ETHICAL DIMENSIONS OF THE APPLICATION OF NEXT GENERATION SEQUENCING TECHNOLOGIES TO CRIMINAL INVESTIGATIONS

Author: National DNA Database Ethics Group

Date: March 2017

1.0 Introduction

1.1 Next Generation Sequencing (NGS) is a term used to describe DNA sequencing technologies whereby multiple pieces of DNA are sequenced in parallel. This allows large sections of the human genome to be sequenced rapidly. The name is a catch-all-phrase that refers to high-throughput sequencing rather than the previous Sanger sequencing technology, which was much slower. NGS is also known as Massive Parallel Sequencing and the terms are often used interchangeably. Within this document the term NGS refers to technologies that provide more wide-ranging information than the standard DNA short tandem repeat (STR) profiling techniques that measure the number of repeats at a specific region of non-coding DNA within an autosomal chromosome.

1.2 NGS sequencing technologies have developed rapidly over the past decade while the costs associated with sequencing have declined. Whilst need and utility, and not merely the availability and affordability of NGS technologies, should be the driver for their introduction into criminal investigations, declining costs increase the feasibility of their introduction. It is therefore timely that the ethical issues associated with the application of NGS in criminal investigations are considered. In this document, the Ethics Group (EG) provides an outline of the NGS technologies that are likely to become available in the next 10 years and a map (albeit not yet an in-depth discussion) of the ethical challenges associated with the application of these technologies for forensic purposes.

1.3 This document has been developed to provide advice and to inform the thinking of those individuals who are minded to consider the development of next generation sequencing technologies and those who are considering the application of these technologies for the investigation of crimes. It is also for use by the public who are interested in these techniques within a forensic context.

2.0 Approach

2.1 As with the application of all new technologies for forensic investigation purposes, all practices involving NGS technologies require ethical consideration, if possible prior to introduction. Ethical concerns identified by the EG and wider stakeholders should be considered in conjunction with arguments put forward by members of the

wider public. Moreover, it is important to bear in mind that there may well be ethical issues associated with *not* introducing NGS technologies if those technologies are available and can impact on preventing and solving crimes and eliminating individuals from investigations and suspicion. Also the right "to enjoy the benefits of scientific progress and its applications", which is stipulated in Article 15 of the United Nations *International Covenant on Economic, Social and Cultural Rights* (1976), could be seen to prescribe an (at least ethical) obligation to consider the potential benefits of technology use in the context of criminal investigation very seriously.

2.2 The Ethics Group will consider the ethical issues associated with NGS techniques and technologies at an early stage of their introduction into police work to help inform consideration. The approach to be taken by the group is to invite those developing and introducing the tests to their meetings to inform the committee about the capabilities and limitations of the tests. Together with other stakeholders we aim to debate and scrutinise the new tests and consider their contribution to the public good as well as their implications on privacy, confidentiality, autonomy, dignity and freedom of individuals. The group will debate these issues over a series of meetings until they form a collective and consensual view of the issues. The Ethics Groups views will be fed to the DNA Database and Fingerprint Strategy Board and Home Office Ministers as well as those developing the technologies.

3.0 Context

3.1 Whilst developing this document, the EG have been cognisant of the following principles which are integral to the ethical consideration of any new technology for criminal investigation purposes.

3.2 *Individual and collective rights and interests*

3.2.1 Ethical considerations of genetic technologies to assist with criminal investigations have often focused on individuals' interests and rights and the necessity for the state to provide justification for interventions which might impact on an individual's autonomy. However, a broader approach to considering ethical issues is required. Moral and ethical thinking should consider collective interests and rights alongside individual ones. The protection of society from harm (including the harm of being a victim of crime) in a way that is consistent with justice to individuals and groups of individuals, is in itself an important moral imperative. It is also important to be especially vigilant to safeguard the interests of groups, whether or not identified by law, who may be particularly at risk from the actions being considered.

3.2.2 An initial consideration for any NGS technology should be *'Should our society be doing this, and if so, why?'*. This should involve consideration of the rights and interests of individuals, groups of individuals and society as a whole, as well as the

effect of the approach being considered on the ‘moral climate’, including trust in public services and respect for the rights and dignities of others.

