nature reviews genetics

https://doi.org/10.1038/s41576-025-00896-1

Review article

Check for updates

Forensic genetics in the omics era

Manfred Kayser

Abstract

Recent advances in forensic genetics, driven by technological innovation coupled with the use of an expanding range of nucleic acid markers, have markedly improved the scope, accuracy and reliability of evidential information obtainable from human biological traces recovered at crime scenes. The majority of these biomarkers have been identified using non-targeted omics approaches, including genomics, transcriptomics, epigenomics and microbiome profiling. Moreover, targeted massively parallel sequencing, in some cases non-targeted whole-genome sequencing, are being applied to the analyses of biological trace material. These approaches and methods are being used for the identification of perpetrators (including monozygotic twins), their relatives or victims of criminal activities; the prediction of phenotypic and behavioural traits of unknown individuals; and the determination of trace characteristics, including tissue type and time of deposition.

Sections

Department of Pathology and Clinical Bioinformatics, Erasmus MC University Medical Center Rotterdam, Rotterdam, The Netherlands. e-mail: m.kayser@erasmusmc.nl

The author dedicates this paper to his colleagues from the former Department of Genetic Identification in recognition of their contributions to the forensic genetic innovations we achieved together over a period of more than 20 years, from its inception as the Department of Forensic Molecular Biology in 2004 until its closure as result of the reorganization of biomedical sciences at Erasmus MC in 2025.

Introduction

Three questions are key for assisting in solving criminal cases by establishing evidential information from human biological traces collected at crime scenes. Who did it? What happened? When did it happen? For 40 years, the analyses of nucleic acid markers in trace material from blood, saliva, semen, vaginal secretion, hair, skin and mixtures thereof have been used to help answer these questions in the field of forensic genetics. For the first question (Who?), the focus remains on DNA-based identification of individuals, mostly trace donors in cases with available suspects, but also victims of criminal activities and missing persons from unidentified human remains, and for the resolution of familial relationships. In cases without suspects, DNA prediction of physical appearance, biogeographic ancestry and age, as well as DNA-based searches for relatives, can help focus police investigations to find potential suspects. The second question (What?) can be addressed by determining the tissue type of a trace with the help of nucleic acid markers. To answer the third question (When?), biomarkers can help estimate the time a trace was deposited at the crime scene. Notably, the degree to which these questions have been investigated, the number of forensic cases they can potentially be applied to and whether they are indeed applied in forensic practice varies.

The field of forensic genetics was born in 1985 with the publication of the milestone discovery made by Sir Alec Jeffreys of individual-specific multi-locus DNA fingerprints[1] and its first application to a forensic case[2]. With the advent of the polymerase chain reaction (PCR), forensic DNA profiling based on autosomal short tandem repeats (STRs), also known as microsatellites, replaced minisatellite-based DNA fingerprinting in the early 1990s, and it remains the gold standard for individual identification today[3,4]. Mitochondrial DNA analysis from the late 1980s[5–7] and Y chromosome STR (Y-STR) analysis from the early 1990s[8,9] expanded forensic identification capabilities, particularly in cases in which autosomal STR profiles are not available, and additionally allowed users to infer maternal and paternal biogeographic ancestry, respectively. Assays based on autosomal single-nucleotide polymorphisms (SNPs), initially in HLA genes[5], were applied for individual identification in criminal casework and missing person cases from the late 1980s.

Over the past two decades, the types and numbers of nucleic acid markers applied in forensic genetics has increased continuously. Autosomal SNPs emerged for identifying individuals[10,11], predicting phenotypic traits[12,13] and inferring bi-parental biogeographic ancestry[14,15] from trace DNA. RNA and DNA methylation markers were developed to estimate a person's age based on technologies suitable for trace analysis[16,17]. RNA, DNA methylation and microbial DNA markers were used to determine the tissue type[18,19] and for trying to estimate the deposition time[20,21] of a trace. Most of these additional nucleic acid markers have been identified by non-targeted omics technologies, including SNP microarrays, gene expression microarrays, DNA methylation microarrays, whole-genome sequencing (WGS) and whole-transcriptome sequencing, some of which are also used directly for trace analysis. In addition, targeted omics technologies, such as targeted massively parallel sequencing (MPS), are increasingly applied for trace analysis (Box 1). These and other developments have broadened the spectrum, and increased the accuracy and reliability, of forensic information obtainable from human crime scene traces (Fig. 1).

