Forensic DNA analysis

A PRIMER FOR COURTS

THE ROYAL SOCIETY

RSE
The Royal Society of Edinburgh
KNOWLEDGE MADE USEFUL

Forensic DNA analysis: a primer for courts
Issued: November 2017 DES4928

ISBN: 978-1-78252-301-7

This primer is produced by the Royal Society and the Royal Society of Edinburgh in conjunction with the Judicial College, the Judicial Institute, and the Judicial Studies Board for Northern Ireland.

To request additional copies of this document please contact:

The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG
T +44 20 7451 2571
E law@royalsociety.org
W royalsociety.org/science-and-law

This primer can be viewed online at royalsociety.org/science-and-law

Contents

Science and the law primers
Foreword

The judicial primers project is a unique collaboration between members of the judiciary, the Royal Society and the Royal Society of Edinburgh. The primers have been created under the direction of a Steering Group chaired by Lord Hughes of Ombersley and are designed to assist the judiciary when handling scientific evidence in the courtroom. They have been written by leading scientists and members of the judiciary, peer reviewed by practitioners, and approved by the Councils of the Royal Society and the Royal Society of Edinburgh.

Each primer presents an easily understood, accurate position on the scientific topic in question, as well as considering the limitations of the science, and the challenges associated with its application. The way scientific evidence is used can vary between jurisdictions, but the underpinning science and methodologies remain consistent. For this reason we trust these primers will prove helpful in many jurisdictions throughout the world and assist the judiciary in their understanding of scientific topics.

The production of this primer on forensic DNA analysis has been led by Lady Justice Rafferty DBE and Professor Niamh Nic Daéid FRSE. We are most grateful to them, to the Executive Director of the Royal Society, Dr Julie Maxton CBE, the former Chief Executive of the Royal Society of Edinburgh, Dr William Duncan, and the members of the Primers Steering Group, the Editorial Board and the Writing Group. Please see the back page for a full list of acknowledgments.

Sir Venki Ramakrishnan
President of the Royal Society

Professor Dame Jocelyn Bell Burnell
President of the Royal Society of Edinburgh

1. Introduction and scope

The aim of this primer is to present:

1. a scientific understanding of current practice for DNA analysis used in human identification within a forensic science context
2. guidance to the Judiciary in relation to the limitations of current interpretation and evaluations that can be made, so that they can be informed when making decisions relating to DNA evidence.
3. The primer has been laid out in sections providing the basic information relating to DNA analysis used in forensic science.

Section 1 provides an introduction to DNA and its use as a forensic science tool as well as the nature of the questions that can be addressed with the most commonly used DNA analysis methods.

Section 2 addresses the following specific questions as they relate to forensic science:

1. What is DNA?
2. How is DNA inherited?
3. What parts of DNA are analysed and how are DNA profiles generated?
4. How are DNA profiles compared and interpreted?
5. How are mixed DNA profiles assessed?
6. What are the limitations to DNA profiling of complex samples?

Some of these areas and questions are expanded upon in the Appendices 1–3. Examples are provided in Appendix 4 and a glossary in Appendix 5.

Section 3 provides a short insight into future areas of development in relation to DNA profiling and Section 4 presents a summary of the current state of the art, including current limitations.

1.1 DNA and forensic science

DNA profiling was first proposed by Sir Alec Jeffreys in 1984 when he found that individuals could be differentiated on the basis of readily detectable differences in their DNA. DNA profiling was first used in a criminal case in the UK in the investigation of the 1983 and 1986 rapes and murders of Lynda Mann and Dawn Ashworth. In this case, Richard Buckland was exonerated through DNA analysis in 1987 and Colin Pitchfork was subsequently convicted. Since 1987, considerable scientific study and resource has been devoted to the development and refinement of DNA analysis technologies. In 1995 the UK National DNA Database was established to maximise the investigative use of DNA profiles and to identify repeat offenders. On a global scale, most countries now use forensic DNA analysis in one form or another. The main questions that a forensic DNA scientist is asked to address are:

1. Whose DNA is it?
2. From what body fluid has it originated?
3. How did it get there?
4. Have the results been reported in a fair and balanced way?

Provided there is sufficient DNA, the interpretation of a DNA profile from a single individual's sample is straightforward and can provide powerful scientific evidence either to exclude or to include any one individual as a possible source of that DNA. That is done by calculating and presenting the match probability; that is, by calculating statistically how rare any matching DNA profile is in a population.

Technological improvements in DNA analysis resulting in the ability to analyse ever smaller quantities of DNA have led to the main developments in this area. This capability has raised important questions relating to:

1. understanding and controlling contamination
2. the interpretation of complex DNA samples.

A variety of computer software programs have been developed for complex sample interpretation, using a range of statistical methods. In the UK, the Forensic Science Regulator's Codes of Practice and Conduct set out the requirements for the validation of software programs used for complex mixed DNA sample interpretation. This necessitates:

1. defining the type of DNA profiles the software program is being used to analyse
2. demonstrating that the model used by the software is acceptable for these DNA profiles
3. scientifically validating the software program to address specifically the type of casework samples it is being used to interpret
4. issuing a statement of validation completion. This statement must clearly identify the uses for which the method is validated and any weaknesses, strengths and limitations.

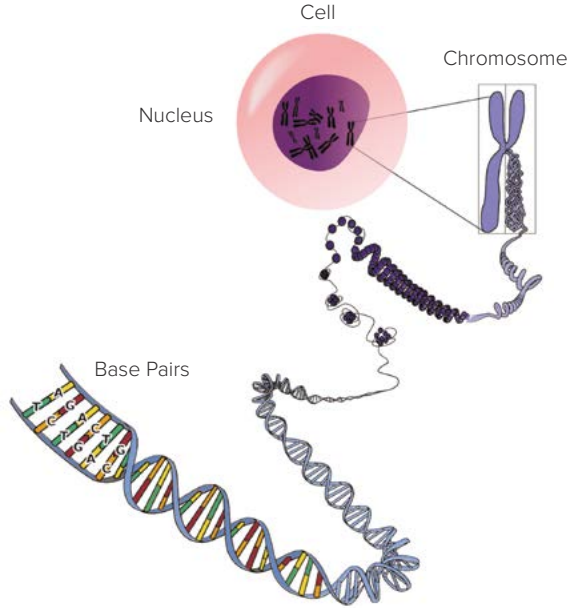
There is a developing scientific research base on the evaluation of how DNA transfers onto an item, and DNA scientists rely on the published scientific literature as well as on their experience and knowledge of the underlying circumstances of each case.

2. Science

DNA is composed of four chemical constituents (labelled A, T, C and G), known as bases, attached to a sugar backbone which can form a strand millions of bases long. There are two such strands in DNA, which run in opposite directions. The bases pair up to form a twisted ladder. Each base pairs exclusively with one other base on the opposite strand: A to T and G to C. This means that when the strands separate, each one can act as a template to reproduce the other precisely. The linear sequence of bases can act as a code, providing the instructions for many biological functions. Figure 1 shows how the bases in DNA are held in paired strands which naturally twist into a double helix structure. Each cell in the human body contains 6,500,000,000 pairs of bases. The full complement, 3 metres in length, is termed the genome. It is packaged into 23 different pairs of chromosomes. During the formation of sperm or eggs, the chromosome pairs are separated with one member of each pair randomly allocated to each sperm or egg. When an egg and sperm fuse during fertilisation, the full set of 23 pairs is re-established. This means that half of a child's DNA comes from the mother and half from the father, and full siblings will on average share half of their DNA.

FIGURE 1

Basic representation of DNA (image adapted from Creative Commons).



Changes in the sequence of bases on the DNA strands (mutations) can arise as a result of errors in DNA replication or repair. As a result an individual might acquire 30–100 mutations relative to their two parents' genomes. This constant influx of new mutations has allowed differences to build up over generations so that the chances of two human genomes being the same are infinitesimally small. An exception is identical twins, who will have identical DNA, except for new mutations.

Forensic DNA analysis focuses on examining specific sections of DNA that are known to be particularly variable between individuals in order to create a DNA profile. The part of the DNA that is examined is called a locus (plural loci), which is a unique site along the DNA of a chromosome characterised by a specific sequence of bases. Currently, an individual's entire genome is not analysed to create his or her DNA profile. This means that part or all of the same DNA profile could be shared by more than one person. The statistical analysis of forensic DNA data therefore focuses on establishing the weight of evidence that should be attached to the similarity between the DNA profile of a person of interest and DNA taken from a crime scene.

Appendix 1 provides a more in-depth focus on DNA inheritance and the use of DNA in forensic science.

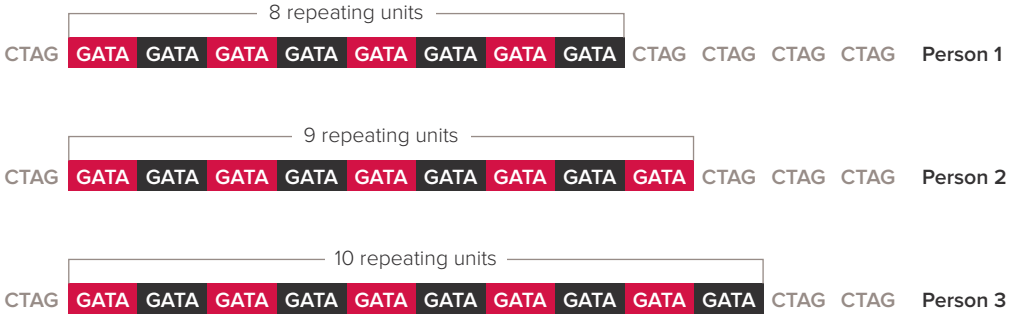
2.1 DNA analysis in forensic science – short tandem repeats

Only small sections of an individual's DNA are analysed routinely for forensic evidence. The parts analysed are called short tandem repeats (STRs). Mutations that affect the number of repeats are relatively common so within a population there are usually several different versions of the DNA at an STR locus with different repeat lengths. The different versions are called alleles (Figure 2).

The frequency of occurrence of a specific allele (ie a specific number of repeating units) at the tested locus in a specific population provides a measure of how common that allele is in that population. This information is essential for calculating match probabilities. If only one STR were analysed, there would be many people with the same allele, purely by chance. It is therefore necessary to analyse a number of different STR loci to ensure that the chance of two unrelated people having matching DNA profiles is very small. Over time, the number of different STR loci analysed has increased as technology has developed. Since 2014 in the UK, 16 loci are examined. In some Scottish cases, 23 loci are examined.

FIGURE 2

STRs of different lengths of repeating units of four bases (represented by GATA) on a single strand of DNA from three different people at the same locus.



2.2 DNA analysis in forensic science – Y chromosome DNA

A second form of DNA analysis involves study of loci found only on the male specific Y chromosome. Y chromosome DNA is inherited by sons from their father with little change between the generations. As a consequence, the profiles generated from Y chromosome DNA are very similar between males with a shared direct male ancestor, with only very rare mutations leading to differences between males who share their Y chromosome. Analysing Y chromosome STRs can be helpful where there is a mixture of DNA from male and female contributors, for example, in a sexual assault case.