3.3 *Trustworthy technology use and institution*

3.3.1 A key factor in mitigating the harms that might arise from the use of NGS technologies in the domain of forensics is to ensure that technologies are used in trustworthy ways and governed by trustworthy institutions. Transparency, as well as open and authentic communication are factors that help build public trust and confidence in the organisations that are involved in the testing (and later on use) of the new technologies. To ensure that the public are well informed about NGS technologies a coordinated approach is required, by the Government who introduce the policies, the police forces who commission the tests and the forensic providers of the services, in order to create an environment of openness and to stimulate public understanding and debate.

4.0 Future Monitoring

4.1 The EG is aware of its own role in exploring the above issues; demonstrating appropriate consideration and making these open to the public. It is inevitable that some of the by-products of these technologies are still unknown therefore the EG will continue to monitor these.

5.0 Table of Ethical Issues arising in connection with Next Generation Sequencing Technologies

		1	2	3	4	5	6	7	8
		Sequencing of all classes of STR alleles (autosomal, X, Y)	mtDNA analysis	SNP-based analysis of degraded samples	SNP-based phenotypic profiling	RNA and mi RNA analysis	Epigenetic analyses: tissue origin and age	Human microbiome analysis / metagenomic analysis of human microbiota	WGS and WES[1]
A	What are the potential public benefits?	*Autosomal*: Additional discriminatory power and increased ability to interpret mixed DNA profiles. It will increase match statistics for complex mixtures and relatedness testing. *Y-STR*: improvement of crime detection and conviction, esp. in sexual crimes. *X –STR*	Ability to obtain a DNA profile from a sample that otherwise would have to be discarded, thus improving crime detection and conviction.	Ability to obtain DNA profile from a sample that otherwise would have to be discarded, thus improving crime detection and conviction.	Additional information (and potentially also evidence), thus improving crime detection and conviction.	Ability to determine tissue type (e.g. menstrual blood v. arterial blood) which can speed up investigations and help save cost.	Additional information (and potentially also evidence), thus improving crime detection and conviction. There are also potential benefits in the identification of unidentified remains and determining the age of an individual.	Additional information (and potentially also evidence), thus improving crime detection and conviction.	Additional information (and potentially also evidence), thus improving crime detection and conviction.

[1] Note that most of the issues raised across practices in columns 1 -7 would apply WGS and WES as well.

		analysis: can provide information on inheritance patterns and in turn assist with complex kinship scenarios, such as incest and human identity.							
B	What are the potential public harms?	*Autosomal*: No additional harms greater than current STR analysis, as long as no full genomes are sequenced. STR sequencing kits might only be available with other SNP tests for identification, phenotype and geography – in which case there would be wider issues. *Y-chromosome*: Assessment of	Assessment of potential public harms depends on how it is used (in individual cases only? Loaded to central database? Available for routine searches?) The police would need to be aware that mtDNA is passed down the maternal line and therefore mtDNA analysis could	n/a	If the probabilistic nature of information is not properly understood, innocent people could be unduly implicated in investigations. This could reduce public trust in new DNA technologies compared to non-DNA technologies such as fingerprint analysis.	n/a	If the probabilistic nature of information is not properly understood, innocent people could be unduly implicated in investigations. This could reduce public trust in new DNA technologies compared to non-DNA technologies such as fingerprint analysis. There is also	Human microbiota can be easily perturbed (for example with antibiotics) leading to false conclusions. If the probabilistic nature of information is not properly understood, people could be unduly implicated or eliminated in investigations. This could reduce public trust in new	Concentration of genetic and genomic information in the hands of public agencies including potentially information about health; increase of bio-surveillance; use of public funds when benefits may be low.