This Review provides an overview on recent advances in forensic genetics. It highlights progress in individual identification of perpetrators, including monozygotic twins, and their relatives. Next, improvements in (epi)genetic prediction of externally visible characteristics of trace donors in cases without suspects, such as appearance traits, biogeographic ancestry, age and certain behavioural traits are summarized. Furthermore, an update on determining trace characteristics, such as tissue type and deposition time, is given. Throughout this Review, the use of non-targeted and targeted omics approaches and technologies based on nucleic acid markers is highlighted.

Identification of persons

Identifying the donor of a biological trace found at a crime scene is key to answering the Who? question. Trace DNA is mostly used for the individual identification of trace donors in cases with suspects, but it can also provide investigative leads that help find potential suspects in cases without suspects, or when all suspects have been excluded from being the trace donor by forensic STR profiling.

Individual identification

Individual identification is achieved by matching an autosomal STR profile obtained from a crime scene trace to a reference STR profile of a known person and providing weight of evidence for such a DNA match[22,23]. Over the past decades, the number of autosomal STRs used in forensic DNA profiling has steadily increased to achieve a very high standard of individual identification evidence. Current commercial human identification kits include more than 20 autosomal STRs[24] and DNA marker(s) to infer biological sex[25]. Forensic STR profiling relies on the person of interest being known with their STR profile to the investigating authorities, either due to being a case suspect or to being included in a national criminal offender DNA database owing to a previous conviction (in some jurisdictions also because of arrest). Since their first establishment three decades ago, criminal offender DNA databases have proven a powerful tool to link a trace STR profile to a possible donor, or to highlight case connections involving the same donor[26]. Over the years, the number of reference STR profiles in criminal offender DNA databases has grown exponentially, with database size directly proportional to the chance of finding matches[27]. Cross-country STR profile searches are practised between forensic databases of 29 EU Member States[26], regulated by the treaty of Prüm. Forensic STR profiling coupled with criminal offender DNA databases remain a staple of individual identification in forensics[26].

The sensitivity of current commercially available machinery and chemistry used for forensic STR profiling enables the generation of complete STR profiles from picogram amounts of trace DNA. This high sensitivity has led to a focus on touched objects for STR profiling[28], which has intensified research in two important areas: DNA transfer and DNA mixtures. More than 25 years ago, the field of 'touch DNA' was born with the first demonstration that DNA fingerprints can be obtained from physical fingerprints and of secondary transfer of DNA from one object to another[29]. Over the years, the understanding of transfer, persistence, prevalence and recovery of DNA (DNA-TPPR) has improved substantially[30,31]. DNA-TPPR is relevant to help address activity-level inquiries, that is, which activity most likely led to the presence of a person's DNA at the crime scene[32,33]. Bayesian networks have

Box 1 | Targeted versus non-targeted omics in forensic genetics

For forensic applications that are based on discrete marker sets preselected to be fit for purpose, targeted omics is most suitable. For example, targeted massively parallel sequencing (MPS) is best for forensic short tandem repeat profiling for individual identification, because the use of criminal offender DNA databases restricts and specifies the short tandem repeat markers applied in crime scene investigations. For forensic DNA phenotyping and forensic tissue identification, some biomarkers have larger effect sizes than others; hence, analysing biomarkers with larger effects by targeted MPS is most practical. Some targeted technologies are also less costly than non-targeted ones, providing economic advantages.

A major forensic advantage of targeted MPS technologies over other targeted DNA analysis methods is that they have a much larger multiplexing capacity, allowing many more markers to be analysed simultaneously. Multiplexing is important in forensic genetics, in which evidence material is often limited, so that as many markers as possible must be analysed simultaneously in one run. Increased multiplexing capacity is also relevant, because forensic marker sets are enlarging to improve various forensic applications. Moreover, targeted MPS assays often have higher sensitivity than non-targeted assays, and targeted MPS allows fully quantitative analysis, which is necessary in some forensic applications such as those based on DNA methylation markers. However, targeted MPS methods also have multiplex capacity limitations, more so for quantitative DNA methylation analysis than for qualitative short tandem repeat or SNP genotyping.

