of our recommendations for FDP evaluation in practice.

[26] This diagram was presented in a talk at the "Future of digital health systems" February 2019.

3.3 Country specific recommendations

This section considers the challenges and solutions related to FDP use in the criminal justice system that fit the particular regulatory, political, cultural, and economic configurations of the national setting for each of the VISAGE represented countries. The recommendations we make in this section apply only if countries have made the decision to permit FDP in their criminal justice system.

Below, we summarise the current regulatory landscape pertaining to FDP use in the criminal justice system for the VISAGE member countries, and discuss how our recommendations could be configured to align with each of these countries' policies and practices. We have paid specific attention to the following elements, which are the specific questions that each country needs to clarify if they legalise the use of FDP in their criminal justice system:

- Under which circumstances should FDP be considered?
- How is a 'serious crime' defined in each country?
- Which governing body could be responsible for deciding which criminal cases, and at which point during a case, FDP should be used?
- Which governing body or oversight organisation could be responsible for evaluation?

3.3.1 Austria

Since the reform of the Austrian *Sicherheitspolizeigesetz* (SPG, Security Police Act) in 2018 (to make it compliant with the new EU General Data Protection Regulation and EU Data Protection Directive for Law Enforcement Authorities), FDP is considered permissible in Austria. The SPG emphasises that any use of FDP must be proportional. This is interpreted by experts to mean that FDP can only be used for extremely serious crimes (murder cases, serial sexual crimes) punishable from a criminal court with a minimum one years' imprisonment. FDP should only be used as a last resort: all other investigative methods must be exhausted, and it must be perceived that FDP could potentially provide a meaningful restriction of suspected perpetrators for further investigation. Only DNA markers suitable for the purpose of identification can be used to infer likely characteristics of perpetrators, i.e., not more than the traits that could be inferred from an eyewitness. Health-related dispositions, or markers pertaining to personality traits, cannot be used for this purpose. FDP must only be used under the authority of the national central criminal police authority in the Ministry of Interior.

The Austrian context for the use of FDP in the criminal justice system is in line with our best practice recommendations. FDP should only be permitted for externally visible characteristics, only for serious crimes, only in exceptional circumstances, and only under the authority of the national central criminal police authority. We endorse the current situation in Austria, whereby the national central criminal police authority is aware of all cases in which FDP is used and the results are evaluated for their criminal relevant suitability. This data should be evaluated on an annual basis as described in section 3.2.5, and reported on their website.

3.3.2 France

As noted in our previous Report,[27] it is not clear to what extent French legislation on forensic DNA testing restricts or permits the use of FDP. However, a recent 2014 High Court case provides the legal basis to perform FDP in criminal cases for morphological characteristics (external physical appearance traits), including (but not necessarily limited to) hair, eye and skin colour. The legal basis for ancestry inference is still ambiguous with some experts taking the view it is illegal, and others arguing that it is legal. It seems that this is based on how 'morphological characteristics' are being defined. The court decision of 2014 leaves FDP in legally ambiguous territory. At the moment, policy discussions are underway to try and resolve the situation. We recommend urgent clarification as to the legal basis of FDP in France. If they rule in favour of permitting FDP, based on our best practice recommendations we recommend that:

- Only appearance, age and BGA should be permitted for prediction testing, and only via tests which have been validated as described in section 3.2.2; recommendation 3.
- FDP should only be permitted for serious crimes as defined by the French penal code ("crimes" in France), which include, for example, murder, intentional homicide, premeditated homicide, and rape.
- FDP should only be used in exceptional circumstances, as defined in section 3.2.2 (recommendation 5). This is in-line with the National Institute of Scientific Police (*Institut National De Police Scientifique*), which only uses FDP for analysis of unknown origin DNA samples when there is no matching profile on the national DNA database.

[27] http://www.visage-h2020.eu/PDF/Deliverable_5.1_for_posting_online_DECEMBER_2018.pdf.

- There is a built-in requirement for FDP use to be evaluated on a national level as described in section 3.2.5 above. We recommend that a suitable oversight body be given the responsibility to oversee this.

3.3.3 Germany

Since the recent reform of the federal Criminal Procedures Act (*Strafprozessordnung*, StPO), FDP is allowed in Germany for skin, eye, and hair colour predictions, as well as age inference.[28] BGA is excluded from the traits that can be tested for. (In Bavaria, BGA can be used in addition to avert imminent danger, but not to solve crimes that have already occurred.) The German context for the use of FDP in the criminal justice system is in-line with our best practice recommendations. There are, however, a number of differences. First, German legislation does not put any restrictions on the type of crime for which FDP can be used, whereas we recommend that FDP should only be used in serious crimes and in exceptional circumstances. Second, German federal legislation restricts the use of FDP to only hair, eye and skin colour, as well as age. We also recommend the inclusion of BGA prediction testing. Such testing can increase the predictive value of other appearance traits.

Alongside the German legislation, we recommend that FDP use be evaluated on a national level as described in section 3.2.5. We recommend that a suitable oversight body be given the responsibility to oversee this.

3.3.4 Poland

In Poland, there seems to be no explicit regulation relating to which forensic DNA tests can be performed. The wording of the Police Act (the Act in Poland that contains forensic DNA regulation[29]) implies that only data from non-coding regions of the genome can be collected, processed and analysed. Having said this, a number of central actors explain that the context of this Act is that it was written entirely in terms of storing DNA profiles in a national database (to be in accordance with Prüm[30]), and therefore does not apply to how forensic DNA analyses are performed. In line with this interpretation, in Poland, FDP is used without restriction, though it is mainly only used for more serious crimes. The Polish context for the

[28] https://de.wikipedia.org/wiki/DNA-Phänotypisierung.
[29] There are also regulations concerning the analysis of biological material in the Polish Code of Criminal Procedure.
[30] The Prüm framework - whose core legislation is Council Decision 2008/616/JHA - requires that EU member countries make available the reference data from their national DNA analysis files to competent and authorised authorities in other member countries, but only those DNA profiles based on the non-coding part of DNA.

use of FDP in the criminal justice system is in-line with our best practice recommendations in that it is restricted to predictions pertaining to externally visible characteristics (appearance, age and BGA); it is only used for serious crimes, which is a prohibited act punishable by imprisonment for not less than 3 years; and is only used in exceptional circumstances. In Poland, when an FDP test is requested, typically either a public prosecutor or the police (without a request from the prosecutor) will ask a laboratory to do the testing. We recommend that this decision be made by the prosecutor. We also recommend that FDP use be evaluated on a national level by an authority authorised to conduct preparatory proceedings in criminal investigations, such as the prosecutor's office.