		potential public harms depends on how it is used (in individual cases only? Loaded to central database? Available for routine searches?). Y-chromosome analysis can provide information about male infertility (which may be unknown to the donor).	lead to biological relatives of offenders being implicated in criminal investigations. Could provide predictive information about the risk of later onset mt disease (would depend on the areas that are analysed).		Racial stereotypes could be reinforced. There is also the risk that out of blind trust in the technology, public funds could be spent on this technology when the actual benefit may be small. Risk that analysis could reveal information on distant genetic relationships and reinforce views about crime running in families.		the risk that out of blind trust in the technology, public funds could be spent on this technology when the actual benefit may be small.	DNA technologies compared to non-DNA technologies such as fingerprint analysis. There is also the risk that out of blind trust in the technology, public funds could be spent on this technology when the actual benefit may be small. Potential to highlight sensitive information about the likelihood of certain diseases.	
C	What individuals or groups would be most affected?	*Autosomal*: offenders, victims, suspects. *Y-chromosome*: People with Y-	Females, but also males if used purely for individual identification, not relatedness.	Offenders, victims, suspects.	Offenders. Ethnic minorities are likely to be most affected if the harm described in 4B	Offenders, Victims, Suspects.	Offenders, Victims, Suspects.	Offenders, Victims, Suspects.	Depending on scenario of use. People from whom subject samples are taken, (arrestees,

		chromosomes whose DNA profile is stored in the NDNAD, Y-chromosome carriers who have a chromosomally male relative (in a direct line) whose DNA profile is stored. X-chromosome: males and females. Likely to be used in incest cases so could provide vulnerable people with information about their genetic relationships.	Biological relatives of offenders		materialised.				suspects); offenders (crime scene samples are sequenced); possibly the entire population.
D	Threat to human rights or moral entitlements?	*Y and X-chromosome*: Undue discrimination. Right to privacy. Proportionality of the use of familial genetic information in	Undue discrimination. Right to privacy. Proportionality of the use of familial genetic information in minor	n/a	Presumption of innocence, right to privacy, public understanding of the technology and acceptance.	n/a	Right to privacy (if analysis discloses also information on accelerated ageing etc).	Right to privacy (if analysis discloses information on diseases).	Right to privacy.

		minor offences. Public understanding of the technology and acceptance.	offences.						
E	Risk of error or injustice?	No additional harms greater than current STR analysis, as long as no full genomes are sequenced.	n/a	Greater risks if technology is not properly validated and quality controlled.	Greater risks if technology is not properly validated and quality controlled.	n/a	n/a	n/a	Depending on how the sequence would be used.

6. Technical Terms and Acronyms

DNA (Deoxyribose nucleic acid)	DNA is found in the cells of an organism and carries that organism's heritable material used in the development, functioning and reproduction of all known living organisms. DNA is a nucleic acid and consists of two stands coiled around each other to form a DNA double helix. Each DNA strand is composed of smaller units called nucleotides. A nucleotide consists of a nitrogen base, a deoxyribose sugar and a phosphate. The sequence of the nucleotides encodes the biological information
Epigenetics	Analyses that look at the chemical modification of the DNA sequence, affecting how DNA is expressed – not at the order of nucleotides as such
Human Microbiome analysis	The collection of microorganisms that inhabit the human body. It includes the genes and genomes of the microbiota as well as the products of the microbiota and environment.
Metagenome analysis	Analysis of the genes and genomes of the microbiota, including plasmids.
Microbiota	The microorganisms of a particular site (e.g. the human gut, throat, etc.)
Mitochondrial DNA (mtDNA)	The DNA information contained not in the cell nucleus but in the mitochondria, the 'power houses' of the cell. mtDNA is maternally inherited. A single mitochondrial genome is much smaller than a nuclear genome. Most human cells contain many more mtDNA genomes than nuclear DNA, which is why mtDNA analysis in forensics can be useful in cases where DNA samples are compromised (e.g. missing persons, natural disasters, etc.)
Single Nucleotide Polymorphisms (SNP)	Variation between individuals at the level of single nucleotides
Whole-exome sequencing (WES)	Whole-exome sequencing, i.e. the sequencing of protein-coding genes in a genome
Whole-genome sequencing (WGS)	A range of techniques aiming at mapping the sequence of entire genomes (i.e. the entire DNA of a person, not only protein coding genes) are subsumed under this term.