When the forensic information, or its accuracy, obtained with targeted MPS methods is limited, and it can be increased with additional and available biomarkers beyond the technology's multiplex capacity, non-targeted omics becomes the preferred approach. Depending on the forensic application, non-targeted omics technologies are applied on reference DNA without restriction on quality and quantity, or on trace DNA, if the quantity and quality of the trace allow. Future advances of non-targeted omics technologies may lead to increased sensitivity and/or decreased costs, which will benefit forensic applications.

become the preferred tools for analysing these data, because they help answer questions on the transfer mechanism in a probabilistic way[32,33].

Often, crime scene objects have been touched by multiple persons, requiring methods for analysing DNA mixtures, which can also arise from the combination of body fluids. Challenges with interpreting mixed STR profiles involve overlapping alleles, imbalanced contributions and allele drop-out (non-observation of existing alleles) or drop-in (observation of non-existing alleles) as a result of low quantities of DNA from minor contributions, the latter being similarly relevant for single-source low-quantity DNA samples[34,35]. To improve the interpretation of STR profiles obtained from mixed DNA samples, including low-input DNA, probabilistic genotyping methods have been developed and are applied in forensic casework based on qualitative (semi-continuous) models that take into account drop-in and drop-out, or quantitative (continuous) models that additionally consider signal intensity (peak height), PCR stutter and other parameters[36–40]. Alternative DNA markers are also investigated for forensic mixture deconvolution[41], such as Y-STRs for male–female mixtures; microhaplotypes[42] based on autosomal SNPs in close proximity, which avoids the STR problem of PCR stutter; or other types of compound DNA markers such as DIP–STRs[43], which pair an STR with a nearby insertion–deletion polymorphism. A different approach is to separate the cells of a mixed sample prior to forensic DNA profiling, for which different technologies have been explored, including laser capture microdissection[44] and DEParray[45] (Menarini Silicon Biosystems). Targeted DNA analysis of single cells is challenging because of the low amount of DNA a diploid human cell contains (~3.3 pg). More PCR cycles and additives in the PCR have been used to increase the sensitivity and probabilistic genotyping is applied to establish consensus DNA profiles from several single cells[46]. Recently, non-targeted single-cell sequencing, that is, single-cell transcriptome sequencing[47] and single-cell DNA sequencing[48], have been explored for SNP-based mixture deconvolution with subsequent biogeographic ancestry inference and individual identification of the separated contributors as well as the identification of the tissue type.

Most forensic DNA laboratories continue to profile STRs using fluorescence multiplex PCR followed by capillary electrophoresis, although some have moved to targeted MPS, for which commercial kits are available too. Because of its fully quantitative nature, targeted MPS benefits mixture analysis, amongst other advantages[49,50] (Box 1). The relatively short fragment length of most forensically used STRs makes short-read DNA sequencing suitable, whereas long-read DNA sequencing benefits long forensic STRs, can enhance mixture analysis[51] and provides other advantages, such as when combining different marker types[52]. Gene-editing technologies have recently started being explored for forensic STR profiling[53]; for instance, STR-seq uses CRISPR–Cas9 for fragmenting DNA to increase the number of sequenced STRs without PCR amplification[54].

Male identification

Male-specific Y-STRs have been applied in forensic genetics for more than three decades, particularly in cases in which autosomal STR profiles cannot be generated, such as in violent or sexual assault cases with male–female mixtures and a large excess of female victim DNA[8,9] (Fig. 2). Although differential lysis can enrich for sperm DNA, probabilistic genotyping software and targeted MPS can improve male–female mixture deconvolution. In cases in which the autosomal STR profile of the male contributor cannot be established from the mixed STR profile, a Y-STR profile can often be generated from the mixed trace, particularly when Y chromosome DNA is detected during quantification of the DNA extract by using DNA quantification kits that contain target-specific assays for human male DNA. Moreover, sperm cells in a trace may be present in low or no numbers, which prevents differential lysis; however, Y-STR analysis can be successfully performed on male epithelial cells. As with autosomal STR kits, the number of Y-STRs in commercial kits has increased over the past decades, with more than 25 Y-STRs in current kits[55]. Most of these are standard Y-STRs with relatively low mutation rates (in the order of 10^{-3} per locus per generation[56]), which makes these kits very powerful for identifying male lineages but not for separating male relatives and individual male identification.