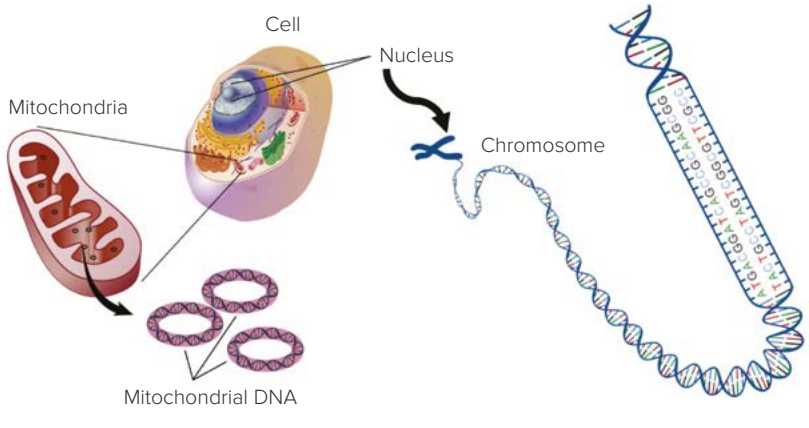
2.3 DNA analysis in forensic science – mitochondrial DNA

It is also sometimes helpful to analyse mitochondrial DNA (mtDNA) which is contained in small structures (called mitochondria) within cells. They are found in the cell body, rather than in the nucleus. The mitochondrial genome consists of only 16,500 bases, arranged in a circle (Figure 3). In contrast to the presence of only two parental copies of the nuclear DNA, there are thousands of copies of mitochondrial DNA in the same cell. Both males and females have mitochondrial DNA but it is exclusively inherited from the mother. All of a mother's children have the same mitochondrial DNA, which is the same as that of all their relatives in the same maternal line. Because of the many copies of mitochondrial DNA present in a cell, this analysis is useful when there is a minute amount of DNA present or when the DNA sample is very old and has broken down. STR profiling and mtDNA / Y chromosome analysis are distinctly different and there are many more individuals who would have matching mtDNA profiles by chance than with STR profiling.

Appendix 2 provides more in-depth information on how DNA is analysed and how a DNA profile is obtained.

FIGURE 3

Mitochondrial and nuclear DNA (image adapted from Creative Commons).



2.4 Comparison of DNA profiles

2.4.1 Collection of DNA samples – avoiding contamination

Biological evidence from a crime scene needs to be collected carefully, transported and stored properly prior to examination. Most biological evidence is best preserved when stored dry and/or frozen. Contamination in the context of DNA analysis can be defined as the introduction of extraneous DNA (or biological material containing DNA) to a sample. The DNA profiling process is extremely sensitive and constant vigilance is required to ensure that contamination does not affect the results. Because of this sensitivity, contaminating DNA may still be observed even with careful precautions, and will routinely be monitored in laboratories. The forensic scientist must use all the information available to them to assess whether a contamination event, if it occurs, has had an impact on the results in a specific case.

2.4.2 Evaluating the statistical weight of matching a single DNA profile

If there is a match between the STR profiles of two DNA samples, then there are three possible explanations:

1. The suspect is the source of the material.
2. The material came from a second person who has an identical DNA profile to that of the suspect.
3. The match is a false positive due to contamination or some other kind of error.

The match probability is an estimate of the likelihood (or chance) of observing the DNA profile obtained if someone other than, and unrelated to, the suspect, was the source of the DNA. An expanded explanation is presented in Appendix 3.

2.4.2.1 Complex DNA profiles

In some instances, the amount of DNA in a sample might be lower than optimal, or it might be of poor quality (degraded) or consist of many contributors (a mixture). In such a situation, particular care must be taken in interpreting the DNA profile. There will always come a point below which no interpretation can deal effectively with the level of variability in a poor DNA profile. There is no simple way of defining the lowest-level profile that should be interpreted. A scientist should always stay within the validated range for his or her interpretation methods using the relevant laboratory equipment and tests and should not attempt to interpret profiles that fall outside this range.

2.4.2.2 Factors to be considered in the evaluation and weight of evidence of DNA profiles

In evaluating matching DNA profiles, it is important to consider how the DNA came to be present in a particular place. Understanding from which material the DNA came can assist in this evaluative process. Current tests for body fluids are not definitive and forensic scientists may not be able to give an opinion as to the body fluid from which a DNA profile originated. Other samples (hair, skin etc) can also provide DNA profiles.

DNA can in some instances be transferred from person A to person B and then onto object 1 ('secondary transfer') or from person A to object 1 to person B and then onto object 2 ('tertiary transfer'). In both cases, traces of person A's DNA might be found on an object even when they have never been in direct contact with that object. It is also perfectly possible that the DNA of person B will not be present on an object with which they have had direct contact. In some cases (but not always) it will be possible to make a comparative assessment between alternative explanations for the presence of the DNA.

Appendix 3 provides more information relating to the evaluation of DNA profiles and the weight which can be put on such evidence, in the light of factors such as transfer and persistence of DNA.

3. The future

Scientists are exploring new DNA methods, which may, for example, enable prediction of an individual's skin, hair or eye colour. These methods, at their current stage of scientific development, would be primarily of use in an investigation for intelligence purposes rather than as evidence presented in court. Methods to examine an individual's entire genome have also been developed and are becoming faster and less costly. The use of different parts of the genome for human identification purposes within the Criminal Justice system has not yet been fully explored.

More accurate chemical testing methods for determining the type of body fluid from which a sample originated are also being developed. Although not yet widely in use, these would enable scientists to be more certain about the type of material (blood, semen, saliva or other cellular material) from which a DNA sample might have originated.

4. Summary

Forensic DNA analysis has been established as a core scientific technique since the mid-1980s and has been used widely in the UK courts and many courts around the globe. Its underpinning science is reliable, repeatable and accurate, and based on validated technology and techniques for both the generation of a DNA profile and the interpretation of that profile. When forensic DNA analysis is adduced as evidence in court, the following matters should be borne in mind when assessing both admissibility and weight of evidence:

- DNA profiles are generated using scientifically accepted techniques and following validated scientific methods.
- When a DNA profile is obtained from one person, the interpretation of that DNA profile is normally straightforward and provides powerful scientific evidence to either exclude or include an individual as a possible source of the DNA.
- DNA profiles can provide exclusionary evidence as well as evidence of association.
- Contamination and errors can occur in the DNA analysis process. Scientists can address case-specific issues through the processes, checks and control samples associated with that case.
- The analysis and interpretation of complex DNA profiles should be undertaken only within guidelines validated by the organisation performing the work. These guidelines should be made available.
- The weight of evidence from complex/mixed DNA profiles is largely estimated using computer software. There are a range of software programs available, which use different assumptions and statistical methods to analyse the complex/mixed DNA profiles and to produce ‘unmixed’ profiles. This means that:

    1. the same data derived from complex/mixed DNA profiles analysed repetitively by the same software can exhibit small differences in the resulting ‘unmixed’ DNA profiles.
    2. the same data derived from complex/mixed DNA profiles analysed by different software programs can exhibit more marked differences in the resulting ‘unmixed’ DNA profiles.

- The choice of software program and why it was used for the specific complex/mixed DNA samples being analysed should be explored with the scientist.
- Any estimate of weight of evidence is calculated with probability estimates: a match probability is a probability estimate, while a likelihood ratio is the ratio of two probability estimates. In the UK, match probabilities smaller than one in one billion are capped at one in one billion. Likelihood ratios greater than one billion are also capped at one billion.
- Tests to determine which body fluid(s) may have produced a DNA profile generally give only an indication of the body fluid and not a definite identification.
- There are some published studies addressing the transfer and persistence of DNA but specific circumstances relating to individual criminal cases are not likely to have been studied.

Appendix 1: Defining DNA and its use in forensic science

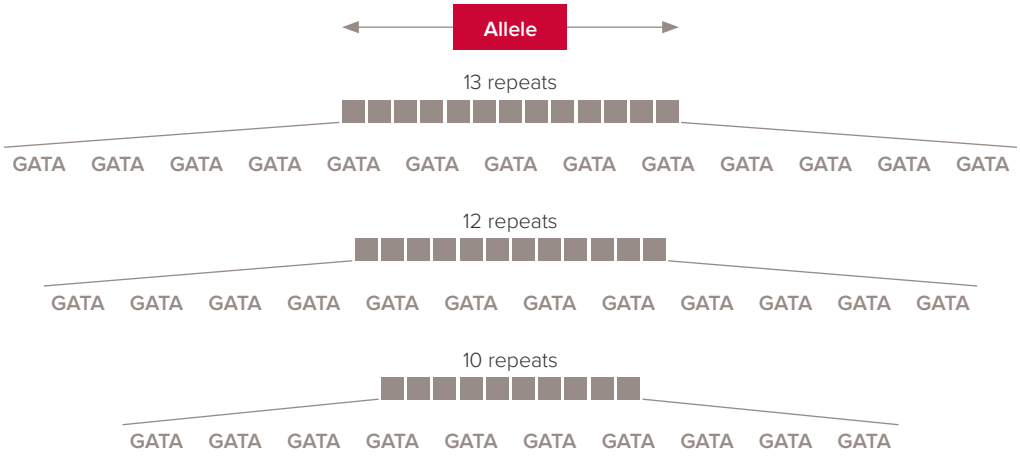
A 1.1 DNA used in forensic science

DNA is composed of four chemical constituents (labelled A, T, C and G), known as bases, attached to a sugar backbone, which can form a strand millions of bases long. Forensic DNA analysis typically assesses specific stretches of DNA (loci) where there are repeating blocks of four bases known as short tandem repeats or STRs. Mutations resulting in the gain or loss of a four-base block are relatively common and as a result the number of four-base blocks present at an STR locus shows considerable variation within a population. Each version of the locus, called an allele, has a specific number of repeats of the four-base blocks. Forensic DNA analysis is concerned with measuring the length of DNA at these sites, which correlates with the number of repeats of the four-base blocks (Figure 4).

In order to determine the length of DNA at any one locus, a technique known as a polymerase chain reaction (PCR) is used to generate many copies of the relevant stretch of DNA from material recovered at the crime scene. These DNA fragments can be separated according to their size using a technique known as electrophoresis.

FIGURE 4

A single strand of DNA illustrating a short tandem repeat (STR) composed of repeats of the four-base pair block GATA. It is the number of repeats of this block that varies between individuals. In Figure 4(a), the DNA 'type' or 'allele' is 13 as there are 13 repeats. In Figure 4(b), the allele is 12 as there are 12 repeats, and in Figure 4(c) the allele is 10 as there are 10 repeats. The locus is the region of the DNA where the STR is located. Each individual will have two copies of each locus – one from each parent, which could be the same or different alleles.