7. Acknowledgment

The EG is grateful to Dr Barbara Daniel and Dr Denise Syndercombe-Court from Kings College London for their contributions to a paper on Ethical Dimensions of NGS which provided a comprehensive overview of the emerging issues with NGS technologies

Annex A – Background document

ETHICAL DIMENSIONS OF NEXT GENERATION SEQUENCING

AUTHORS

Professor Barbara Prainsack, NDNAD EG Member & Kings College London
Dr Barbara Daniel, Kings College London
Dr Denise Syndercombe-Court, Kings College London

Conflicts of Interests of Authors:

Professor Barbara Prainsack
Professor Prainsack has no conflicts of interest.

Dr Barbara Daniel
Dr Daniel's has no conflicts of interest.

Dr Denise Syndercombe-Court
Dr Syndercombe-Court's department employs a Knowledge Transfer Partnership associate who is jointly funded by Illumina who is looking at the concordance between NGS and the existing methodologies for the markers that are currently stored on the National DNA Database.

Dr Syndercombe-Court is a member of the European Forensic Genetics Network of Excellence (EuroForGen), funded by the EU which undertakes research using NGS techniques and considers the legal and ethical issues associated with new technologies across Europe.

1.0 Brief & Course of Action

1.1 Many novel NGS techniques are now ready to be validated and implemented for a variety of forensic uses. The manufacturers are currently seeking guidance before they finalise and lock down their systems for:

- Scientific/technical standards requirements that feed into the configuration/ format of the system(s) and validation studies
- Legal requirements for the access to, use/non-use of genetic information that could disclose medical, ancestry or physical characteristics
- Ethical principles/requirements on the access to and use of genetic information that could disclose medical, ancestry or physical characteristics

1.2 Technology has progressed at a rapid pace. Ethical dimensions still need to be discussed and agreed to clarify the boundaries for NGS technologies and for future developments. NGS technologies are already used in the medical domain, such as in the area of gene and cancer therapy. In the forensic domain, a particular area of interest are advances in research in markers for physical appearance, including ageing. This requires ethical deliberation.

1.3 Kings College London (KCL) is actively engaged in research and is collaborating with various suppliers and manufacturers of this technology and have presented work they have done to date. As KCL is the current leading UK organisation for the NGS technology, it was proposed that National DNA Database Ethics Group (EG) member, Dr Barbara Prainsack met with Dr Barbara Daniel, both of KCL, to discuss the topic and to prepare the following discussion paper for the EG.

ETHICAL DIMENSIONS OF NEXT GENERATION SEQUENCING (NGS)

2.0 Background

2.1 For a long time the main DNA-based technology used for policing and forensic purposes were short tandem repeat (STR) profiles. STR profiling utilises the fact that individuals (except monozygotic twins and multiples) have different numbers of repeats of chains of nucleotides at the same location of their genome; the more different loci are compared, the chance that two different people coincidentally have the same number of repeats in all loci decreases. STR profiles in databases consist of combinations of numbers representing the STR count at given loci. It is not possible to make useful inferences regarding a person's externally visible traits or disease risk based on an STR profile.

2.2 Spurred by the rapid advance of high-throughput technologies, genetic and genome-based analysis has become much cheaper and faster in recent years. While police and forensic databases still use STR profiles, other genome-based technologies have become available and are used to support the investigation of individual cases. There are clear ethical challenges related to these developments. The purpose of this paper is to map these challenges to aid discussions in the NDNAD EG.

3.0 SNP analysis

3.1 Single nucleotide polymorphisms (SNPs) refer to variations at the level of single bases (nucleotides). They represent the most common type of genetic polymorphism in humans. Common and rare SNPs have been shown to underpin many diseases and traits.

3.2 *Traits and ancestry*. Findings of genome-wide association (GWA) studies have provided a wealth of information relevant to forensically useful phenotypes and traits. Many traits useful for forensics are highly

heritable (over 80%) including eye and hair colour. For skin colour predictive values are currently significantly poorer.[2] Commercial kits are already available that include autosomal SNPs targeted at identification and ancestry estimation.[3]

3.3 A number of genetic models are emerging for the probabilistic prediction of quantitative traits by combining SNP results and haplotyes. SNPs can also give indications of a likely genetic ancestry (so-called ancestry informative markers, AIMs). In admixed societies such as UK, composite and specific genetic ancestry of people's respective grandparents can be of forensic value in criminal investigations particularly for examples in cases of unidentified crime victims.