3.3.5 Spain

Beyond the legal basis of the Spanish DNA database there is no explicit regulation related to the use of FDP in Spain, nor any forensic DNA analysis. As such, FDP is not prohibited. FDP is practiced for BGA, for pigmentation testing (eye, hair and skin colour), and recently also for age prediction. A statutory body exists (*Comisión Nacional para el Uso Forense del ADN* (CNUFADN) - National Commission for the Forensic application of DNA), entrusted with reviewing and regulating all aspects related to the national forensic DNA database.

Policy discussions are still on-going in Spain regarding the need for explicit legislation around the use of FDP in the criminal justice system. The CNUFADN have, however, recently approved recommendations on the use of FDP in the criminal justice system, which endorse the use of internationally approved phenotypic and ancestry predictive markers. The recommendations state the following:

- Each laboratory should perform internal validation studies prior to the use of FDP markers. They also recommend accreditation under the current legislative ISO 17025 standard for genetic markers routinely used in criminal investigation.
- Specific proficiency exercises for using FDP should be developed to standardise and check analyses, and to allow for inter-comparison of results between laboratories, similar to the recent *Grupo de Habla Española y Portuguesa de la ISFG* (GHEP-ISFG; Spanish and Portuguese Speaking Working Group of the International Society for Forensic Genetics) exercise.[31]
- FDP should only be used when other investigative routes have been exhausted (i.e., no DNA match in the national DNA database) and in a way which minimises the intrusion of human rights.

[31] https://ghep-isfg.org/wp-content/uploads/medios/comision_GHEP-MPS_Praga2019.pdf.

- FDP should not be used as conclusive proof of identification.
- FDP should only be used in serious crimes for sentences over five years in prison.
- FDP should only be used with express authorisation (judicial and from the prosecution).

These recommendations are in-line with our best practice recommendations in that FDP is recommended to be restricted to predictions pertaining to externally visible characteristics (appearance, age and BGA); only for serious crimes (defined as "a serious crime against life, integrity of persons, freedom, sexual liberty or indemnity, terrorism, or any other serious crime that carries a serious risk to life, health or the physical integrity of people" as per Article 129 bis of the Criminal Code (*Ley Orgánica* 10/1995, *de 23 de noviembre, del Código Penal*); only in exceptional circumstances; and only with express judicial or prosecution authorisation. We also recommend that FDP use be evaluated by the CNUFADN in line with Section 3.2.5.

3.3.6 Sweden

In Sweden, the use of FDP is permitted. The legal documents explicitly governing FDP are the Police Data Act and its accompanying materials. The Swedish Police Authority interpret these legal texts as meaning that they are free to apply a broadened use of DNA analysis methods to crime scene traces (from unknown individuals), including age, appearance and BGA inference, in order to generate forensic intelligence leads, with no set limitation of the application. FDP must only be applied in the investigation of serious crimes, however, and only if it is deemed to be an absolute necessity to solving the crime; and its use must be proportional. FDP must only be used to generate investigative leads, meaning that the results must not be stored in databases. Despite this permissive legal situation, FDP testing is not currently performed in Sweden. The Swedish National Forensic Centre plans to start performing FDP analyses in-house during 2020. Until this is achieved, the Swedish National Forensic Centre has been sending samples abroad for FDP analysis. This occurs only rarely and for high profile cases. Eye colour, hair colour and BGA are all tested. The decision to use FDP in a criminal investigation comes from the person in charge of the case, who can be either a police officer of a prosecutor, depending on the stage of the investigation; no Magistrate is needed.

The Swedish context for the use of FDP in the criminal justice system is in-line with our best practice recommendations in that it is restricted to predictions pertaining to externally visible characteristics (appearance, age and BGA), it is only used for serious crimes, and in

exceptional circumstances. We also recommend that FDP use be evaluated on a national level by a relevant body, which would need to be decided upon at a national level.

3.3.7 The Netherlands

The Netherlands has explicit legislation, the Dutch Code of Criminal Procedure, which permits FDP for appearance and BGA inference in the criminal justice system.

FDP for BGA, eye colour and hair colour predictions are currently permitted under the Code of Criminal Procedure. Additional EVCs[32] can be added to this legislation by Royal Decree once it is demonstrated by scientific research and technological development that an EVC can be predicted from DNA, and a DNA test has been produced and forensically validated. Skin colour is likely to be the next EVC to become approved, the legal procedure is currently underway. FDP for age inference is not explicitly permitted at this time. Primarily, the Netherlands Forensics Institute (NFI; an agency of the Ministry of Justice and Safety) and the Forensic Laboratory for DNA Research (FLDO) at the Leiden University Medical Centre, which also acts as an official "second opinion" laboratory for the Ministry of Justice and Safety, are the institutions which conduct FDP testing for the Dutch police. Laboratories performing an FDP analysis are overseen (with certification and documentation) by the Dutch Accreditation Council to ensure scientific quality, professionality, and reproducibility, and compliance in terms of data protection and storage. In terms of FDP use, FDP is permitted only for serious crimes, which are defined as those which attract a four-year prison, and only in exceptional circumstances.

Policy and practice related to the use of FDP in the Netherlands criminal justice system is in-line with our above recommendations for the implementation and use of FDP in the criminal justice. Though some open questions remain. Specifically:

- One open question relates to whether age should be considered an EVC. If age is considered an EVC, age prediction testing could be legally permitted via a Royal Decree once it is demonstrated by scientific research and technological development that it can be predicted from DNA, and a DNA test has been produced and forensically validated.[33] There are different opinions about this. For example, whilst the Minister of Justice stated that FDP can only be conducted for EVCs visible "from

[32] We note here that in the Dutch legislation, appearance inference is categorised as an EVC; BGA inference is not. It is not clear whether age prediction inference is considered an EVC.
[33] This would require methylation DNA testing (the testing used to infer biological age) to be categorised as a DNA test.

birth" – a statement which would exclude age inference testing from being conducted - this concept is not adhered to in the strictest of sense since hair and eye colour, which have been legally allowed already, can change from birth. We recommend, that the legal permissibility of age prediction testing should be clarified in the Netherlands.