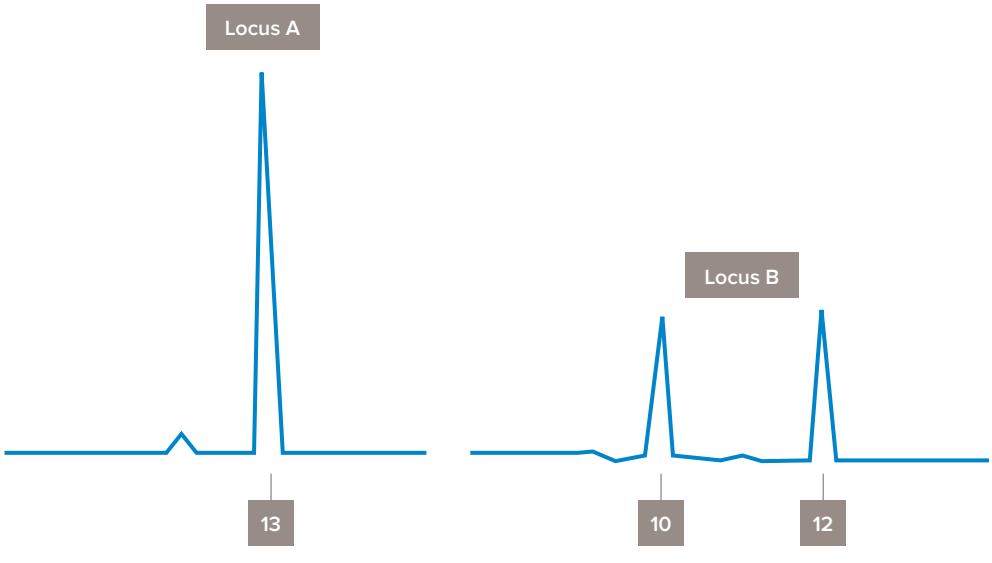


Resulting DNA profiles are represented as a numerical code (corresponding to the number of repeats of units of four bases on each allele at each STR locus), and the length of each STR is visualised on a chart known as an 'electropherogram'. On this chart, the horizontal axis shows the length of the DNA fragments and the vertical axis shows their relative abundance. Figure 5 is a schematic of part of an electropherogram showing two loci A and B. At locus A, there are two STR alleles of length 13 (one allele of length 13 from each parent) and at locus B there are two alleles of length 10 and 12 (again one allele from each parent, this time of different lengths). Because the two '13' alleles at locus A are the same length, they occur at exactly the same position on

the DNA profile chart. When there are two copies, there is twice as much of the '13' DNA present, and so the height of the peak, which represents the amount of DNA present, is about twice as tall as if there were one '13' allele present. Examining different loci and determining the alleles (a process known as 'genotyping') generates a person's DNA profile. The allele frequency is how often that number of repeating units at a particular STR locus occurs in a given population. For example, if allele 13 at locus A occurs ten times in 100 individuals, then its frequency would be ten in 200 alleles (100 people with two alleles each – one from their father and one from their mother). The statistical analysis of forensic DNA data focuses on establishing the weight of evidence that should be attached to the similarity between the DNA profile from a person of interest and material recovered from a crime scene or from a complainant/complainer.

FIGURE 5

Diagram of the alleles representing the STRs from each of the two copies of DNA present (one contributed by each parent) at two loci A and B.



A 1.2 Current DNA profiling methods

The principal method of forensic DNA analysis is to consider the profile of the STRs. If only one STR section of DNA were analysed, many people would share the same DNA profile. Therefore, it is necessary to analyse a number of different STRs to ensure that the chance of two unrelated people's STR profiles matching is acceptably small. Over time, the number of STRs analysed in human DNA profiling has been increased to the point that the chance of two unrelated people sharing the same DNA profile has become infinitesimally small. Table 1 illustrates the evolution of the numbers of STRs analysed. There are various commercial analytical kits containing the chemicals required for the analysis of groups of STRs at the same time. These kits are called multiplexes. In addition to the STRs, each of the systems also includes a test to determine whether the sample comes from a male or a female.

TABLE 1

The STR DNA profiling systems used in the UK.

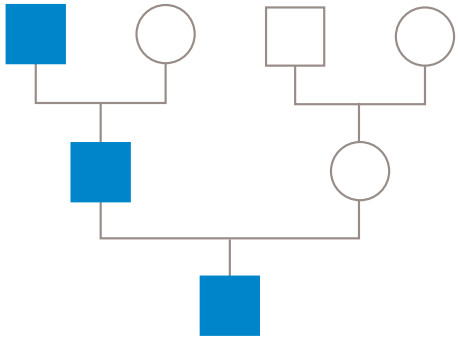
Years used	Number of STRs analysed	The commercial kits (multiplexes) used for the analysis of groups of STRs present at different loci
1995 – 1999	6	**SGM** (Second Generation Multiplex): Few of the DNA profiles held on the National DNA Database are SGM profiles – where possible, a sample matching an SGM profile would be upgraded to SGM Plus® or a later system.
1999 – 2014	10	**AmpFlSTR® SGM Plus®** (Second Generation Multiplex Plus): Many of the DNA profiles held on the National DNA Database are SGM Plus® profiles. SGM Plus® profiles contain all the STRs in the SGM grouping plus four more. This amplification kit has not been in routine use since 2014.
2014 – present	16	The names of the multiplexes used in the UK are: **PowerPlex® ESI 17**; **AmpFlSTR® NGM (Next Generation Multiplex) SElect™**; **Investigator ESS (European Standard Set) Plex SE**. All are collectively referred to as DNA 17 multiplexes and contain the same 16 STRs, which include the 10 SGM Plus® STRs plus six more, and a gender identifier.
2014 – present (in Scotland)	23	**AmpFlSTR® GlobalFiler™**: GlobalFiler contains the 16 STRs in ESI 17, NGM SElect and ESS Plex SE, plus an additional five STRs and two Y chromosome markers, plus a gender identifier.

A 1.3 Y STR

A second form of DNA analysis involves the analysis of DNA found in one particular chromosome found only in males, called the Y chromosome. Analysing Y chromosome STRs can be helpful where there is a mixture of DNA from male and female contributors. For example, if a sample contains a large amount of female DNA and there is only a small amount of male DNA present, then examining the Y chromosome gives just the male contributor's DNA profile rather than a mixture (Figure 6).

FIGURE 6

Diagram of Y STR links between males. Squares represent males, circles represent females.



Y-Chromosome DNA
(passed on complete from fathers to sons)

A 1.4 Mitochondrial DNA

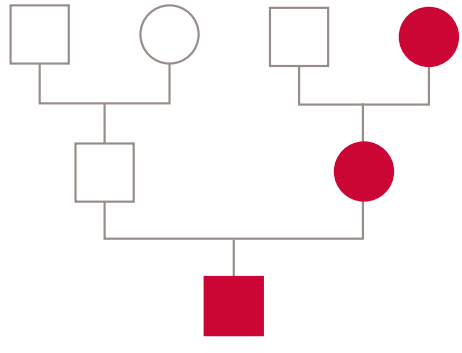
A third technique is the analysis of mitochondrial DNA (mtDNA). Both males and females have mitochondrial DNA which is always inherited from the mother. All of a mother's children have the same mitochondrial DNA, which is the same as that of all their relatives in the same maternal line (Figure 7).

Many copies of mitochondrial DNA are present in each cell, so mitochondrial DNA analysis is useful when there are very small amounts of DNA present (such as in hair shafts without roots), or when a DNA sample is very old and has broken down. In mitochondrial DNA analysis, scientists assess part of the DNA sequence rather than the length of a region of repeated blocks. As with Y chromosome analysis, and in contrast to nuclear DNA profiling, there are always more individuals who would have the same mitochondrial DNA profile. This is because relatives in the same female line over many generations share the same mitochondrial DNA. An example of mitochondrial DNA from two people who are unrelated maternally (with the differences underlined) might be:

Person 1: ACCGGTTGCAAG
Person 2: AGCGGTACCAAG

FIGURE 7

Diagram of mtDNA links between mother and children. Squares represent males, circles represent females.



Mitochondrial DNA
(passed on complete from mothers to sons and daughters)

Appendix 2: DNA analysis in forensic science

A 2.1 Samples generally analysed for DNA profiling

Forensic DNA analysis relies upon comparing DNA profiles. A DNA profile is produced from body fluids and/or other cellular material deposited during the course of the commission of a crime (for example, hair, blood, saliva from a discarded cigarette or drinks can, semen from an intimate swab from an alleged rape complainant/complainer). Such samples are called 'questioned samples'. The DNA profile from the questioned sample is compared to the DNA profile of one or more known samples from:

- suspect(s)
- complainant(s)/complainer(s)
- other people with regular access to the location from which the crime-scene samples were collected
- other relevant people such as family members (in missing person's investigations, paternity testing and mass disaster events).

Generally, for known samples, mouth (buccal) cells are collected rather than drawing blood. Buccal cell collection involves wiping a swab against the inside cheek of an individual's mouth to collect skin cells. The swab is generally frozen for storage. Known samples are collected from people already known to the investigation or from people found following a DNA database search.

A 2.2 How should DNA samples be collected and preserved for analysis?

A 2.2.1 Sample collection

The biological material present at crime scenes first needs to be detected. Body fluids may be identified visually, by chemical analysis/test/reaction or with the use of different types of light source. Sometimes the approach to targeting for testing is more intuitive and relies on the scientist's expectations of where a person may have handled an object depending on the circumstances of each case. One of the most common methods for collecting biological material from hard surfaces (such as a broken window or a knife) is using a swab. The swab is moistened with sterile DNA-free water, and then rubbed over the surface to be sampled. This might be followed by a second swab to ensure any remaining material is collected. Biological material might be collected from fabrics by cutting out a stain or by using sticky tape to collect surface material (such as from the collar of a shirt).

In choosing sampling sites for material not visible to the eye, such as cells left by handling an object, the forensic examiner will use their knowledge of the circumstances to determine where to collect the material. For example, if an assailant has grabbed a bag, the area of the bag which was grabbed will be sampled, or if an assailant has tied a ligature around a complainant's/complainer's neck, the areas where the ligature will have been handled most in tying the knot will be sampled. In allegations of rape or sexual assault, a complainant/complainer will be medically examined, and will have intimate samples, such as from the vagina or anus, and swabs of any skin areas alleged to have been touched or licked by the perpetrator, taken by a medical practitioner.

A 2.2.2 Sample preservation

Most biological evidence is best preserved when stored dry and/or frozen. These conditions reduce the rate at which DNA will break down, and prevent mould and bacteria from growing. Samples are packaged carefully, often using 'tamper-evident' bags that show a visible warning if someone has attempted to open them. They are then transported to the forensic laboratory, where they are inspected and signed for on arrival. Inside the laboratory, the samples are generally frozen, although very heavily stained, wet items might be dried in a controlled environment. Drying will assist with preservation but would generally only be used for large, heavily stained, wet items. The DNA is then chemically extracted and purified from the biological material and stored in sealed tubes either in a refrigerator at 4°C or a freezer at -20°C.

A 2.3 How is a DNA profile generated?