3.4 Important advances are currently being made in the area of SNP-based inferences for male pattern baldness, hair structure, and facial features (see Kayser *et al*. 2015[i] for an overview) although the current research on determining body height from genetic markers has not revealed particularly useful sets, despite the high hereditability.

3.5 *Identification despite degraded DNA samples*. Resolving profiles from poor quality DNA is a frequent necessity from DNA sample 'stains' recovered from crime scenes. For a while it was believed that SNPs are less sensitive to DNA degradation than STRs as polymerase chain reaction (PCR, a DNA amplification method) will often 'stutter' incorrectly across poly-nucleotide repeats in poor quality DNA (although this can be limited by use of smaller PCR amplicon products containing the STR loci of interest). It was originally thought that

---

[2] See e.g. Identitas V1 Forensic Chip which includes several markers allowing inferences regarding EVC and ancestry; see identitascorp.com; https://www.newscientist.com/article/mg21728995-500-dna-identichip-gives-a-detailed-picture-of-a-suspect/; see Kayser 2015 for an overview.

[3] E.g. by *Illumina* (http://www.illumina.com/areas-of-interest/forensic-genomics.html); *Thermo Fisher* Scientific (https://www.thermofisher.com/uk/en/home/industrial/human-identification/next-gen sequencing-for-forensics.html).
The strategy of these two companies is different in that Illumina aims to replace the capillary electrophoresis (CE) method by NGS, while Thermo Fisher Scientific seeks to supplement CE by NGS. Also sther companies (Qiagen and Promega) are developing strategies.

research looking at panels of 100,000's of SNPs would allow selection of around seventy SNPs with moderate to high minor allele frequencies across all populations and that this would be sufficient for unequivocal identification matching of a given individual (Budowle 2007) and that SNPs would replace STR-based analysis at least for complex mixtures (especially for low-copy DNA where there is drop out and drop in of STR alleles). Replacement of the current technology is not considered likely today, partly because of the worldwide investment in STR databases, but also because improvements in chemistries now allows more probative analysis of degraded DNA. It is much more likely that identification SNPs will be used as a supplement to STRs in the future.

3.6 Last but not least, 'ultra-deep' sequencing on the basis of SNPs has also enabled the differentiation between sets of identical twins (who always have the same STR profile).

4.0 Next generation sequencing (NGS)

4.1 Very broadly speaking, there are three main approaches used within next generation sequencing: (1) the use of specific gene panels (including anything from ten to several hundreds of selected genes), (2) whole exome sequencing looking at the protein coding part of the DNA, which contains about 85 % of known disease causing variants (WES) (Choi *et al*. 2009)[ii] and whole genome sequencing (WGS).[4]

4.2 It is likely that whole genome sequencing of individuals will be cost effective for forensic application within a few years. Whilst focused sequencing of mitochondrial, Y and other already forensically employed STR loci has obvious advantages for reduced cost and simplicity, the massive datasets produced by NGS technologies pose considerable problems for data storage (Brion *et al*. 2010[iii] and Weber-Lehmann *et al*. 2014)[iv]. These approaches are also likely to offer a more useful

[4] We leave shotgun sequencing – i.e. the reconstruction of a long strands of DNA on the basis of random 'shots' at different parts of it that will then be assembled together – out of this overview as it is not suitable for forensic analysis; two sequencing experiments of samples from the same person, e.g. from a crime scene stain and a subject sample, may yield different results because they look at different targets

approach when dealing with low template DNA.[5] PCR is a competitive process and more likely to be successful if only a limited number of SNPs are targeted. Currently, looking beyond STRs for identification requires additional material to be available for analysis, which may be problematic when the crime scene sample is compromised. NGS solutions will enable different approaches – identification, phenotypics, geographic ancestry and age – to be reported simultaneously. Some companies are producing different sets to comply with legislative differences between different jurisdictions.