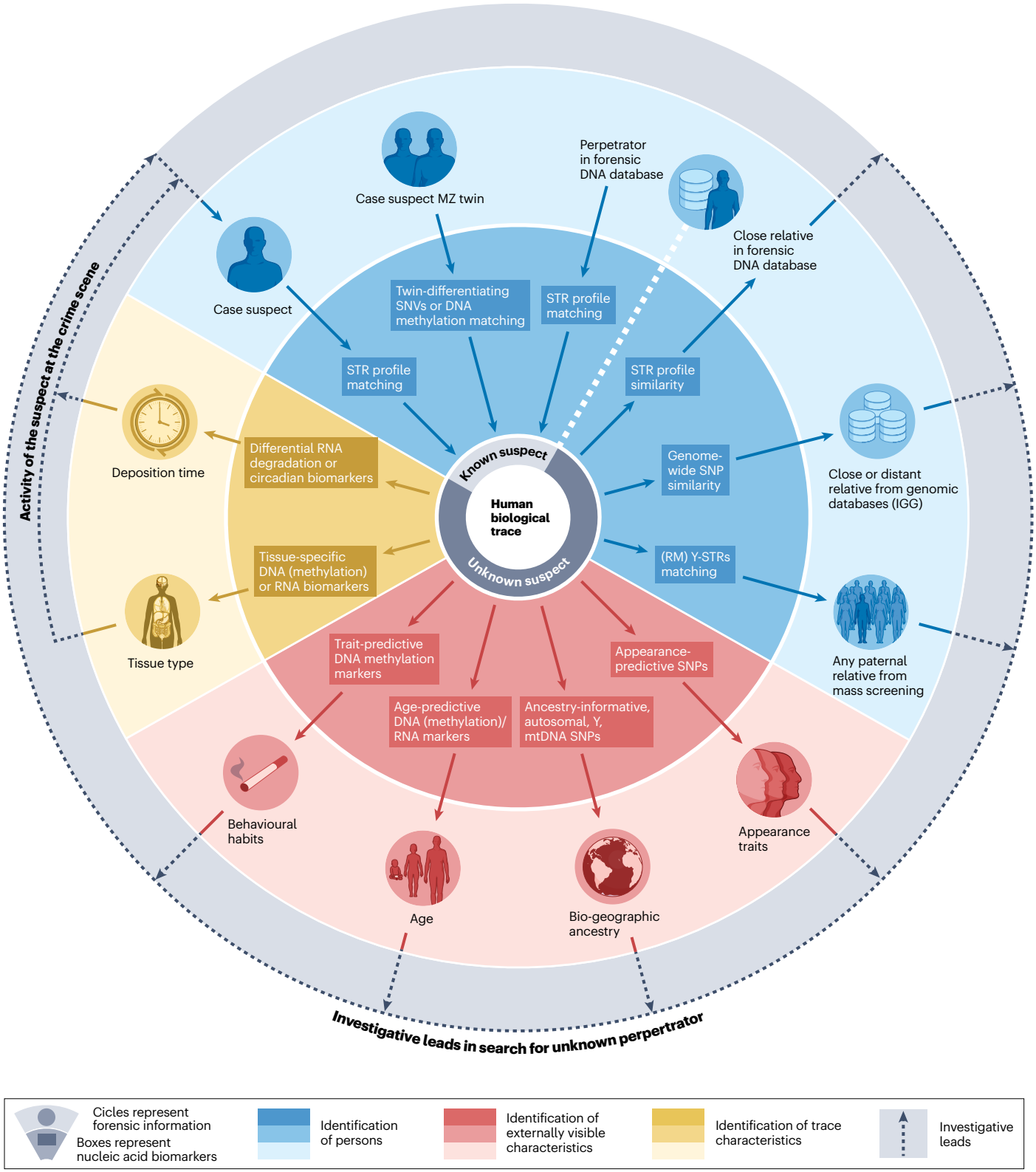
Activity of the suspect at the crime scene
Case suspect MZ twin
Perpetrator in forensic DNA database
Close relative in forensic DNA database
Case suspect
Twin-differentiating SNVs or DNA methylation matching
STR profile matching
STR profile matching
STR profile similarity
Known suspect
Human biological trace
Unknown suspect
Deposition time
Differential RNA degradation or circadian biomarkers
Genome-wide SNP similarity
Close or distant relative from genomic databases (IGG)
Tissue-specific DNA (methylation) or RNA biomarkers
(RM) Y-STRs matching
Tissue type
Any paternal relative from mass screening
Trait-predictive DNA methylation markers
Appearance-predictive SNPs
Age-predictive DNA (methylation)/ RNA markers
Ancestry-informative, autosomal, Y, mtDNA SNPs
Behavioural habits
Appearance traits
Age
Bio-geographic ancestry
Investigative leads in search for unknown perpetrator
Cicles represent forensic information
Boxes represent nucleic acid biomarkers
Identification of persons
Identification of externally visible characteristics
Identification of trace characteristics
Investigative leads