A DNA profile is generated from the analysis of a submitted sample or from a sample taken from a known individual. Once an item of evidence from a crime scene has been presented to the forensic scientist for DNA analysis, the following general steps are undertaken (steps 4 to 9 are also undertaken to generate a DNA profile from a sample taken from a known individual):

1. laboratory examination of the submitted item to locate any body fluid(s) present
2. recovery/sampling of body fluid
3. evaluation of the collected sample
4. DNA extraction
5. establishing how much DNA is present within the extracted sample (quantification)
6. copying (known as amplifying) of the STR regions many times using a chemical process called PCR (polymerase chain reaction)
7. separation of PCR products by size
8. detection of PCR products
9. data interpretation.

The DNA profile looks like a chart with different coloured peaks rising from a baseline (Figure 8).

FIGURE 8

STR DNA single person profiles from (a) good-quality DNA and from (b) poor-quality DNA. Going from left to right along the horizontal axis, the size of the DNA fragments gets larger. The vertical height of each peak shows how much DNA of that size is present. The numbers in the good-quality profile refer to the 16 STR loci and the gender marker, (labelled 17). Different colours are used for loci where the expected DNA fragments are of similar size, allowing alleles at one locus to be readily distinguished from alleles of similar length at a different locus. The larger DNA targets are missing from the poor-quality sample because the DNA sample was broken down and could not be profiled. The PCR process relies on fully intact DNA across the locus of interest, so detection of longer fragments is more sensitive to degradation of the sample DNA.

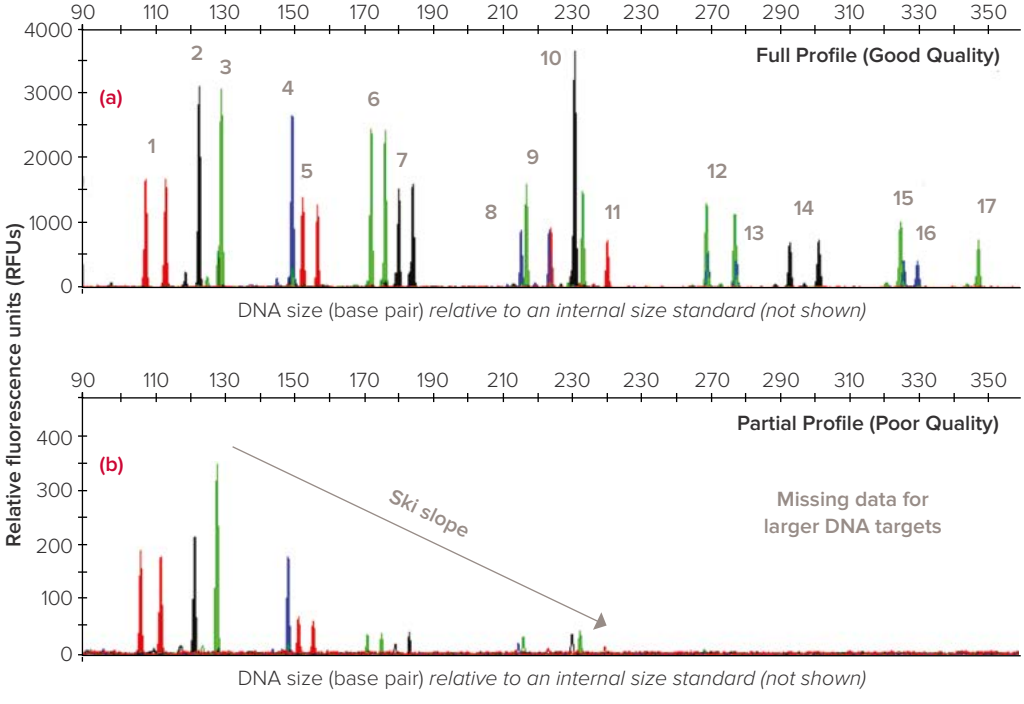


Figure courtesy of Margaret Kline, National Institute of Standards and Technology (NIST). From J M Butler (2012) *Methodology book,* Figure 10.3

The scientist can use this chart (called an electropherogram) to determine whether the sample is from a male or female, and whether it is from a single individual or from multiple individuals. In a male, the gender marker (known as amelogenin) shows two different sized peaks. In a female, only one of these peaks is seen.

At each STR locus, the number of peaks observed on the graph will give an indication of the number of individuals whose DNA has contributed to the profile. One individual will have either one or two peaks (alleles) at each STR locus.

The vertical scale of the graph represents the amount of DNA detected at each STR locus and the scientist will use the height of the peaks as an approximation for the amount of DNA of that particular type.

If the DNA profile has arisen from a mixture of DNA from two people, then three or four peaks would be expected to be observed at a number of the STR loci. As the number of contributors to the mixture increases, the number of peaks seen at each STR locus will tend to rise. However, as the number of contributors to a mixture increases, the chances are that those contributors will share some of the same peaks and so the scientist may not be able to determine with certainty how many people's DNA is present.

When mixed samples are obtained then the number of peaks at each STR locus can become difficult to determine. In some instances, there may be sufficient difference between the amount of DNA contributed by one person in a mixture and that of others, so that the entire DNA profile of the person contributing the most DNA to the mixture (the major profile) can be unequivocally determined. If all of the individuals have contributed about the same amount of DNA to the mixture, then there will be no discernible difference in height between the peaks originating from one individual versus another.

The scientist will also check whether or not the DNA profile is of the quality and clarity they would expect, given that they have already determined approximately how much DNA was present in the sample. If the DNA profile is not as good as expected, given the amount of DNA that was analysed, the scientist may choose to repeat the analysis. However, if a minute amount of DNA was available for the analysis yet a very strong and clear DNA profile was obtained, the scientist will want to double check that contamination has not occurred.

A 2.4 Interpreting DNA profiles

When viewing DNA profiles, scientists first judge whether the overall quality of the data is appropriate for reliable interpretation. In a fresh, good-quality DNA sample, the scientist will observe large peaks, which are a similar height to each other. They may be able to confidently evaluate how many people have contributed to a mixed DNA profile from good-quality DNA profiles.

A complex DNA profile is one in which one or more of the following conditions occur:

1. Less than the optimal amount of DNA present (low template).
2. A mixed DNA profile where the number of contributors is unclear.
3. The DNA has degraded, which means it has broken down into small pieces insufficient for a full profile to be produced.
4. There are chemical components stopping the DNA profiling process from working efficiently (inhibition).

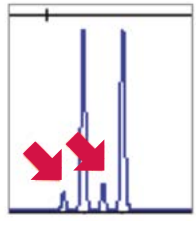
In a complex DNA profile, the scientist will often observe small peaks that are close to the baseline of the graph. This will mean that there will be ambiguity regarding what constitutes a true allele rather than an artefact of the analysis. There will also be uncertainty in defining the number of possible contributors to a mixture. If the DNA is old, or has been in a warm, humid environment, it will have started to degrade into smaller pieces, and larger STRs (longer alleles) might give disproportionately low peaks or even be missing (Figure 8(b)) giving a characteristic ‘ski slope’ appearance. All of these effects are increased if there is less than an optimal amount of DNA present in the sample to start with. As the quantity and quality of DNA decreases, some STRs show only one instead of an expected two peaks and some will give no results at all. These are known as partial DNA profiles. Figure 9 shows portions of DNA profiles with ‘noise’ artefacts marked. These artefacts will generally be excluded from the comparison between a crime-scene sample and a known sample.

Scientists must judge whether all of the parts of the DNA profile can be confidently assigned to one person or, in the case of a mixture of DNA from multiple people, how many different people's DNA might be present. Depending on the quantity and quality of DNA present, the interpretation process might be straightforward with no ambiguity, or it might leave room for a range of opinions. In a fresh known sample such as a mouth swab taken from a person of interest, scientists can distinguish which of the peaks are known and understood technical artefacts with certainty, as they will be at a very low level compared to the alleles in the DNA profile.

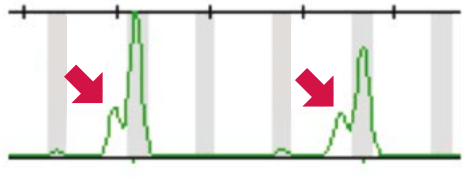
FIGURE 9

Sections of STR DNA profiles showing a range of technical artefacts.

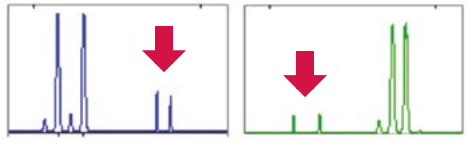
Stutters:
these are low peaks (highlighted by yellow arrows), generally one repeat unit smaller than the true peak; they are caused by slippage during the process of making copies of the DNA



A peaks:
the copying process usually adds one single 'A' base at the end of every DNA fragment, but when this has not happened completely, the result is a 'shoulder' to the left of the main peak, which is one base smaller than the main peak



Pull-up:
during the detection part of the process, a strong signal in one colour can cause small peaks of the same size to appear in the adjacent colour



As the amount of good-quality DNA decreases and/or the number of people's DNA present in the sample increases, the level of certainty in distinguishing sample DNA from noise decreases. Take for example, a DNA profile originating from two people, where person A has contributed most of the DNA and therefore has high peaks, but person B is only present at a very low level, with low peaks. It would not be possible to determine whether a small peak near to the baseline is a stutter from person A's strong profile or might be part of person B's very weak profile.

In general, scientists are aided in their interpretation of DNA profiles by computer software and by data produced during extensive testing (validation) of the analytical processes. The results from validation testing, which would include the maximum level at which each artefact (illustrated in Figure 9) is observed, are used to generate 'standard operating procedures'. Standard operating procedures are written guidelines to ensure that the scientists within an organisation make consistent interpretation decisions, supported by analytical data. While different organisations will have different procedures, each will have scientifically validated their methods. The way in which methods should be validated is prescribed in the Forensic Science Regulator's Codes of Practice and Conduct[1] and associated guidance[2]. These Codes also set out the required quality standards; for DNA analysis, accreditation to an international standard (ISO 17025) is required. The accreditation process includes independent external scrutiny of each organisation's methods and competence, to ensure they meet the required standards. This external scrutiny is provided by the United Kingdom Accreditation Service (UKAS). Appendix 3 provides a detailed analysis of how DNA profiles are compared.

1. Available at: https://www.gov.uk/government/publications/forensic-science-providers-codes-of-practice-and-conduct-2017
2. Available at: www.gov.uk/government/publications/forensic-science-providers-validation

A 2.5 What is DNA contamination and how can it be controlled?

Contamination can be defined as the introduction of DNA, or biological material containing DNA, to a sample after a (trained) responsible official has control of the crime scene.

Because the DNA profiling process is extremely sensitive, constant vigilance against contamination is required. A police officer or crime-scene examiner collecting evidence can contaminate samples if proper care is not taken. Examples of ways in which contamination could occur at a crime scene include a crime-scene examiner not changing gloves between handling different exhibits, or talking without a properly fitted face mask. Likewise, the scientist analysing the DNA can inadvertently add his or her own DNA to the sample. For this reason, detailed guidance has been published by the UK Forensic Science Regulator on avoiding DNA contamination at crime scenes, in laboratories, and in sexual assault referral centres and police custody.