- There is currently no built-in requirement for FDP use to be evaluated at a national level. We recommend that a suitable oversight body in the Netherlands take on this role in line with our recommendations as described in section 3.2.5.

3.3.8 United Kingdom

In the UK, there is no explicit legislation governing which techniques can be used for forensic DNA analyses for crime scene DNA stains. The legislation that exists governs only the collection, processing and storage of DNA for forensic purposes. As such, according to the letter of the law, FDP is permitted. As is customary in the UK, the use of FDP is governed by various regulatory oversight bodies. The National DNA database (NDNAD) Strategy board is the overarching governance committee.[34] The Strategy Board comprises representatives of the National Police Chief's Council, the Home Office, the Biometrics and Forensics Ethics Group, the Association of Police and Crime Commissioners, the Forensic Science Regulator (or her representative), the Information Commissioner's Office, the Biometrics Commissioner (or his representative), representatives from the police and devolved administrations of Scotland and Northern Ireland, and such other members who may be invited. The task of the Board is to provide governance and oversight for the operation of the NDNAD. As yet these regulatory bodies have engaged little with the issue of FDP, and no recommendations, guidelines or regulations regarding the technology have been developed.

At present, FDP can be requested by police and forensic providers at their discretion. No formal requests need to be made to the courts. Any use by the police would require strict oversight: accreditation of any forensic laboratory process has to be received from UKAS (the national accreditation body for the United Kingdom), and any use of FDP (or any other forensic technique) has to first be negotiated and discussed with UKAS. However, as far as we know, FDP is used rarely, and at present, not by the police. Private companies do have the technology but are not presently using it. We recommend that FDP use in the UK should only be permitted for appearance, age and BGA predictions, only for serious crimes (Schedule 1 of the Serious Crime Act 2007 states serious crimes uploaded to the NDNAD include

[34] https://www.gov.uk/government/groups/national-dna-database-strategy-board

(attempted) murder, manslaughter, other suspicious death, rape or other sexual offences, serious robbery, terrorism, explosives, high value fraud, blackmail, abduction/kidnapping, arson, or grievous bodily harm/wounding), and in exceptional circumstances. We also recommend that FDP use is evaluated at a national level by the NDNAD Strategy board.[35]

[35] Other scholars have called for the Strategy Board to have a similar role in terms of evaluating the UK national DNA database (Amankwaa and McCartney 2019).

3.4 Recommendations for unidentified person cases

FDP could play an important role in the identification of human remains. An FDP indication of the characteristics an unidentified body had in life could assist with both directed investigations, or in database-driven cold case matching. In such cases, there is much less of a concern about stigmatisation or discrimination, as the individual subject to FDP is not normally suspected of being a criminal. Rather, it is generally in all best interests, including the relatives of the missing, to identify a missing person's remains. While some concerns remain, such as those associated with the over-interpretation of findings that could mislead investigations or tend toward conclusions beyond the strength of the evidence, many of the recommendations applied to the use of FDP in criminal cases, aimed to provide safeguards around the use of this technology, do not apply to such unidentified person cases. Moreover, unidentified person investigations are often governed under different laws and regulations, and depending on the jurisdiction may not be under the regulation of a Penal Code, meaning that investigations take place outside of law enforcement. It may be useful for the FDP profile of unidentified human remains to be placed in searchable databases indefinitely, until the case is solved.

Given the above, it seems appropriate that the use of FDP in unidentified person cases is less strictly regulated. We recommend the following:

- Only phenotypic tests that can predict appearance, biogeographical ancestry (BGA), and age (i.e., externally visible characteristics) from DNA should be permitted for FDP use (see section 3.2.2, recommendation 2).
- FDP should only be permitted for those phenotypic traits whose predictive marker and prediction models have been validated as per section 3.2.2, recommendation 3.
- FDP should only be permitted for those phenotypic DNA tests that have been technically and forensically validated as per section 3.2.2, recommendation 3.
- Context specific education and training in interpreting FDP findings is required for all law enforcement officers and/or other officers who handle unidentified person cases.
- The outcome/interpretation of an FDP analysis is communicated very carefully to the officer who requested the analysis (and not reported to anyone outside of law enforcement or requesting authority) as described in section 3.2.1 recommendation 4.

4. Best practice examples

4.1 Written report of FDP findings

The table below highlights different best practice examples of reporting FDP findings from across various forensic laboratories in VISAGE represented countries. The format of each of these reporting styles is different. This is because the reporting of FDP findings is contingent on the features, systems and infrastructures of each jurisdiction's criminal justice system. As such, there is no one-size-fits-all approach. There are, however, a number of common themes which we can distil from these reporting practices.

1. Try to avoid statistical calculations and models in the reporting document. Statistical calculations can be difficult for non-specialists to understand and could potentially lead to misinterpretation of the findings.
2. Where decision trees are available to assist with reporting, these should be used.
3. Written policy – either in the form of legislation, or as national oversight body guidance - is the best way to ensure consistent and appropriate best practice across different institutions and time.
4. Consider omitting genotype data from the FDP report given to law enforcement officers. This data is not relevant to the interpretation of the findings nor the forensic investigation. This also supports the principle of data minimisation and ultimately genetic confidentiality.
5. Consider the inclusion of maps and/or images in the final report by weighing up the likely benefits (ease of understanding visual information) against the risks (increased risk of misinterpretation, or simplifying the findings from those which are predicted to those which are certain).
6. State clearly that FDP findings are predictive only.