4.3 The developed sets will almost certainly include almost all of the currently available STR identification markers. An additional advantage will come from the increased allelic variation provided through sequence differences within an STR repeat revealed with NGS. In addition to improving match statistics the approach will allow for increased precision when interpreting complex mixtures and when calculating relatedness statistics. NGS will therefore enable STR-based identification of known offenders with increased predictive value, and also provide high confidence ancestry, phenotype and relatedness predictions for those samples where there is no match on a DNA database, all at a manageable cost from a single sample.

5.0 Epigenetics

5.1 Variations not only in the DNA sequence but in how genes are regulated and expressed offer new ways of identification, which could help distinguish monozygotic ('identical') twins or multiples who share the same DNA sequence (Li *et al.* 2013)[v]. Typically these approaches are looking for differences in a chemical modification of the sequence, rather than a difference in the sequence; the methylation of the cytosine residue is one such modification commonly used. These epigenetic changes will vary within genes controlling the function of

[5] Low template DNA, also referred to as low copy number DNA (LCN), signifies DNA from samples that contain less than the 250pg (>100pg) necessary to produce a complete profile using the standard number of PCR cycles.

different tissues, offering insights into tissue origin, but they are also environment driven and are helpful in determining the age of the person who has left a stain (Bocklandt *et al*. 2010)[vi].

6.0 Metagenomics

6. 1 The current focus of metagenomic microbial studies has been on animals with the expectation that techniques will be transferable to humans in the future. The main studies of the human microbiome for forensic application have been based on biological warfare agents and more recently saliva (Hasan *et a,l* 2014)[vii]. Analysis currently focuses on pure cultures within the laboratory, therefore consideration has to be made about the extra information obtained from forensic samples from different sources and possibly multiple contributors.

6.2 The use of shotgun amplification in a metagenomics approach, instead of the traditional 16S rRNA, produces a much wider analysis of the genome and as a result a vast increase in data. Current difficulties include assigning reads obtained from reference genomes because large areas of the genome have not yet been sequenced, therefore making this data presently unavailable. As investigation and research in to this area progress, these 'orphan reads' will be able to be aligned. At that point, more information will be available for other researchers to use.

7.0 Ethical challenges

7.1 If SNP data correlating with known (not externally visible) phenotypes are held in national databases, then governments could query the databases to assess if associations for aggressive behaviour or criminally relevant traits or phenotypes are evident. When research in this field advances, profiles of 'risky' individuals, even in the absence of (re-)offending, could then be retained for longer periods than those of others. Similarly, if SNP data were divulged to third parties (such as employers or insurance companies), discrimination on the basis of

supposed genetic risks could ensue. At present, most countries have legislation in place that prohibits the use of forensic DNA samples or profiles for any other purpose than forensic identification[viii].

7.2 Moreover, SNP analysis could yield more in-depth information about the possible distant genetic relatedness between individuals whose profiles are stored in the database. While the approach, known as 'familial searching' or 'genetic proximity testing' in the context of STR profiles, has helped to solve several cases in the US and the UK, its efficacy could improve further with the use of SNP testing. This, however, would also exacerbate the concerns voiced in connection with the approach. Furthermore, genetic proximity testing could reinforce views about the alleged prevalence of criminality in certain families; reveal to relatives that a genetic relative has a profile on the database; or even reveal a genetic link (or lack thereof) between individuals unaware of it. For example, this might reveal paternity information that the parties involved had not asked for and that potentially could disrupt social and familial structures (Haimes *et al*. 2006[ix] and Greely *et al*. 2006[x]). Finally, it has been argued that the use of genetic proximity testing, reinforces existing demographic disparities in the criminal justice system, in which arrests and convictions differ widely based on race, ethnicity, geographic location, and social class (Greely *et al*. 2006[xi], Bieber *et al*. 2006[xii] and Kim *et al*. 2011[xiii]).