It is important that all disposable items (for example swabs) and all chemicals and kits used in the analytical process are free from DNA before use. An international standard for DNA-free items has been published[3] and all purchasing of items by police and laboratories should take account of its requirements. It is critical that police ensure that, during the arrest of suspects and their processing in custody suites, cross-contamination between suspects or from a suspect to a police officer and hence potentially to a complainant/complainer or crime scene is avoided. For example, a suspect in relation to a sexual assault should not be transported in the same vehicle as was previously used to transport the complainant/complainer. Similarly, if multiple suspects are arrested, they should be processed, detained and forensic samples taken separately. If intimate samples are to be taken, they should be taken by different medical practitioners in different facilities.

Laboratories recognise that contamination can occur between people, consumables and other items in the forensic process. The working practices of labs are geared to prevent contamination and to detect it, should it happen. The use of appropriate controls and testing provides assurances that the general risks of contamination are minimised.

3. ISO 18385:2016 Minimizing the risk of human DNA contamination in products used to collect, store and analyze biological material for forensic purposes – Requirements. Available at www.iso.org/standard/62341.html

Even with all these precautions, the sensitivity of DNA profiling methods means that sometimes contaminating DNA will still be seen. This may either be as a complete or nearly complete profile, or merely one or two peaks (alleles). In the latter case, this type of very minimal contamination is known as 'drop-in.' A suggestion that contamination has adversely affected any particular case is dealt with by assessing the information available relating to the continuity of the specific evidence and evaluating particular scenarios.

A 2.6 What is the National DNA Database and what types of samples does it contain?

The UK National DNA Database (NDNAD) was established in April 1995 and is managed and operated by the Home Office on behalf of UK police forces. DNA databases can generate investigative leads in cases without suspects, and can also enable linking of serial crimes involving biological evidence. Two data sets exist, which are searched against each another:

(1) DNA profiles from offenders who have been convicted or in some cases arrested for a recordable offence.

(2) DNA profiles from evidence recovered from crime-related samples.

In addition, the DNA profiles of crime-scene staff, many police officers, all forensic science laboratory staff, many staff involved in manufacture of the reagents and consumables used in laboratory processes and some external experts are retained on elimination databases and are checked to ensure that these individuals did not inadvertently contaminate the results. These DNA elimination databases are separate from the National DNA Database. DNA samples from volunteers and missing persons are also held, but again, separately from the National DNA Database.

Over time, the number of STRs used to generate the profiles stored on the NDNAD has increased as the technology has developed. Neither Y chromosome STR data nor mitochondrial DNA sequences are held or searched against the National DNA Database.

Appendix 3: Comparison of DNA profiles in forensic casework

A 3.1 How DNA profiles are compared and the calculation of the likelihood ratio and match probability

A 3.1.1 Comparison process

The best approach to DNA interpretation, which should be followed, is for the scientist to interpret the crime sample first, and to document their findings. Exceptions should be rational and documented. Only after the crime sample has been interpreted should the scientist interpret any known samples, before making a comparison between the two. This is to reduce the risk of confirmation bias. When the crime sample DNA and the known sample DNA have been interpreted in isolation, they can be compared, to see if they may be from the same or a different source. This comparison refers to the origin of the DNA only. How and when the DNA was deposited is a matter for further evaluation.

A 3.1.2 Match probability

If there is a perfect match between the STR profiles of two DNA samples, then there are three possible explanations:

1. the suspect is the source of the material at the crime scene
2. the material came from a second person who has an identical DNA profile
3. the match is a false positive due to a sample switch or some other kind of error.

The match probability is an estimate of the likelihood of observing that profile if someone other than, and unrelated to, the suspect was the source of the DNA.

In order to calculate the weight of evidence if a match is observed, a frequency estimate for the profile is generated from representative data sets and appropriate statistical correction factors are applied so that the figure presented in court is fair and reasonable and does not overstate the strength of the evidence. About 1 in 1,000 individuals within a population has an identical twin. If there is no information as to whether a suspect has a twin, an upper limit of 1 in 1,000 should be assumed, although typically such information is available. In the UK, the lowest match probability that is reported is one in a billion, even though the actual calculation might result in an even smaller chance of a match, such as one in a trillion or even less. The reasons for this ‘cap’ on match probability are that:

1. it becomes difficult to test the assumptions required in the calculation to the point where even smaller match probabilities can be assured to be accurate
2. The real meaning of numbers in the trillions or beyond is difficult to comprehend.

Assuming that the match probability has been calculated accurately and in accordance with the laboratory's standard operating procedures, its interpretation still requires care.

Suppose the match probability is 1 in 3 million. This is the probability that a randomly chosen individual has the particular DNA profile revealed by analysis of the crime-related samples. However, this means a little over 20 people in the UK would be expected to have the same profile. The 'defence lawyer's fallacy' is to argue that there is therefore a 1 in 20 chance of the suspect being the source of the material. However, this is only true if all members of the population (including the suspect) had an equal probability of committing the crime and leaving biological material at the scene or the suspect was only identified by searching a national database of DNA profiles of randomly chosen individuals. Typically, other sources of evidence have been used to lead prosecutors to a suspect.

In contrast, the 'prosecutor's fallacy' argues that the match probability implies a 1 in 3 million chance that the suspect is innocent. Again, the statement is false. The probability of guilt or innocence given the DNA profile match is dependent on a wide variety of non-DNA factors that are unique to each case. The highest level of confidence in a match occurs when the match probability is so low that there is unlikely to be any other individual within the population with the same DNA profile.

There can, however, be other credible reasons to find a particular person's DNA at a crime scene and these should be explored. Such reasons may include whether there was evidence of any other explanation for the presence of the suspect's DNA, whether the article on which the DNA was found was associated with the offence, how readily removable the article in question was, whether there was some geographical association between the offence and the suspect and whether the DNA in question was more likely to be there by primary or by secondary transfer. As a consequence, a DNA 'match' alone should never be used to imply a suspect's involvement in a crime.

A 3.1.3 The likelihood ratio

The likelihood ratio (LR) divides the probability of obtaining the observed genetic similarity under a hypothesis associated with the prosecution view (which is generally that the suspect has contributed the DNA) by the probability of obtaining the observed match under an hypothesis relating to the defence view (for example that the suspect and actual donor of the DNA are unrelated). In simple cases, the likelihood ratio is one over the match probability for well-amplified profiles coming from one person. In more complex cases, the evidence may be evaluated under a range of hypotheses.

Likelihood ratios are generally accepted as being the most appropriate method for evaluating the evidential strength of DNA profiles. The calculation allows for different explanations for the observed evidence. For example, in a mixed DNA sample containing DNA from three people, the prosecution hypothesis might be that the DNA originated from the suspect, the complainant and an unknown, unrelated individual. The defence, however, might claim that the DNA originated from the complainant, the suspect's brother and an unknown, unrelated individual. Or indeed it might be that the DNA originated from three completely different people. The calculations for these different defences will be different and will give different results.

Scientists should communicate clearly the propositions they have considered, including the number of contributing individuals, any assumptions they have made regarding known or assumed contributors and any assumptions concerning the relationship between individuals eg that they are unrelated, as these will all affect the calculation of evidential strength. The likelihood ratio also enables scientists to account for artefacts, low-level contamination and other complexities such as low-template DNA effects, which are discussed in the following section.

As for match probabilities, there is a 'cap' placed on the likelihood ratio in the UK and the largest likelihood ratio that would be reported is one billion.

A 3.2 Low-template, degraded and compromised DNA profiles

In the past, 'low-template DNA analysis', sometimes referred to as Low Copy Number or LCN DNA analysis, was used to refer to methods where the sensitivity of the analysis was specifically boosted by altering the analysis method, to enable results to be gained from lower quantities of DNA. However, due to the very high analytical sensitivity of all current methods used for DNA analysis, specific technical boosts to sensitivity are now rarely employed.

DNA profiles can now be obtained routinely from just a few cells. The optimal amount of DNA is the amount of DNA that will yield a full DNA profile without the potential for interference from artefacts or other technical issues in the interpretation process. Sometimes, the amount of DNA from each contributor in a sample is lower than optimal because the DNA is a mixture from more than one individual – the total amount of DNA (by weight) added to the chemical reaction might contain enough of the major contributor of the mixture to generate a good-quality profile, but there might be insufficient DNA from a minor contributor to enable a high-quality profile to be generated from this (minor) contributor.

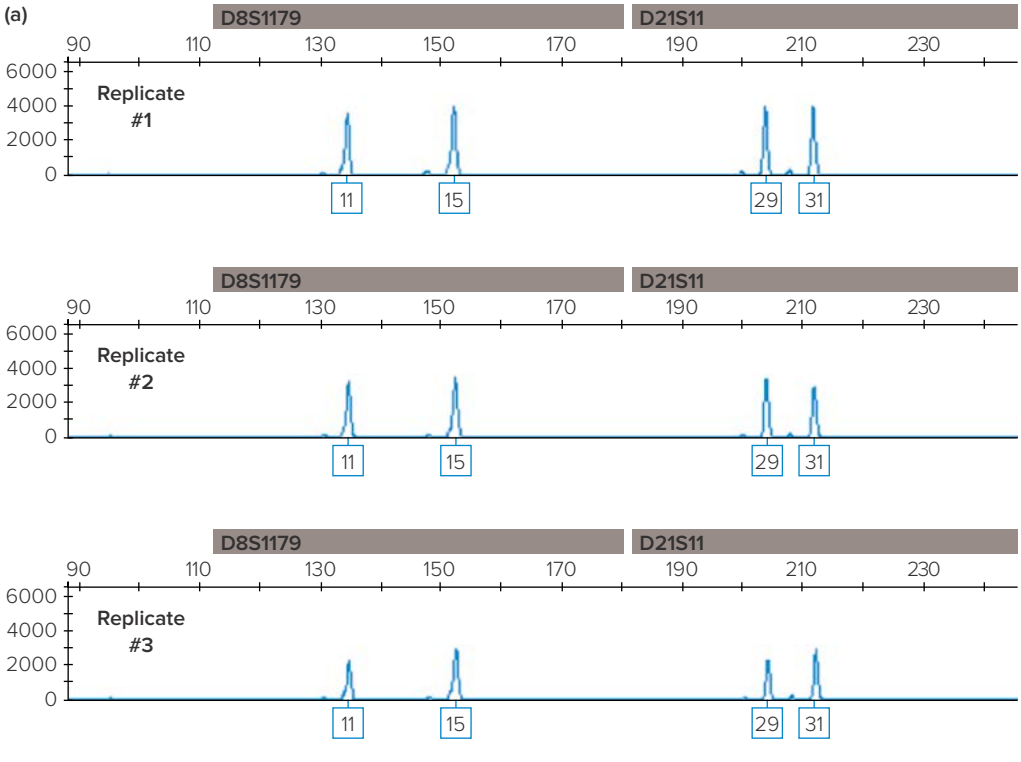
Alternatively, there might have been a sufficient quantity of DNA, but it is of poor quality, with many short segments and little of the required length for analysis. This is termed degraded DNA. DNA degrades (or breaks down) gradually as it ages, but the process is quicker if the biological material stays warm and wet. For these reasons, the amount of DNA measured and added to the chemical reaction cannot be used as a standalone guide to whether a DNA profile should be regarded as 'low template'. When the quantity of good-quality DNA is lower than optimal, particular care must be taken in interpreting the DNA profile. The optimal level of DNA is determined through the validation processes of the laboratory based on the multiplexes used to generate the profiles.