Country	Reporting style
France (INPS)	Report conclusions contain no statistical analysis. BGA results are presented in tables/plot (principal component analysis); appearance conclusions are presented with illustrations. In the report, it is emphasised that: 1/ differences can be observed between genotype data and the phenotype of the DNA contributor. The appearance can be altered voluntarily (e.g. hair colouring, coloured contact lenses, wig wearing) or involuntarily (e.g. hair colour darkening by aging, health condition). 2/ the results are probabilistic inferences (they do not predict the traits with certainty). The accuracy of each method used is given. 3/ FDP is not an identification method. It is intelligence information and that the potential identification of the suspect has to be carried out with conventional STR typing. After sending the report, we try as much as possible to contact the officer/Magistrate to ensure that results are not misinterpreted and explain these results. It is often BGA results that raise more questions. At the moment, statistics are excluded from the report. They can be provided upon request.
Sweden (Swedish National Forensic Centre)	Reports (*forensiska uppslag*, meaning "investigative leads") are clearly marked as "intended for intelligence purposes". The focus is on narrative descriptions in the following form: "Given the results, the person is considered (*bedöma* i.e., "judged or estimated") to have brown eyes and brown to black hair". Statistical probabilities, figures or maps are not to be presented in the initial report but may be given or presented upon request. We always encourage personal communication with the client regarding our investigative lead reports (irrespective of if DNA or any other discipline) to discuss the outcome and what it means. In many cases we will make a call prior to or after sending the report.
Poland (Central Forensic Laboratory of the Police)	Information about the technology and kit used is provided (Illumina FGx or self-developed NextDNA Intelligence Tool - FDP and BGA Ion Torrent technology panel). The report includes investigated markers: • FGx: Autosomal, Y and X STRs and iSNPs. • NextDNA Intelligence Tool: FDP and BGA SNPs. FDP traits are reported with percentage values. Due to a lack of global recommendations and guidelines we use decision trees from (Walsh et al. 2011, Walsh et al. 2013) to inform our statistical analysis of this information. Additionally, a photo of predicted eye and hair colour is printed on a dye sublimation printer (DS-RX1 DNP) to give a good picture of the predicted trait. BGA is explained in a brief conclusion. No multidimensional scaling or graphs are included.

The Netherlands (NFI: Netherlands Forensic Institute)	The methodology is explained in brief and the conclusions with no statistical analysis, illustrations or maps are given. The most probable eye- or hair colour and ancestry is indicated in bold. Scientific background literature and analysis websites are given in footnotes.
The Netherlands (Manfred Kayser, Erasmus MC for international police clients: budget regulations in the Netherlands practically prevent others than NFI to do casework for Dutch police)	Bi-parental BGA using autosomal DNA markers is reported via plots from multidimensional scaling analysis of identity-by-state distances calculated between pairs of individuals from the genetic data showing the case sample together with all samples from the reference dataset used, with narrative explanation and a conclusion statement on the most likely geographic region(s) of bi-parental ancestry. Paternal BGA with high-resolution Y-SNP analysis is reported as Y haplogroup, with tables and frequency maps from the literature on geographic distributions of the observed Y haplogroup, and narrative explanation and a conclusion statement on the most likely geographic region(s) of paternal ancestry. Maternal BGA with high-resolution mtDNA analysis is reported as mtDNA haplogroup, with tables and frequency maps from the literature on geographic distributions of the observed mtDNA haplogroup, and narrative explanation and a conclusion statement on the most likely geographic region(s) of maternal ancestry. Eye, hair and skin colour is reported using categorical probabilities provided in a table, with narrative explanation and a conclusion statement on the most likely eye, hair, and skin colour categories. A brief introduction and detailed methodology are included in all reporting, including genotyping and statistical methods, reference data and prediction models are included, a list of references, and a final conclusion statement to narratively summarise the most likely BGA and appearance information obtained from the tested sample. No actual genotype data are included in the report.
Austria (Medical University Innsbruck)	Results of forensic DNA testing, including FDP, are reported in two documents, the Laboratory Report (LR) and the Expert Opinion (EO). The LR provides some basic introduction into the scientific background, markers and methods used; describes the workflow; and lists the obtained results (haplotypes and genotypes in tables). The EO describes the details of the case (sample information, chain of custody) and interprets the findings stated in the LR using narrative explanation and conclusion statements. Autosomal BGA prediction is reported based on two-dimensional Principal Component Analysis plots and Cluster Analysis. The plots are not shown but narrative is used to explain the findings. Paternal (Y) and maternal (mitochondrial) BGA predictions are reported by the distribution of the

	estimated haplogroups. No maps are shown but narrative is used to explain the findings. Appearance prediction is reported using the estimated probabilities (shown in the EO) for the respective traits and using the interpretation keys provided in the literature. Photographs are not shown but narrative is used to describe the findings.
Spain (*Universidade de Santiago de Compostela*)	Likelihood ratios are reported. No illustrations are used for appearance predictions (although these appear in Snipper for skin tone, no reference is given to these). No maps are used for BGA prediction, but two-dimensional Principal component analysis (PCA) plots and the optimal (K-value) STRUCTURE cluster diagrams are given with averaged reference population columns and individual casework sample columns. Multiple populations from the same continental region are merged and no distinction is made between them, even though this is an increasing trend in genomic databases (e.g. gnomAD lists Finnish and non-Finnish Europeans separately). Reduced population comparisons are sometimes made to clarify patterns. Admixed American population samples are not used as reference samples unless they have been shown to be completely free from signals of 'non-American' co-ancestry (reduces 1000 genomes to 18 Peruvian samples). Actively considering/trialling use of all available genetic data (i.e. using appearance data) to bolster ancestry likelihoods. **Why?** Appearance: Photographs of skin tone do not adequately represent its visual appearance. Eye and hair colour are categorised into a limited number of classes and intermediate pigmentation patterns can create a 'conceptual disjoint' between an illustration of a prediction class and the observed patterns. Ancestry: Frequency distribution heatmaps are an amalgam of several variables (loci) so are not applicable to autosomal ancestry informative markers; they misrepresent a tighter than actual distribution as observers overlook cooler heat colours across large regional distributions; we do not have the software to produce them and are sceptical about those made by the Thermo Fisher Precision ID module. Principal component plots are visually intuitive but 'flatten' the true distribution of clusters into a two-dimensional plane. Non specialists have been positive about the use of a black casework point on a coloured cluster of reference population points – but overlapping clusters can create problems as they can exaggerate the possibility of classification error. PCA plots only analyse the main (principal) component of variation, and so are based on portions of genetic data generated. STRUCTURE plots help to underpin low likelihood ratios which jointly signal admixture (low likelihood ratios, mixed cluster membership columns). They are visually intuitive and highlight any lack of divergence between

	reference populations used and the marker's power (e.g. Middle East and European samples lacking a clear distinction of genetic cluster patterns).