Fig. 1 | Overview of genetic and epigenetic analyses for crime scene investigation. Different types of genetic or epigenetic analyses of crime scene traces can be applied in cases with and without suspects for the identification of persons, externally visible characteristics and trace characteristics. Trace analyses in cases without suspects provide investigative leads to help find potential case suspects (indicated by arrows). Trace analyses in cases with suspects provide individual identification (indicated by arrows). Trace analyses for the identification of trace characteristics provide information on the activity of the DNA-identified trace donor at the scene of crime (indicated by arrows). IGG, investigative genetic genealogy; mtDNA, mitochondrial DNA; MZ, monozygotic; RM Y-STRs, rapidly-mutating Y chromosome short tandem repeats; SNV, single-nucleotide variant; STR, short tandem repeat.

In contrast to autosomal STRs, the so-called product rule of multiplying single-locus allele frequencies to estimate match probabilities cannot be used for Y-STRs, given their localization on the non-recombining part of the Y chromosome. In current forensic practice[57], Y-STR match probabilities are equated with Y-STR haplotype frequencies obtained from population reference databases, most notably the publicly accessible Y-Chromosome STR Haplotype Reference Database (YHRD)[58]. However, in principle, and because of their limitations in size, geographic coverage and sample selection to avoid any relatives, population databases cannot be regarded as representative of all plausible alternative suspects in every single case[59]. Recently, a mathematical framework for estimating Y-STR match probabilities based on the suspect's pedigree was developed[59], which obviates frequency estimates from population databases; a respective software tool is currently in development.

Moving from male lineage to male individual identification is key in improving forensic Y chromosome analysis. Over the past 15 years, rapidly-mutating Y-STRs (RM Y-STRs) with an about tenfold higher mutation rate than standard Y-STRs were identified[60,61] and developed as markers to differentiate male relatives[62,63]. Although commercial Y-STR kits include only a few RM Y-STRs, several non-commercial RM Y-STR genotyping assays have been developed. The currently most comprehensive tool, RMplex, targets 30 Y-STRs with increased mutation rates, including all currently known 26 RM Y-STRs[64]. RMplex-based Y-STR typing achieved average male relative differentiation rates of 43% for father–sons, 66% for brothers, 76% for cousins, >95% for relatives separated by ≥6 meiosis and 100% for relatives separated by ≥12 meiosis[65]. As (Y-)STR mutation rates increase with increasing repeat numbers[66], some RM Y-STRs exceed the maximal read lengths of short-read sequencing technologies, making long-read sequencing interesting for RM Y-STR analysis. Recently, WGS was applied for Y chromosome-based male relative differentiation[67], yielding an observed father–son separation rate of 58% based on Y chromosomal single-nucleotide variants (SNVs), although this finding needs to be confirmed by sequencing more father–son pairs. The expected — as estimated based on the Y chromosomal SNV mutation rate — male relative differentiation rates are 69% for father–sons, 90% for brothers and 97% for cousins[67]. Moving to WGS with long-read sequencing is expected to further increase male relative-differentiation rates.