Section A 3.1 laid out the general approach to interpretation, and Figure 8(b) shows the loss of part or all of the information at one or more STRs that can occur. These effects all happen to some extent, and in a less predictable manner, when the input level of DNA is lower than optimal. In addition, the impact of DNA contamination can be greater when the amount of DNA in the evidence sample is very small – if there is very little DNA to begin with, then even a minute amount of contaminating DNA could 'take over', with the result that only the contaminant DNA and not the source DNA is seen.

One useful way to determine whether the profile is from a low level of input DNA is to analyse it two or three times, and to look at the level of reproducibility between the replicates. If the pattern of peaks remains similar between replicates (such as in Figure 10(a)), then there is sufficient DNA present to interpret reliably using standard methods. If, however, each replicate gives a very different pattern of taller and smaller peaks and some peaks are missing (such as in Figure 10(b)), then the scientist must either reject the profile as being insufficiently reproducible for reliable interpretation, or must employ special interpretation methods that have been thoroughly tested (validated) to deal with such low levels, accounting for the high degree of variability seen.

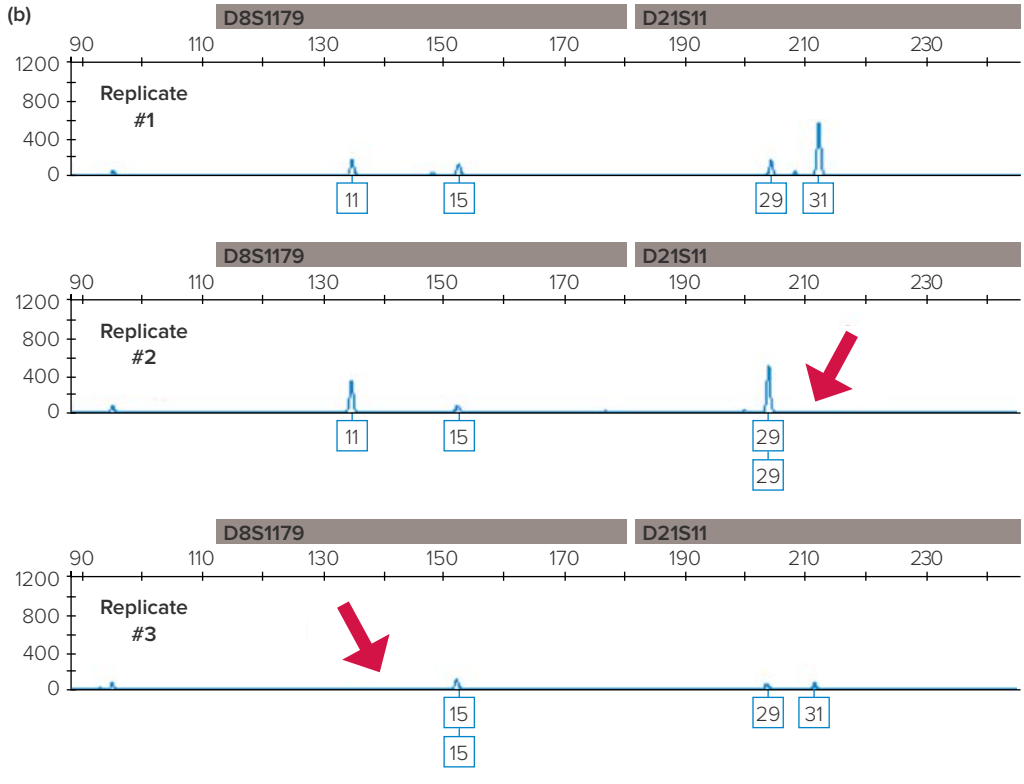
FIGURE 10

A portion of a DNA profile involving replicate tests from low-template DNA (a) where reproducible peak heights were observed in the replicate samples and (b) where drop-out occurred, shown by arrows.



There will always come a point below which no software or method of interpretation can deal effectively with the level of variability in extremely low-level DNA profiles and such profiles should not be interpreted. There is no simple way of defining the lowest-level profile that should be interpreted. A scientist should always stay within the validated range for his or her interpretation methods using the relevant laboratory equipment and tests and should not attempt to interpret profiles that fall outside this range.

(Source: www.cstl.nist.gov/strbase/LTDNA.htm (Identifiler 31 cycles, sample 1) 100 pg (left) and 10 pg (right))



A 3.3 Assessing the weight of evidence of DNA profiles

There are a range of software programs available to assist scientists in calculating the weight of evidence resulting from genetic similarity between a known sample and a crime sample. However, the interpretation method does not solely consist of the software, but also the standard operating procedures of the laboratory, which are based on validation data (including the demonstration of repeatability, reproducibility and accuracy) and the judgment of the scientist, for example in:

1. assessing whether a DNA profile is suitable for statistical evaluation, including interpretation of the various quality controls employed
2. assessing the optimal software to use for the profile(s) in question
3. ensuring that at least two suitable alternative propositions are clearly stated (occasionally there can be more than two alternatives)
4. evaluating the output from the software used
5. evaluating the combined meaning of the various biological stains, amounts of input DNA, profiles from crime samples, known samples and controls in the context of the case.

Software used for calculating the weight of evidence from DNA profiles can be divided into three types, as shown in Table 2. Whichever software is used, the interpretation method, including the software, must be validated (including the demonstration of repeatability, reproducibility and accuracy) for the types of DNA profiles that are being interpreted. The Forensic Science Regulator's Codes of Practice and Conduct publication has set out the required approach to validation. It starts with clearly defining what the method is to be used for and ends with a 'statement of validation completion', which sets out the strengths and weaknesses of the method, what it can be used for and any limitations. A scientist presenting evidence using software to aid their interpretation should therefore be able to state clearly the types of DNA samples for which their method (including the software) is validated.

There has been rapid development in methods for calculating the weight of DNA evidence in recent years, which has increased the range of complex profiles that can be evaluated. Each weight of evidence method:

1. makes different assumptions
2. uses a different subset of the raw or processed data comprising a DNA profile
3. employs different statistical models.

This means that when the weight of evidence from a complex DNA profile is estimated using the different software approaches, even if the hypotheses being tested are identical, different values for the likelihood ratio will be obtained, as each is an estimate of probability. Weight of evidence software using binary or discrete methods (Table 2) does not take account of the height of the peaks in the DNA profile, so would not use peak height to distinguish between the DNA contributors. Continuous methods (Table 2) incorporate more information from the profile, such as peak heights and artefacts. Frequently, but not always, this approach will provide stronger likelihood ratios for true contributors.

New refinements and developments in computer software are ongoing. Given that the current software programs essentially use very different statistical methods, when the same DNA data are analysed by different software systems, different results can be obtained. However, for many DNA profiles, no difference will be seen, as any software would calculate the likelihood ratio to be over a billion, and all results are capped at this level. It will only be where DNA profiles are incomplete or at low levels that differences might be seen.

TABLE 2

Types of DNA mixture interpretation software.

	Software type
	Binary
Typical uses	Simple DNA mixtures (two or three persons' DNA present). There is sufficient DNA present so that low-template DNA issues do not need to be considered.
How does it work?	The scientist evaluates which of the DNA peaks are from the source DNA. The software does not use information about the height of the various DNA peaks (although the scientist will already have considered peak height information), nor does it consider the possibility of unpredictable effects as described for low-template DNA. Therefore, this type of software is not suitable for evaluation of evidential strength where one or more of the profiles shows low-template effects. The software makes a straightforward calculation of the estimated match probability or likelihood ratio.

	Discrete variables	Continuous variables
	DNA mixtures (up to four persons' DNA present) or DNA mixtures where there might be low-template DNA issues to interpret.	Complex mixtures with low-template DNA issues to interpret. Can be used to interpret DNA mixtures from at least three different people. Some software might be capable of interpreting mixed profiles with DNA from more than three contributors.
	The scientist evaluates which of the DNA peaks are from the source DNA. The software does not take account of the various peak heights, but it does make allowances for low-template DNA effects. The software estimates the likelihood ratio, making allowances for low-template DNA effects.	The scientist rules out a small number of technical artefacts (see Figure 5), but does not need to evaluate which of the DNA peaks are from the source DNA and which are due to other technical artefacts (such as stutter). The software is programmed to know how these artefacts and low-template DNA effects vary in different samples, and in estimating the likelihood ratio, takes account of all the possibilities for each peak: whether it is really part of the source DNA or an artefact, whether it might be part of person A or person B's profile, whether the DNA might be broken down into smaller lengths (degraded) and so on. Some software of this type requires data from the laboratory using it, to ensure it reflects correctly how the effects vary between samples in that laboratory's processes. So before the laboratory uses the software for casework, it will analyse samples with known DNA profiles at various dilutions and feed the data from this analysis into the software. This enables the software to model the characteristics of the laboratory's process, for example stutter heights.

TABLE 2 (CONTINUED)

	Software type
	Binary
Reproducibility	If the same DNA profiles were evaluated on different occasions, the same result would be expected. If different software of this type were used by different scientists, it would be expected that there would be a close agreement (less than one order of magnitude difference) between the results.
Examples of interpretation software available	Many forensic science laboratories will have developed their own spreadsheets to perform calculations of this type.

	Discrete variables	Continuous variables
	If the same DNA profiles were evaluated on different occasions, the same result would be expected. If different software of this type were used by different scientists to analyse the same data, it would be expected that there would be a close agreement between the results (generally within one order of magnitude). Any differences would be due to variations in how the software is set up to deal with low-template DNA results. If the same data were analysed by this type of software and by software using a continuous method, the results might be more markedly different and the reasons why the scientist believes their method is scientifically validated and appropriate for the samples being analysed in the case should be explored.	This type of software often uses simulations (thousands of different estimations of the result) to give a final overall evaluation that is the best 'fit' for the DNA profile data. If the software was used to analyse the same set of DNA profiles on several different occasions, it would produce slightly different results each time. These variations are normal, and because they are very small in comparison to the overall result, they do not have a significant impact. If the same data were analysed by different software packages these might produce results that are more markedly different, and so the reasons why the scientist believes their method is scientifically validated and appropriate for the samples being analysed in the case should be explored.
	LRmix Studio; Resolve; LikeLTD 4+; LabRetriever; LiRa.	STRmix; TrueAllele; LiRA-HT; DNA View Mixture Solution; LikeLTD 6.+; European Forensic Mixtures.

A 3.4 Factors to consider in the evaluation of DNA

The assessment of weight of evidence as described in Section A 3.3 addresses the source of the DNA but does not consider how or when the DNA was deposited. We all transfer DNA to objects that we touch, objects onto which we sneeze, cough or bleed, and onto each other through social or sexual contact. Transfer events require three things to be in place in order to be considered – source, opportunity and mechanism. Each of these parameters is considered by scientists in the context of each case and the information supplied.