4.2 Appropriateness of FDP requests

Best practice case example: written by Ron Rintjema and Jelle Tjalsma, the Netherlands National Police Services Agency

(Ron Rintjema and Jelle Tjalsma are cold case advisors experienced with FDP use in policing)

Key point:

- A forensic DNA scientist who has experience working with crime scene stains, as well as a knowledge of police practices, should always offer advice to a police officer about whether FDP testing is technically possible and/or potentially useful to the investigation prior to FDP testing being conducted.

Our perspectives reflect our roles as two police officers who have used FDP findings in our police work, as well as provided advice to other police officers on the use of FDP testing in a number of criminal cases in the Netherlands – a country in which FDP testing is permitted for hair colour, eye colour and ancestry inference. Based on our observations and experience, police officers are often focused on collecting any evidence they can, and as quickly as possible, to aid in their search for a criminal perpetrator. However, before a decision can be made about whether to incorporate the use of FDP, there are a range of technical questions which need asking: whether a crime scene stain is good enough quality for FDP testing (for example, relating to low volume crime stains), and what the testing can actually tell police in terms of the specific perpetrator (particularly with mixed stains). At the NFI in the Netherlands, before a crime scene stain sample is tested using FDP, the forensic DNA scientists speak to the police officers about such issues. This is not necessarily the case for other laboratories – but we think it should be. Specifically, we think a forensic DNA scientist who has experience with working with crime scene stains, as well as a knowledge of police practices, should always offer advice to a police officer about whether FDP testing is technically possible and/or potentially useful to the investigation prior to FDP testing being conducted.

Best practice case example, written by members of the INPS, Lyon, France

Key points:

- Feasibility and verification studies should be performed prior to FDP analysis. These should be communicated to the requesting law enforcement officer/Magistrate.

Before we conduct FDP, we perform a feasibility study. During this study we communicate with the requesting law enforcement officer/Magistrate, as well as the laboratory which conducted the initial DNA STR-profile, in order to ensure that:

- no match was identified between the DNA sample and any sample in the national DNA database;
- the request for FDP relates to a serious crime;
- the DNA profile analysis suggests that the DNA sample is not mixed, and originates from only one contributor; and
- the DNA extract is still available in enough quantity to perform the analysis.

Best practice case example: written by Reinhard Schmid, Ministry of the Interior, Austria

Key point:

- Before making a final decision to use FDP, the technical feasibility and potential utility to inform the criminal case should be taken into consideration.

In Austria, FDP is only used as a last resort, and only in very serious cases when all other avenues of investigation have been explored (including checking the DNA profile on the national DNA databases, the Prüm network[36] and any other databases, as well as exploring other biometric and electronic leads). The decision to use FDP is made by the Ministry of the Interior in co-operation with the investigating authorities and the forensic DNA scientists who would be conducting FDP. Through this co-operation we aim to address the question "does it make sense to use FDP?". The investigating authorities provide comprehensive background information and clarification as to how an FDP finding could be useful in reducing a suspect population for a particular crime. The forensic DNA scientists will provide information on whether they have sufficient DNA material of suitable quality to conduct the analysis. Together these considerations will be taken into account, along with any current evidence which has already been gathered, before a final decision to use FDP is made. It is important to note that FDP is just a small part of the parcel when it comes to the investigation, and there are, and are likely to be, very few cases for which FDP would be a suitable option.

[36] See footnote 31.

4.3 Communicating FDP findings

Best practice case example: written by Ron Rintjema and Jelle Tjalsma, the Netherlands National Police Services Agency

Key points:
Steps should be made to ensure police officers interpret FDP findings correctly.

- The FDP written report given to police officers from forensic DNA scientists should be accompanied by a verbal discussion which should be conducted openly and honestly.
- Lines of communication between the scientists and police should remain open so the police can continue to ask questions relating to outstanding queries.
- The amount of information shared between forensic DNA scientists and police must be carefully considered. Police sharing too much case information with scientists may increase the potential for unconscious bias; scientists sharing too little information with police may lead to scientists missing something that could guide police officers in the correct direction.

Our perspectives reflect our roles as two police officers who have used FDP findings in our police work, as well as provided advice to other police officers on the use of FDP testing in a number of criminal cases in the Netherlands. In our view, when FDP findings are reported to police officers, key to this is that police officers interpret the findings correctly, especially because police officers often have a low knowledge of science. Indeed, our first experience with FDP was during the Marianne Vaatstra rape and murder case in the Netherlands, when we requested an ancestry inference test on the crime scene sample, but for which we had little knowledge of the science behind the testing. We went to talk to the scientist who conducted the FDP test face-to-face, and asked them how to interpret the findings. Integral to our understanding of the findings, was always asking "do I understand this correctly?" It was also important that we kept the lines of communication open between the scientist and ourselves so we could continue asking questions relating to any outstanding queries of the relevance of the findings to the criminal case as the case progressed.

We also want to note that while it is important for the scientist to explain their findings openly and honestly, our experience has raised concerns about how much evidence should flow from police officers to the forensic laboratory. Our worry is that if police officers do not give enough information to the scientists, then the scientists might miss something that might guide police officers in the correct direction. However, if police officers give forensic scientists too much information, it might bias their interpretation of their findings. For instance, if a police officer is trying to solve a hideous crime, and the knowledge about this crime is shared intricately with the forensic scientist, the scientist might be too drawn into wanting to solve the case, leading to interpretation biases. We say this reflecting on our own experience of the Milica van Doorn rape and murder in the Netherlands. For example, consider a test which reports the presence of a E1B (Y-chromosome) haplogroup in the crime scene stain, which suggests Eastern European ancestry. If police officers share information that an eyewitness at the crime scene saw a man of apparent Turkish descent cycling away, the forensic scientist might report that there is a good chance the suspected perpetrator may be Turkish. Yet the science does not say this, it only reports the likelihood of Eastern European descent. Overall, there is no ideal way to handle this because people communicate differently, and each case is distinct. Though it is important that both the reporting forensic scientist and the police are aware of this potential for bias.

Best practice case example: written by Ricky Ansell, Johannes Hedman, NFC Sweden

Key points:

Steps should be made to ensure police officers interpret FDP findings correctly.