Relative identification

The identification of relatives of trace donors in cases without suspects by so-called familial searching can be achieved with different approaches, marker types and technologies. Relatives can be identified with autosomal forensic STR profiling, if they are included in the national criminal offender DNA database[68]. However, the success of this approach is restricted to very close relatives one or two generations apart, such as parent–offspring or sibling relationships. In cases with a male trace donor, as mentioned above, male relatives can be identified via Y-STR analysis through mass screenings[9], in which standard Y-STRs are used first for male lineage identification and, if matches are observed, this initial step is followed by RM Y-STR analysis to differentiate close from distant male relatives (Fig. 2). In some countries, Y-STR profiles are included in criminal offender DNA databases, enabling male relative identification through database searches[69].

To overcome the marker limitation of forensic autosomal STRs for familial searching, tools based on thousands of autosomal SNPs were developed based on targeted MPS with and without capture enrichment and have started to be applied in forensic practice for familial searching[70] and missing person identification[71]. Moreover, if the trace DNA quality and quantity are high enough, non-targeted genomics approaches such as WGS or SNP microarrays can be used to find relatives up to third cousins[72], or even more distant ones[73], if they are included in genomic databases. This can provide investigative leads via DNA-motivated genealogy research, a process referred to as investigative genetic genealogy, forensic genetic genealogy or forensic investigative genetic genealogy[74]. As these consumer DNA databases were not established for forensic purposes, proprietary companies have implemented opt-in or op-out procedures for their customers to regulate access by law enforcement agencies[75]; conversely, some consumer genomics companies do not allow forensic applications at all. Many hundreds of cases have been solved with this approach to date, according to the Mendeley database, mostly in the USA, where these companies are based and most of their customers live[76,77], but also in other countries, such as Sweden[78] and Norway[79]. Because poor input DNA quality and quantity generate errors in SNP microarray data[80], which can affect downstream applications such as relative identification[80], careful selection of trace samples based on suitable DNA quality and quantity is crucial.

Monozygotic twin identification

Monozygotic twins typically share the same forensic STR profile; hence, STR profiling is not informative in criminal cases involving one or both monozygotic twins. Various molecular approaches have been investigated for monozygotic twin differentiation in the forensic context[81]; among them, the genomic and the epigenomic approaches are most notable. In both, non-targeted omics is used to identify twin-differentiating DNA markers in the reference DNAs of both twins that are subsequently analysed in the trace DNA with targeted methods, such as targeted MPS.

The genomics approach, which has been applied successfully to several forensic cases[82,83], is based on somatic mutations (typically autosomal SNVs) that occurred in the zygote at an early stage of embryonic development. As a result, the mutant allele is present in one but not in the other twin. The number of twin-differentiating SNVs varies between pairs. For instance, a deep-sequencing study of 381 monozygotic twin pairs showed that they differed on average by 5.2 early-developmental mutations, and 15% of twin pairs had a substantial number specific to one of the twins[84]. These mutations may not be present in all cells and not in all tissues (somatic mosaicism), so their identification requires ultra-deep WGS (for example, 100x coverage). Moreover, depending on the cell types of reference and trace DNAs, the mutant allele identified in