7.3 Whilst some panels of SNPs and other markers have been shown to provide accurate predictions of what someone looks like, or where they are likely to have originated from, concerns have been voiced with regard to making inferences about externally visible traits (EVC) and other traits from DNA material. All forensic interpretation is made using a probabilistic approach and it is vital that any prediction is made with a full understanding of the likely error rates, that the tests are fully validated with blind testing and are presented for intelligence use only, in order to avoid over-interpretation of the data. It has been argued that because of its probabilistic nature, phenotypic profiling is especially prone to misinterpretation (Cho & Sankar *et al*. 2004[xiv]). It has also been argued that due to its probabilistic nature, insofar as 'predictions'

of EVC are concerned, it is of utmost importance that criminal investigators keep in mind that the perpetrator may look very different from what the test result suggests (i.e. he may have blue eyes despite the test 'predicting' brown eyes with 68 per cent likelihood), and that intelligence led mass screening should not be based on EVC predictions (M'Charek *et al*. 2012)[xv].

7.4 There are already examples where EVCs have been used to produce 'photofits' of potential suspects that may be erroneous (and examples of facial comparisons provided by companies using this approach oddly have managed to predict the same hair style). Such an approach, currently based on very limited information and which certainly does not build in an ageing effect, can be very dangerous, although the separate predictions can be useful intelligence. Despite that, an EVC approach used in an historic murder in Spain, solved by traditional STR typing, accurately predicted the North African ancestry, dark eyes and dark hair, but the skin tone of the individual arrested was not as predicted. We already know that skin tone predictions are more prone to error and it may be better to use an approach where one can be more certain about the prediction – not pale skin, for example – along with the certainty of any prediction.

7.5 Also in the realm of STR profiles, new challenges lie ahead. The EG has already looked into the issue of identifying male lineages as a result of new Y-chromosomal markers (Y STR markers) being proposed for use in DNA profiling in England and Wales. The EG advised in favour of the use of Y-STR information in connection with serious crimes without opening the door to routine/speculative searches of genealogical links between males.[6] The use of NGS, and newer STR approaches, will provide this genetic information by default, however, and so its governance must be considered.

[6] Specifically the EG advised that: "Provided that searches on the basis of small numbers of markers (including Y-STR marker/s) will be governed by the same rules as familial searching on the basis of autosomal markers – i.e. if they were used for the most serious crimes only, and only after approval from the NDNAD Strategy Board in each individual case, then the EG does not see a need for a public consultation on this matter. If, however, searches including small numbers of Y-STR markers were used in a wider range of cases (e.g. volume crimes), then there may be a case for a public consultation." (EG's DNA20+ Position Paper, October 2014)

7.6 But also currently available autosomal (i.e. non-chromosomal) STR markers can provide 1st and 2nd degree relationships to be ascertained if necessary, and the NGS platforms will allow the possibility of going further without a separate workflow to reveal relationships that most people would be unaware of and could result in suggestions of a 'criminal gene'.

7.7 Also the use of human models in a metabiomic approach will pose ethical challenges. It is a sensitive technique that produces large volumes of data, of which some will have the potential to highlight sensitive information, such as the likelihood of certain diseases. For instance, a few studies have looked into gut microbiota in the faeces of animals to determine links between gut microbiota and obesity. This moves on from looking at the microbes that are present in a sample to which ones are active and the active processes that are occurring within a human body.

7.8 A way to address many of the ethical concerns is to create analysis pipelines as a safeguard to ensure that only forensically useful information is obtained rather than sensitive information. In the future, discussions will be needed in order to determine whether information which has been identified as forensically useful but also as sensitive can be included in forensic analyses. Due to the uncertainty related to the capabilities of these methods and the vast areas where studies need to be carried out, at present it is believed that such methods should be used as a theoretical investigative tool rather than being used to produce active conclusions. This is, however, a developing area and will need a watching oversight.

8.0 Error rates

8.1 The error rates of these new technologies have not been fully explored, or not sufficiently within a worldwide population. While we may be able to predict that someone has blue eyes with 91% certainty, if you are Dutch, how does that translate across populations? Moreover, eye colour is one of the best EVC methods available. Blind studies of

geographical ancestry are very promising but again cannot provide certainty because of individuals of mixed ancestry and the lack of research that has been undertaken in some areas of the world. It is important that predictions do not produce false positives and the uncertainties must be made clear; already some companies are producing 'heat maps' of likely ancestry that may be used to drive an inappropriate investigation, and revealing this descriptive information to the public in a high profile case is likely to undermine the usefulness of the approach in the future if revealed to be wrong.