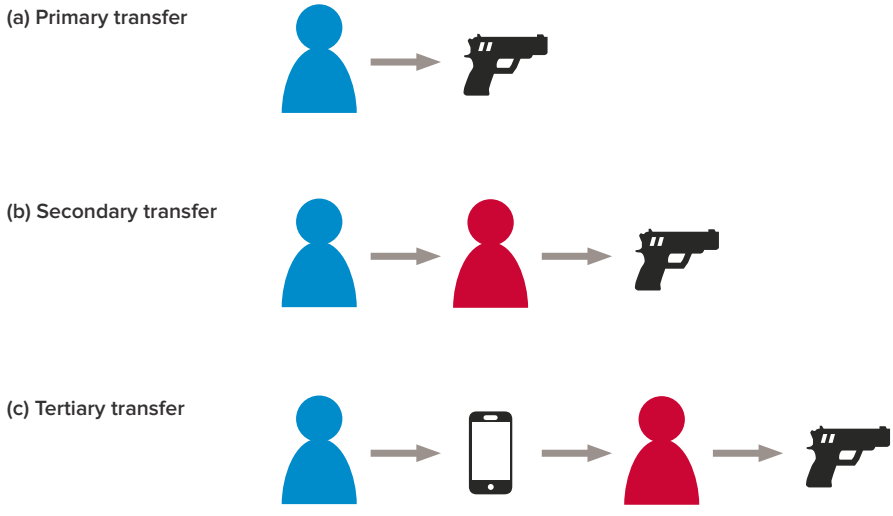
The first consideration by the scientist will concern whether or not it is possible to attribute the DNA profile to a particular body fluid. Depending on the type of sample, it might be possible to say that the DNA came from blood, semen or saliva. Scientists use chemical tests or special lighting to gain an indication of what body fluids might be present. However, the tests for these body fluids vary in their sensitivity and specificity, and there might be more than one body fluid present. If, for example, there is a very tiny, weak blood stain, but a very strong DNA profile is obtained, it is also possible that the DNA might not have originated from the blood stain, but might have been from someone who subsequently touched the blood stain. Therefore, there are occasions when a forensic scientist will not be able to give an opinion on the body fluid of origin. If the forensic scientist has sufficient information to infer that the DNA came from an identifiable body fluid, this information might assist with assessing by what activity the DNA came to be present in a sample.

Often DNA is transferred by touch rather than from a specific body fluid; this is known as ‘touch DNA’. DNA can persist for many months on an item, and determining when it was deposited is not possible. Not all touches will result in a DNA transfer and the amount of DNA we transfer in each situation will depend on a variety of factors, including:

1. person to person variability
2. how long it has been and what we have done since washing our hands/body
3. the intensity of contact (for example, a brief touch or a robust handshake)
4. whether surfaces are wet or dry, rough or smooth, absorbent or non-absorbent.

FIGURE 11

Illustration of possible (a) primary, (b) secondary and (c) tertiary DNA transfer.



In certain circumstances, DNA can be transferred from person A to person B and then to an object, leaving traces of person A's DNA on the object when they might have never been in direct contact with that object. This is known as secondary transfer. Tertiary transfer (person A to object 1 to person B to object 2) has also been demonstrated. Possible secondary and tertiary transfers are illustrated in Figure 11.

In Figure 11(a), person A touches the gun with primary transfer of A's DNA to the gun. In Figure 11(b), person A touches person B, who then touches the gun. It would be possible to see person B's profile, a mix of person A and person B, or just person A's profile on the gun. This secondary transfer is more likely if the contact from A to B and from B to the gun happens soon after each other. In Figure 11(c), person A touches the mobile phone, which is then touched by person B, who touches the gun. If person A's DNA profile were observed on the gun, this would have occurred by tertiary transfer. Although there is a low expectation of observing tertiary transfer, it is more likely if the contacts from A to the phone, from the phone to B and from B to the gun happen very soon after each other.

Questions as to how long after a transfer of material occurs can DNA related to that transfer still be recovered, or how much DNA will be transferred given a specific type of contact, are currently largely unknown. Each transfer possibility is dependent on the specific circumstances of the alleged activity and, as such, in looking at transfer scenarios, the scientist would rank possibilities rather than saying activity A is true and activity B is false. Because each case is different, there is not always directly applicable research or data related to each specific set of circumstances. The published research addresses different questions, and the experiments have been carried out in different ways, so it is not always possible to compare them directly. However, the published scientific research includes the following general principles regarding the transfer and persistence of touch DNA:

1. It is not possible to determine when the DNA was deposited.
2. DNA could persist for many months depending on a range of variables.
3. Secondary (or tertiary) transfer can occur such that a person's DNA might be on an object they have never touched.
4. Secondary or tertiary transfer without also leaving the transferring person's DNA has been demonstrated but only when the transfers occur immediately after each other. Transfer of DNA remains the subject of continuing research.
5. With each transfer we would normally expect a loss of available DNA, but the quality of the DNA profile cannot rule out a particular type of transfer since the end result will always depend on the available starting material.

In some instances, the scientific findings cannot provide any assistance in assessing how or when DNA came to be present, but in other cases, considering specific case circumstances, a comparative assessment can be made between alternative explanations. The scientist, in carrying out such an evaluation, should state their assumptions clearly. Having knowledge of the specific circumstances of the case after the interpretation and comparison process has been concluded, is considered important and will facilitate the scientist in effectively evaluating these transfer scenarios, and such evaluations should be contextualised in the latest relevant research.

A 3.5 The current understanding of error rates in DNA

If a match is observed between a suspect and crime-scene evidence, then three possibilities exist: (1) the suspect deposited the sample, (2) the suspect did not provide the sample but has the profile by chance, and (3) the suspect did not provide the sample and the matching result is a false positive due to a sample switch or some other kind of error or transfer. Genotyping errors (such as can occur from analysing very small traces of biological material) can also lead to imperfect matches.

TABLE 3

Table 3: Error rates in submission of samples to the National DNA Database. A year's data have been considered in compiling these figures.

Source of error	General frequency of occurrence (as a proportion of samples processed)
Police force handling of suspect sample or suspect sample records	Less than 1 in 2,000
Police force handling of crime sample or crime sample records	Less than 1 in 50,000
Forensic laboratory handling of suspect sample or suspect sample records	Less than 1 in 10,000
Forensic laboratory handling of crime sample or crime sample records	Less than 1 in 5,000
Forensic laboratory error in known sample DNA profile interpretation*	Less than 1 in 10,000
Forensic laboratory error in the interpretation of mixed profile crime samples DNA*	Less than 1 in 500

* Each error in the laboratory interpretation error rates quoted refers to a single part of the DNA profile only. A full DNA profile in England, Wales and N. Ireland consists of 32 numbers and 'XX' for a female or 'XY' for a male. Each of these errors would relate to only one of the 34 alphanumeric values, with the remaining 33 being correct.

Many quality assurance measures are in place to prevent or reduce the possibility of error in performing DNA testing. All laboratories analysing DNA for evidential purposes must comply with stringent quality standards. Each is externally assessed at least annually to ensure they comply with the international standard set by the Forensic Science Regulator and must declare if they are not compliant. However, errors can still occur, as in any process where there is an element of human intervention. There are quarterly checks of the quality of DNA profiling laboratories submitting to the National DNA Database. This enables an estimate of general error rates to be made (Table 3). These are errors that have been detected through the systems and processes designed for that purpose. For example, a 'near miss' check is run regularly on the National DNA Database, to ensure that any profiles that are extremely similar but differ in a single designation are identified. It is important to note that the error rates in Table 3 are for submission of samples to the National DNA Database; in a case coming to court, additional quality checks are made during and after the comparison between the suspect's sample and the crime sample, both by the scientist reporting the results to the court and by a second scientist. The error rates in Table 3 are therefore higher than would be expected for cases coming to court. In any particular case, the important question is whether an error was made in that case – a realistic suspicion of error in a case can be explored in more detail by examination of the records and quality controls in that case.

Appendix 4: Some case examples

Body fluid attribution

Example 1: The complainant/complainer has been stabbed and a DNA profile has been obtained from a large area of heavy staining that has the appearance of blood, on a suspect's clothing. The forensic scientist carried out a test to check if the stain on the clothing was consistent with being from blood. Although this test is not perfect (there is a possibility of false positives from other substances), it gave a very strong indication that the stain was in fact blood. A mixed DNA profile was obtained from the stained area, of which the major contribution matched the complainant/complainer and the minor contribution matched the suspect. In this example, it is reasonable to assume that the major component of the DNA profile was from the heavy blood staining and that the minor component might have been from the habitual wearer of the clothing.

Example 2: An allegation of rape has been made by a complainant/complainer. The suspect claims not to have had intercourse with the complainant/complainer, but only to have touched her external genital region. He says he masturbated two hours previous to the alleged incident. A swab taken from high within the vagina of the rape complainant/complainer has been examined for semen, and a significant number of sperm heads were visually observed. A process whereby sperm are separated from all other cells has been carried out, and the DNA extracted from the sperm fraction matches the suspect. The DNA from the other cells is a mixture, with the major part matching the complainant/complainer. Because the sperm were chemically separated from the other cells and because a significant number of sperm heads were observed, it is possible to say with confidence that the DNA extracted from this fraction was from sperm. Because the swab was from high within the vagina, the scientific findings would be highly unlikely if the defendant merely touched the complainant/complainer after having masturbated earlier. The scientific findings are much more probable if sexual intercourse with ejaculation into the vagina occurred than if the external genital area was touched by hand after masturbation.

DNA transfer

Example 3: A knife has been recovered, which might have been used in a stabbing. The blade has been cleaned and there is no visible bloodstaining. The forensic scientist carried out a test for blood, but found none. The surface of the blade was swabbed to sample for DNA, and the handle of the knife was swabbed separately. The swab from the blade produced a weak DNA profile matching the complainant/complainer, and the swab from the handle produced a weak mixed DNA profile, of which the major component matched the suspect. In this example, it cannot automatically be assumed that the complainant's/complainer's profile from the blade originated from blood. Questions such as 'where was the knife found?' become highly relevant: if it was a kitchen knife from the complainant's/complainer's house, then his DNA could have been there because he had handled the knife recently, and not because it was used to stab him. If the suspect had previously had access to the complainant's/complainer's kitchen (eg if the defendant also lived there or was a regular visitor), then the finding of his DNA on the handle could be explained by contact with the knife at some time in the past. If it was a kitchen knife from the suspect's house, then the finding of his DNA on the handle is to be expected, but the finding of the complainant's/complainer's DNA might or might not have relevance, depending on whether the suspect and complainant/complainer had had previous contact and whether the complainant/complainer had been at the defendant's house.