- Scientists should explain their written FDP report to police officers verbally and encourage police to ask questions about any elements they do not understand. officers to make contact if they have any further questions.
- FDP reports should be marked "intended for intelligence purposes only" so they are not confused with STR profile hit reports.

When communicating FDP findings to the Police, we put the emphasis on verbal statements rather than on figures and graphics such as maps. Figures such as heat maps demand a thorough understanding of the underlying science, which cannot be expected from the receiving police investigator. For that reason, we do not provide figures, maps or statistical probabilities in the initial report. Instead we encourage the client to contact us in person for more information, giving us the opportunity to explain the complexity of graphics and meaning of given probabilities. In many cases we will also make a call prior to or after sending the report, to initiate personal contact. The reports are clearly marked "intended for intelligence purposes only", in order not to confuse them with STR profile hit reports. The narrative descriptions of the outcome of FDP analysis is in the following form: "*Given the results, the person is considered/estimated to have brown eyes and brown to black hair*". As we learn more about how the message is received, we are open to updating the reports and communication accordingly.

Best practice case example: written by Reinhard Schmid, Ministry of the Interior, Austria

Key point:

- Forensic DNA scientists should provide a verbal explanation to accompany their written FDP report for police officers so that police officers clearly understand what the results mean in terms of the criminal investigation.

We receive a very detailed written report from the forensic scientists conducting FDP which explains the meaning of the findings, along with their weaknesses and sources of error. However, what is more important, this is the most important point, is that these findings are described verbally to the investigating officers so that they clearly understand what the results mean in terms of their investigation, and what possibilities could arise if these findings are misinterpreted.

Best practice case example: written by Ron Rintjema and Jelle Tjalsma, the Netherlands National Police Services Agency

Key points:

- When FDP findings inform DNA dragnets, a multidisciplinary group of experts should be brought together (police, ethicists, public engagement specialists etc) to help with decision-making relating to how to engage the public.
- Before police engage with the public, they should be trained on how best to communicate with the suspect population using culturally sensitive approaches.

Our perspectives reflect our experience working on two rape and murder cases in the Netherlands – the Marianna Vasststra and the Milica van Doorn case. In the Milica van Doorn case, ancestry testing, together with other police evidence, suggested the perpetrator was of Turkish descent. A decision was made to conduct a DNA dragnet search in the local area, focusing on 133 men of Turkish descent. A huge effort was put into how to communicate this search to the Turkish population. This effort was based on techniques developed in the Marianne Vaaststra case. Specifically, we invited a range of external experts from, for example, communication, ethics, law and psychology and introduced them to the FDP findings and the other police evidence on the case. We then discussed how best to implement the dragnet search in practice. This bringing together of multidisciplinary experts is not typical practice during criminal investigations. Police officers tend to make decisions within the force and do not share their knowledge with others. But it was very successful and added a layer of ethical and social expertise which sat between the police officers and the public community.

We were very careful with how we proceeded. Before we spoke to the public, specialists in Turkish culture were brought in to teach police officers how best to communicate with the Turkish population using culturally sensitive approaches: Where do we need to be very careful? What can we say and what can't we say? What do we need to take into consideration before we speak to the public? Letters were then written to individuals in the community and talks were held with key figures. Moreover, when going door-to-door asking Turkish individuals to participate in the dragnet search, police officers were given information about DNA, what the results of the tests might be and how to respond to questions from the community. Afterwards, police officers reported that they were happy with the process, noting that people invited them into their homes, offered them food, and showed a general willingness to assist in the police investigation.

4.4 Evaluation

Best practice case example: written by Members of the INPS, Lyon, France

Key points:

- A record is kept of all FDP analyses conducted by law enforcement, including both administrative and scientific information.
- We email the requesting law enforcement officer/Magistrate to try and request feedback on the usefulness of the FDP findings for the criminal investigation.

At the INPS we keep a record of all FDP analyses we have conducted for law enforcement. This record is kept in an excel sheet and currently contains about 60 cases. For each case we include two types of information. First, administrative information about the laboratory reference case, the contact details of the law enforcement officer/Magistrate who ordered the FDP analysis, the nature of the crime committed, the request date for FDP analysis, the expected result deadline requested by the magistrate, and the date we sent the final FDP report to the requesting law enforcement officer/Magistrate. Second, scientific information related to type of material being analysed (DNA extract or object found at the crime scene) and if FDP results were provided or not (the latter may be the case if the results are undetermined). Genetic data is not stored in this table nor transmitted to the law enforcement officer/Magistrate. We email the requesting law enforcement officer/Magistrate to try and request feedback on the usefulness of the FDP findings, whether the suspected perpetrator has been identified, and/or information on whether the reported FDP predictions were correct. Our goal is three-fold:

- to improve our FDP analysis in instances when FDP findings appear incorrect (e.g. by adjusting our interpretation threshold for admixed BGA);
- to better advise law enforcement officers/Magistrates interested in FDP analysis by providing concrete examples of FDP findings helping an investigation;
- to better communicate the benefits and limits of FDP to law enforcement, Magistrates, the Head of our Institute, the scientific community, and the public.

There are limitations with our approach to evaluation. For example, law enforcement officers and Magistrates do not always provide feedback to us. However, when they have, most findings suggest FDP utility; in some cases, FDP findings facilitated the identification of the DNA donor. Interestingly, in one case, the FDP result was unexpectedly discordant to the eyewitness testimony, and, following this result, the police investigation later showed that the DNA trace was not linked to the crime.

References

Amankwaa, A. O., and C. McCartney. 2019. "The effectiveness of the UK national DNA database." *Forensic Sci Int. Synergy* 1:45-55. doi: 10.1016/j.fsisyn.2019.03.004. https://www.researchgate.net/publication/331945001_The_effectiveness_of_the_UK_national_DNA_database. Accessed 8th April 2020.

Baskin, D. R. , and I. B. Sommers. 2010. "Crime-show-viewing habits and public attitudes towards forensic evidence: the 'CSI effect' revisited." *The Justice System Journal* 31 (1):97-113. https://www.ncsc.org/~/media/Files/PDF/Publications/Justice%20System%20Journal/CRIME-SHOW-VIEWING%20HABITS%20AND%20PUBLIC%20ATTITUDES.ashx. Accessed 8th April 2020.