8.2 Reliable software: In the case of NGS data, it is very difficult to analyse the sequence manually. The reporting of NGS results thus depends on the availability of reliable software. An additional complication is that reliability thresholds for clinical use are not likely to be sufficient for forensic analysis, and that the standards and parameters for validation cannot, as is typically the case in the context of medical applications, remain entirely up to commercial companies and thus unavailable to scrutiny by public authorities (Borsting & Morling 2015)[xvi].

8.3 With the potential availability of sequence that can be looked at to reveal potential susceptibilities to disease, once such a condition has been revealed the question as to whether there should be searches of medical databases (many of which will be freely available to all, offering individual although anonymised, data) needs to be considered. A recent forensic ethics workshop was of the view that forensic and medical databases should be separately compartmentalised and that there should be no cross interrogation. This is, however, not a new area for forensic geneticists – even prior to DNA analysis disease, or disease susceptibility, was often revealed to the individual scientist, but never taken further due to the code of practice held by the individual or institution. Should we be looking for information that will influence appearance? One might argue that some genes that predict height will also be related to conditions that produce extremes of height in an individual, sufficient to be considered a disease, but should we target genes that might reveal the presence of a cleft palate in an individual, or albinism, for example? However the likely availability of such

information in every analysis is something that will need a consideration of governance structures.

9.0 References

[i] Kayser M. Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. *Forensic Science International: Genetics* (2015) *18*, 33-4

[ii] Choi M, Scholl UI *et al*. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proceedings of the National Academy of Sciences.* (2009) 106(45),19096-19101.

[iii] Brion M, Quintela I *et al*. New technologies in the genetic approach to sudden cardiac death in the young. *Forensic science international*. (2010) *203*(1),15-24.

[iv] Weber-Lehmann J, Schilling E *et al*. Finding the needle in the haystack: Differentiating “identical” twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Science International: Genetics* (2014) *9*, 42-46.

[v] Li C, Zhao, S et al. Differences of DNA methylation profiles between monozygotic twins’ blood samples. *Molecular biology reports* (2013); *40*(9), 5275-5280.

[vi] Bocklandt S, Lin W *et al*. Epigenetic predictor of age. *PloS one*. (2010) *6*(6), e14821-e14821.

[vii] Hasan NA, Young BA *et al*. Microbial Community Profiling of Human Saliva Using Shotgun Metagenomic Sequencing. *PLoS ONE* (2014) 9(5): e97699.

[viii] Wienroth M, Morling N, Williams R. Technological Innovations in Forensic Genetics: Social, Legal and Ethical Aspects. *Recent Advances in DNA and Gene Sequences* (2014) 8 98-103.

[ix] Haimes, E. Social and ethical issues in the use of familial searching in forensic investigations: insights from family and kinship studies. *The Journal of Law, Medicine & Ethics* (2006) *34*(2), 263-276.

[x] Greely HT, Riordan, DP. Family ties: the use of DNA offender databases to catch offenders' kin. *The Journal Of Law, Medicine & Ethics*. (2006) *34*(2), 248-262.

[xi] Greely HT, Riordan, DP. Family ties: the use of DNA offender databases to catch offenders' kin. *The Journal Of Law, Medicine & Ethics*. (2006) *34*(2), 248-262.

[xii] Bieber, FR. Turning base hits into earned runs: improving the effectiveness of forensic DNA data bank programs. *The Journal of Law, Medicine & Ethics* (2006) *34*(2), 222-233.

[xiii] Kim J, Mammo D *et al*. Policy implications for familial searching. *Investigative genetics* (2011) *2*(1), 1.

[xiv] Cho MK & Sankar P. Forensic genetics and ethical, legal and social implications beyond the clinic. *Nature genetics* (2004) *36*, S8-S12.

[xv] M’charek A, Toom V & Prainsack B. Bracketing off populations does not advance ethical reflection on EVCs: a reply to Kayser and Schneider. *Forensic Science International: Genetics* (2012) *6*(1), e16-e17.

[xvi] Børsting C & Morling N. Next generation sequencing and its applications in forensic genetics. *Forensic Science International: Genetics* (2015) 18: 78-89.