Example 4: An illegal firearm is found wrapped in a plastic bag at a lock-up rented by the suspect. The prosecution alleges it is the suspect's gun, but he claims to have no knowledge of it. The trigger of the gun is swabbed; this is chosen for swabbing because it is to some extent protected from accidental contact by the trigger guard, and because it would be expected to be an area of the gun that would be touched by a person using the gun. This yields a low-template DNA profile matching the defendant, with no other contributing DNA from any other person. The handle and barrel of the gun were also swabbed but no profiles were obtained. The suspect claims that the plastic bag was his and that the DNA must have been transferred from him to the bag and from the bag to the gun. In assessing the scientific findings, the scientist will consider the current level of knowledge regarding transfer and persistence of DNA, and the physical transfers that would need to take place under each scenario. The steps and considerations are summarised in Table 4.

TABLE 4

Transfer stages for DNA to be detected on gun trigger.

Transfers	Suspect's scenario	Prosecution scenario
1	DNA is transferred from the defendant to the bag (this will not be all over the bag, but at the points of contact). To assess this scenario fully, multiple areas of the bag were swabbed and profiled but the suspect's DNA was not detected other than on the handles.	DNA is directly transferred from the defendant to the trigger of the gun
2	The DNA on the bag must line up with the trigger of the gun, and get between the trigger guard and the trigger. The DNA must then transfer from the bag to the trigger. This would happen more easily if the DNA were in a moist state, such as just after it was deposited.	

In this example, the simplest transfer would be direct transfer to the gun. The suspect's scenario is also possible, but there are additional requirements for it to occur (the right part of the bag managing to get between the trigger guard and the trigger etc).
The scientist would therefore be likely to give an opinion that the findings are more likely if the prosecution scenario is true than if the defence scenario is true.

Appendix 5: Glossary

Accuracy: the degree of agreement or conformity of a measured value with its actual true) value.

Allele: one of two or more versions of a genetic sequence at a particular location (locus) in the genome.

Allele drop-in: allelic peak(s) in an electropherogram that are not reproducible across multiple independent analyses; also one hypothesis used to explain the observation of one or more allelic peaks in an electropherogram that are inconsistent with the assumed/known contributor(s) to a sample.

Allele drop-out: failure of an otherwise detectable allele to produce a signal above the analytical threshold because the allele was not present or was not present in sufficient quantity in the DNA sample.

Allele frequency: the number of times that an allele appears in a data set; the proportion of a particular allele in a population.

Allelic ladder: in STR testing, a measurement calibration tool, consisting of the most commonly observed alleles, used for assigning an allele designation to a peak in an electropherogram at a particular genetic locus.

Amelogenin gene: located on the X and Y chromosome used in the determination of sex from a DNA sample.

Base: a chemical unit within DNA that forms part of its structure. There are four bases that are linked together to make up the long strands of the DNA helix: adenine, thymine, cytosine and guanine, known as A, T, C, G respectively.

Base pair: two complementary bases on opposite strands of the DNA double helix joined by chemical bonds called hydrogen bonds; base pairing occurs between A and T and between G and C.

Billion: one thousand million.

Buccal swab: a relatively non-invasive technique of scraping the inside of a mouth with a cotton swab or similar collection device to collect cells from the inner cheek lining; a common method for collecting and preserving samples for DNA testing from known individuals.

Cell: the basic building block of an organism; humans have approximately 100 trillion cells in their body, most containing DNA.

Chromatid: thread-like strand into which a chromosome divides longitudinally prior to cell division. Each contains a double helix of DNA.

Chromosome: long continuous strand of DNA found in the nucleus of cells.

Complainant/complainer: Terminology used to describe the person who instigates a criminal investigation within the legal framework. The latter is used in Scotland, while the former is used in the rest of the UK.

Deoxyribonucleic acid (DNA): a genetic material of organisms, usually double-stranded; composed of large chemical molecules called nucleic acids composed of smaller chemical molecules called nucleotides identified by the presence of deoxyribose, a sugar, and four chemical bases; DNA is a fairly stable molecule, and variations in DNA sequence between individuals permits DNA profiling to distinguish individuals from one another.

Detection limit: the smallest amount of some component of interest that can be measured by a single measurement with a stated level of confidence.

DNA database: a computer repository of DNA profiles.

DNA degradation: the fragmentation, or breakdown, of DNA by chemical, physical, or biological means; a common occurrence when biological samples containing DNA encounter warm moist environments or excessive UV light.

DNA profile: a string of values (numbers or letters) compiled from the results of DNA testing at one or more genetic loci; a count of the STR lengths contributed from the maternal and paternal copies of DNA at each locus tested; can be from a single source or a mixture from multiple contributors.

Double helix: the native form of DNA, which looks like a twisted ladder; two linear strands of DNA assume this shape when held together by complementary base pairing, analogous to the rungs on the twisted ladder.

Electropherogram: a graphical representation of a DNA profile, where the horizontal axis represents the size of the DNA fragments analysed and the vertical axis represents the relative abundance of the DNA fragments analysed.

Gene: the basic unit of heredity; a sequence of DNA nucleotides on a chromosome passed from parents to offspring that influences various traits.

Genetics: branch of biology that deals with the heredity.

Genome: the entire DNA sequence found in a cell; the human genome consists of approximately 6,500,000,000 pairs of bases.

Genotype: the genetic make-up of an organism as characterised by its DNA sequence. With STR DNA testing, a locus genotype generally consists of two alleles, inherited from an individual's mother and father.

Haplotype: a group of genes or DNA sequences inherited together from one parent.

Likelihood ratio: the probability of the evidence under one proposition divided by the probability of the evidence under an alternative, mutually exclusive proposition; the magnitude of its value expresses the weight of the evidence. A larger likelihood ratio occurs if the 'top' scenario is the more likely to have occurred.

Loci: plural of locus.

Locus: a unique physical location of a gene (or a specific sequence of DNA) on a chromosome; In Scotland the 'locus' is the name given to a crime scene.

Low Copy Number (LCN) DNA testing: the analysis of a small quantity of DNA often conducted by increasing the number of PCR amplification cycles.

Low-level or **low-template DNA:** usually defined as less than approximately 100 picograms (pg) or about the amount in 15 human cells.

Major profile: The profile derived from the predominant DNA source in a mixed sample.

Match probability: a conditional probability used to address the question 'given that a particular DNA profile has been generated from evidence related to a case and an identical DNA profile has been generated from a sample taken from the person of interest, what is the chance of the same DNA profile also being generated from another person at random?'

Matching profile: genetic profiles that show the same alleles at all loci tested and with unexplainable differences.

Mitochondrial DNA (mtDNA): a small, circular DNA molecule located in the mitochondria of a cell that consists of approximately 16,500 base pairs; the presence of hundreds of copies of mtDNA in each cell make it useful for analysing samples originating from limited or damaged biological material.

Multiplex PCR: co-amplification of multiple regions of a genome enabling information from the different target sequences to be collected simultaneously.

Mutation: any change in DNA sequence.

Partial profile: a DNA profile for which complete results are not obtained at all tested loci.

Polymerase chain reaction (PCR): an *in vitro* process that yields millions of copies of the desired DNA through repeated cycling of a reaction involving the DNA polymerase enzyme.

Precision: a measure of the closeness of results when experiments are repeated.

Probabilistic genotyping: use of statistical modelling informed by biological data, statistical theory, computer algorithms and/or probability distributions to infer genotypes and/or calculate likelihood ratios.

Pull-up: an artefact that may occur during analysis of fluorescently labelled DNA fragments when signal from one dye colour channel produces artificial peaks in another, usually adjacent colour, at a similar position on the horizontal axis in an electropherogram; sometimes referred to as bleed-through or spectral calibration failure.

Short tandem repeats (STR): multiple copies of an identical (or similar) DNA sequence arranged in direct succession where the repeat sequence unit is between two base pairs and six base pairs in length. The number of repeat units can vary between individuals.

Stutter product: a minor peak primarily appearing one repeat unit smaller than the primary STR allele; this results from strand slippage during the amplification process; usually <15% of the height of the true allele peak.

Touch DNA: DNA that is transferred to or from surfaces via contact.

Validation: The process of providing objective evidence that a method, process or device is fit for the specific purpose intended.

Weight of evidence: refers to either match probability, likelihood ratio or exclusionary evidence.

X chromosome: one of the sex chromosomes; normal females possess two copies and males one copy.

Y chromosome: one of the sex chromosomes; normal males possess one copy and females none

The members of the groups involved in producing this primer are listed below. The members acted in an individual and not organisational capacity and declared any conflicts of interest. They contributed on the basis of their own expertise and good judgement. The Royal Society and the Royal Society of Edinburgh gratefully acknowledges their contribution.

Primer leads

Lady Justice Anne Rafferty DBE

Professor Niamh Nic Daéid FRSE

Editorial board

His Honour Judge Paul Farrer QC

Sir Alec Jeffreys CH FMedSci FRS

Sir Paul Nurse FMedSci FRS

Professor Veronica van Heyningen CBE FMedSci FRSE FRS

Writing Group

Dr John Butler

Professor James Chalmers

Professor Gilean McVean FMedSci FRS

Dr Gillian Tully

Acknowledgments

This project would also not have been possible without contributions and support from a range of individuals. In particular we wish to thank:

The Rt Hon Lord Thomas of Cwmgiedd, Lord Chief Justice of England and Wales, 2012 – 2017

The Rt Hon Lord Carloway, Lord President of the Court of Session and Lord Justice General

Sir John Skehel FMedSci FRS, Vice President and Biological Secretary of the Royal Society

Primers Steering Group

Lord Hughes of Ombersley (Chair)

Professor Dame Sue Black DBE FRSE

Lord Bracadale QC

Dr William Duncan

Sir Charles Godfray CBE FRS

Dame Ottoline Leyser DBE FRS

Dr Julie Maxton CBE

Professor Angela McLean FRS

Professor Niamh Nic Daéid FRSE

Lady Justice Anne Rafferty DBE

Sir Muir Russell KCB FRSE

His Honour Judge Mark Wall QC

The Royal Society is a self-governing Fellowship of many of the world's most distinguished scientists drawn from all areas of science, engineering, and medicine. The Society's fundamental purpose, as it has been since its foundation in 1660, is to recognise, promote, and support excellence in science and to encourage the development and use of science for the benefit of humanity.

The Society's strategic priorities are:

- Promoting excellence in science
- Supporting international collaboration
- Demonstrating the importance of science to everyone

For further information
The Royal Society
6 – 9 Carlton House Terrace
London SW1Y 5AG

T +44 20 7451 2571
E law@royalsociety.org
W royalsociety.org/science-and-law

Registered Charity No 207043

The Royal Society of Edinburgh (RSE), Scotland's National Academy, is a leading educational charity which operates on an independent and non-party-political basis to provide public benefit throughout Scotland. Established by Royal Charter in 1783 by key proponents of the Scottish Enlightenment, the RSE now has around 1600 Fellows from a wide range of disciplines. The work of the RSE includes awarding research funding, leading on major inquiries, informing public policy and delivering events across Scotland to inspire knowledge and learning.

For further information
The Royal Society of Edinburgh
22 – 26 George Street
Edinburgh EH2 2PQ

T +44 131 240 5000
E info@theRSE.org.uk
W rse.org.uk

Scottish Charity No SC000470

ISBN: 978-1-78252-301-7
Issued: November 2017 DES4928