Cashman, K. , and T. Henning. 2012. "Lawyers and DNA: Issues in Understanding and Challenging the Evidence." *Current Issues in Criminal Justice* 24 (1):69-83. https://www.researchgate.net/publication/268819486_Lawyers_and_DNA_Issues_in_Understanding_and_Challenging_the_Evidence. Accessed 8th April 2020.

Cassa, C. A., S. K. Savage, P. L. Taylor, R. C. Green, A. L. McGuire, and K. D. Mandl. 2012. "Disclosing pathogenic genetic variants to research participants: quantifying an emerging ethical responsibility." *Genome Res* 22 (3):421-8. doi: 10.1101/gr.127845.111. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3290777/. Acceessed 8th April 2020.

Chambrone, L., and G. C. Armitage. 2016. "Commentary: Statistical Significance Versus Clinical Relevance in Periodontal Research: Implications for Clinical Practice." *J Periodontol* 87 (6):613-6. doi: 10.1902/jop.2016.150554.

Cino, J. G. 2017. "Tackling technical debt: Managing advances in DNA technology that outpace the evolution of law." *Am. Crim. L. Rev.* 54:373-421. https://www.omicsonline.org/open-access/tackling-technical-debt-managing-advances-in-dna-technology-thatoutpace-the-evolution-of-law-2169-0170-1000183.php?aid=70590. Accessed 8th April 2020.

Cole, S. A. . 2015. "A surfeit of science: The "CSI effect" and the media appropriation of the public understanding of science." *Public Understanding of Science* 24 (2):130-146.

Cole, S. A., and R. Dioso-Villa. 2009. "Investigating the 'CSI effect': media and litigation crisis in criminal law." *Stanford Law Review* 61 (6):1335-1373.

de Cerqueira, C.C.S. , V. Ramallo, T. Hünemeier, S. de Azevedo, M. Quinto-Sanchéz, C. Paschetta, C. Cintas, M. González, L. Schüler-Faccini, M. Cátira Bortolini, and R. González-José. 2016. "Predicting Physical Features and Diseases by DNA Analysis: Current Advances and Future Challenges." *Journal of Forensic Research* 7 (4). doi: doi:10.4172/2157-7145.1000336. https://www.omicsonline.org/open-access/predicting-physical-features-and-diseases-by-dna-analysis-current-advancesand-future-challenges-2157-7145-1000336.php?aid=79947. Accessed 8th April 2020.

Fischhoff, B., and A. L. Davis. 2014. "Communicating scientific uncertainty." *Proc Natl Acad Sci U S A* 111 Suppl 4:13664-71. doi: 10.1073/pnas.1317504111. https://www.pnas.org/content/pnas/111/Supplement_4/13664.full.pdf. Accessed 8th April 2020.

Goddard, K. A., E. P. Whitlock, J. S. Berg, M. S. Williams, E. M. Webber, J. A. Webster, J. S. Lin, K. A. Schrader, D. Campos-Outcalt, K. Offit, H. S. Feigelson, and C. Hollombe. 2013. "Description and pilot results from a novel method for evaluating return of incidental findings from next-generation sequencing technologies." *Genet*

Med 15 (9):721-8. doi: 10.1038/gim.2013.37. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3927794/. Accessed 8th April 2020.

Guillen, M., M. V. Lareu, C. Pestoni, A. Salas, and A. Carracedo. 2000. "Ethical-legal problems of DNA databases in criminal investigation." *J Med Ethics* 26 (4):266-71. https://jme.bmj.com/content/medethics/26/4/266.full.pdf. Accessed 8th April 2020.

Hannum, G., J. Guinney, L. Zhao, L. Zhang, G. Hughes, S. Sadda, B. Klotzle, M. Bibikova, J. B. Fan, Y. Gao, R. Deconde, M. Chen, I. Rajapakse, S. Friend, T. Ideker, and K. Zhang. 2013. "Genome-wide methylation profiles reveal quantitative views of human aging rates." *Mol Cell* 49 (2):359-367. doi: 10.1016/j.molcel.2012.10.016. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3780611/. Accessed 8th April 2020.

Holmgren, J. A., and J. Fordham. 2011. "The CSI effect and the Canadian and the Australian Jury." *J Forensic Sci* 56 Suppl 1:S63-71. doi: 10.1111/j.1556-4029.2010.01621.x.

Howes, L. M. . 2017. "Sometimes I give up on the report and ring the scientist: Bridging the gap between what forensic scientists write and what police investigators read." *Policing & Society* 27 (5):541–559.

Karberg, S. . 2018. "Phenotyping: The hidden mugshot in a culprit's DNA " *European Biotechnology*. https://european-biotechnology.com/up-to-date/backgrounds-stories/story/phenotyping-the-hidden-mugshot-in-a-culprits-dna.html. Accessed 8th April 2020.

Kayser, M., and P. M. Schneider. 2009. "DNA-based prediction of human externally visible characteristics in forensics: motivations, scientific challenges, and ethical considerations." *Forensic Sci Int Genet* 3 (3):154-61. doi: 10.1016/j.fsigen.2009.01.012.

Lindsey, S., R. Hertwig, and G. Gigerenzer. 2003. "Communicating statistical DNA evidence." *Jurimetrics* 43:147-163. https://www.researchgate.net/publication/27282335_Communicating_statistical_DNA_evidence. Accessed 8th April 2020.

Lipphardt, V., V. Toom, A-C. Mupepele, and T. Lemke. 2017. Open Letter on Critical Approaches to Forensic DNA Phenotyping and Biogeographical Ancestry. https://www.researchgate.net/publication/316527261_OPEN_LETTER_ON_CRITICAL_APPROACHES_TO_FORENSIC_DNA_PHENOTYPING_AND_BIOGEOGRAPHICAL_ANCESTRY. Accessed 8th April 2020.

Liu, F., K. van Duijn, J. R. Vingerling, A. Hofman, A. G. Uitterlinden, A. C. Janssens, and M. Kayser. 2009. "Eye color and the prediction of complex phenotypes from genotypes." *Curr Biol* 19 (5):R192-3. doi: 10.1016/j.cub.2009.01.027. https://www.sciencedirect.com/science/article/pii/S0960982209005971. Accessed 8th April 2020.

Lynch, M. et al. . 2008. *Truth Machine: The Contentious History of DNA Fingerprinting*. Chicago: University of Chicago Press.

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 Sci Int Genet* 6 (1):e16-7; author reply e18-9. doi: 10.1016/j.fsigen.2010.12.012. https://www.researchgate.net/publication/49799165_Bracketing_Off_Population_Does_Not_Advance_Ethical_Reflection_on_EVCs_A_Reply_to_Kayser_and_Schneider. Accessed 8th April 2020.

Nance, D. A., and S. B. Morris. 2005. "Jury Understanding of DNA Evidence: An Empirical Assessment of Presentation Formats for Trace Evidence with a Relatively Small Random Match Probability." *The Journal of Legal Studies* 34 (2):395-444. doi:10.2139/ssrn.462880.

Nelkin, D. , and S. Lindee. 1995. *The DNA Mystique: The Gene as a Cultural Icon*. New York: W.H. Freeman.

Ossorio, P. N. 2006. About face: forensic genetic testing for race and visible traits. *J Law Med Ethics* 34 (2):277-292. doi:10.1111/j.1748-720X.2006.00033.x.

Oliver, K., T. Lorenc, and J. Tinkler. 2019. "Evaluating unintended consequences: New insights into solving practical, ethical and political challenges of evaluation." *Evaluation*. doi: 10.1177/1356389019850847.https://pure.york.ac.uk/portal/en/publications/evaluating-unintended-consequences-new-insights-into-solving-practical-ethical-and-political-challenges-of-evaluation(49b8efd0-0699-4623-8dbe-4fd1670ed2f7).html. Accessed 8th April 2020.

Prainsack, B. . 2009. "Book review: Truth machine: The contentious history of DNA fingerprinting by M. Lynch, S.A. Cole, R. McNally and K. Jordan Chicago, University of Chicago Press." *Critical Policy Studies* 3 (1):143-145.

Rosenberg, N. A., J. K. Pritchard, J. L. Weber, H. M. Cann, K. K. Kidd, L. A. Zhivotovsky, and M. W. Feldman. 2002. "Genetic structure of human populations." *Science* 298 (5602):2381-5. doi: 10.1126/science.1078311. https://www.researchgate.net/publication/10980474_Genetic_Structure_of_Human_Populations. Accessed 8th April 2020.

Sainani, K. L. 2012. "Clinical versus statistical significance." *PM R* 4 (6):442-5. doi: 10.1016/j.pmrj.2012.04.014. https://onlinelibrary.wiley.com/doi/full/10.1016/j.pmrj.2012.04.014. Accessed 8th April 2020.

Samuel, G., and B. Prainsack. 2018a. "Forensic DNA phenotyping in Europe: views "on the ground" from those who have a professional stake in the technology." *New Genetics and Society*. doi: 10.1080/14636778.2018.1549984. https://www.researchgate.net/publication/329314232_Forensic_DNA_phenotyping_in_Europe_views_on_the_ground_from_those_who_have_a_professional_stake_in_the_technology. Accessed 8th April 2020.

Samuel, G., and B. Prainsack. 2018b. The regulatory landscape of forensic DNA phenotyping in Europe. VISAGE. http://www.visage-h2020.eu/Report_regulatory_landscape_FDP_in_Europe2.pdf. Accessed 8th April 2020.

Samuel, G., and B. Prainsack. 2019. Societal, ethical, and regulatory dimensions of forensic DNA phenotyping. VISAGE. http://www.visage-h2020.eu/PDF/Delliverable_5.2_for_online_publication_vo1.pdf. Accessed 8th April 2020.

Schweitzer, N. J., and M. J. Saks. 2007. "The CSI effect: popular fiction about forensic science affects the public's expectations about real forensic science." *Jurimetrics* 47:357–364.

Scudder, N., D. McNevin, S. F. Kelty, S. J. Walsh, and J. Robertson. 2018. "Forensic DNA phenotyping: Developing a model privacy impact assessment." *Forensic Sci Int Genet* 34:222-230. doi: 10.1016/j.fsigen.2018.03.005.

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 Sci Int Genet* 22:e1-4. doi:10.1016/j.fsigen.2016.01.010. https://www.researchgate.net/publication/291385761_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_materi. Accessed 8th April 2020.

Toom, V. 2018. Cross-Border Exchange and Comparison of Forensic DNA Data in the Context of the Prüm Decision. Policy Department for Citizens' Rights and Constitutional Affairs, Directorate general for internal policies. https://op.europa.eu/en/publication-detail/-/publication/36524b0c-7b46-11e8-ac6a-01aa75ed71a1/language-en/format-PDF/source-120515586. Accessed 8th April 2020.

van der Bles, A. M., S. van der Linden, A. L. J. Freeman, J. Mitchell, A. B. Galvao, L. Zaval, and D. J. Spiegelhalter. 2019. "Communicating uncertainty about facts, numbers and science." *R. Soc. open sci.* 6. doi: http://dx.doi.org/10.1098/rsos.181870. https://royalsocietypublishing.org/doi/10.1098/rsos.181870. Accessed 8th April 2020.

Walsh, S., F. Liu, K. N. Ballantyne, M. van Oven, O. Lao, and M. Kayser. 2011. "IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information." *Forensic Sci Int Genet* 5 (3):170-80. doi: 10.1016/j.fsigen.2010.02.004. https://www.academia.edu/24657351/IrisPlex_A_sensitive_DNA_tool_for_accurate_prediction_of_blue_and_brown_eye_colour_in_the_absence_of_ancestry_information. Accessed 8th April 2020.

Walsh, S., F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, and M. Kayser. 2013. "The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA." *Forensic Sci Int Genet* 7 (1):98-115. doi: 10.1016/j.fsigen.2012.07.005. https://www.fsigenetics.com/article/S1872-4973(12)00181-0/fulltext. Accessed 8th April 2020.

Williams, R., and M. Wienroth. 2017. "Social and Ethical Aspects of Forensic Genetics: A Critical Review." *Forensic Science Review* 29 (2):145-169. https://eprint.ncl.ac.uk/file_store/production/238608/AFC65F5D-B92C-49DB-8373-E33DB2696F9C.pdf. Accessed 8th April 2020.