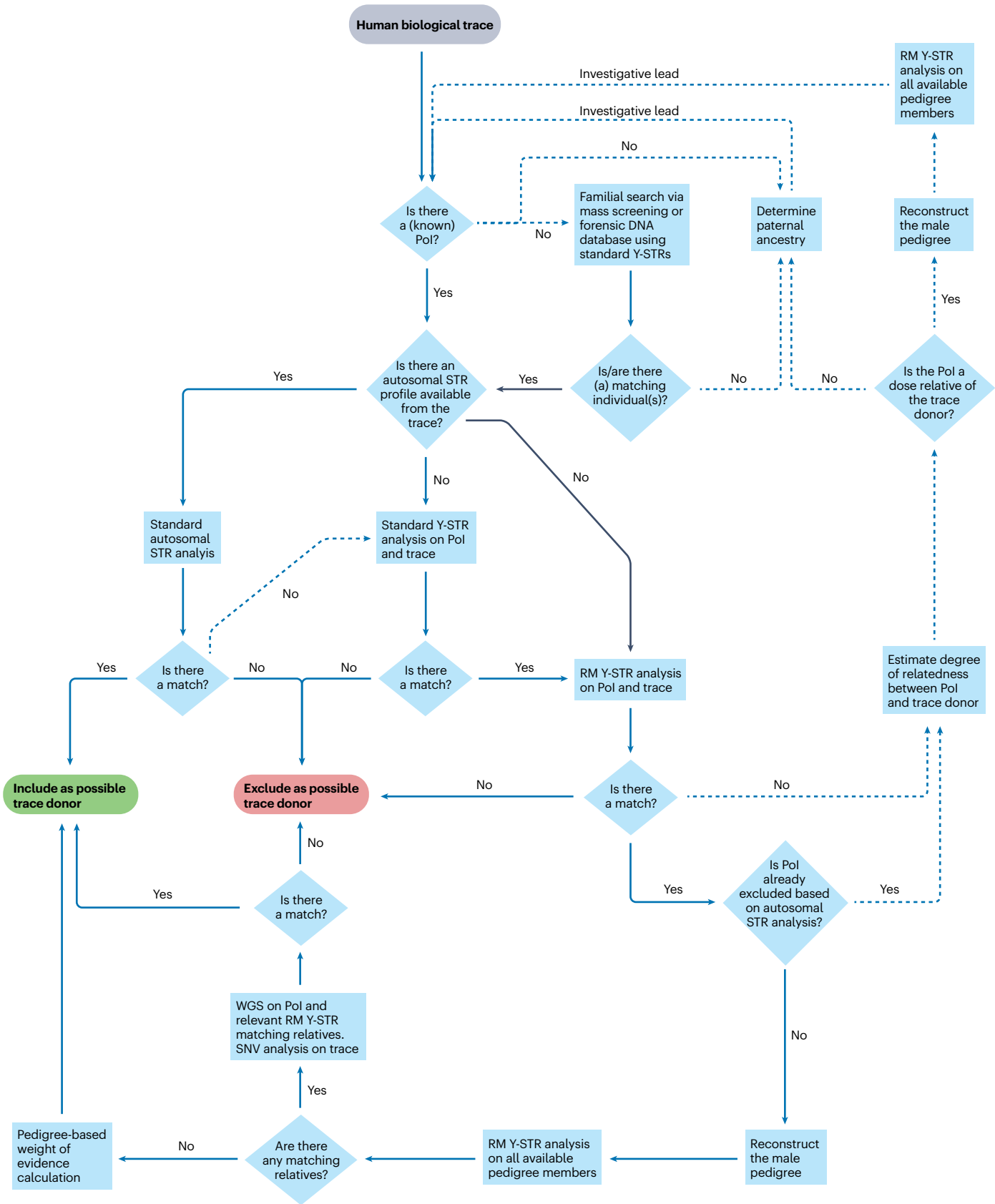
Human biological trace
Investigative lead
RM Y-STR analysis on all available pedigree members
Investigative lead
No
Is there a (known) PoI?
No
Familial search via mass screening or forensic DNA database using standard Y-STRs
Determine paternal ancestry
Reconstruct the male pedigree
Yes
Yes
Is there an autosomal STR profile available from the trace?
Yes
Is/are there (a) matching individual(s)?
No
No
Is the PoI a dose relative of the trace donor?
Yes
No
No
Standard autosomal STR analyis
Standard Y-STR analysis on PoI and trace
No
Yes
Is there a match?
No
No
Is there a match?
Yes
RM Y-STR analysis on PoI and trace
Estimate degree of relatedness between PoI and trace donor
Include as possible trace donor
Exclude as possible trace donor
No
Is there a match?
No
No
Yes
Is PoI already excluded based on autosomal STR analysis?
Yes
Yes
Is there a match?
WGS on PoI and relevant RM Y-STR matching relatives. SNV analysis on trace
No
Yes
Pedigree-based weight of evidence calculation
No
Are there any matching relatives?
RM Y-STR analysis on all available pedigree members
Reconstruct the male pedigree

Fig. 2 | Overview of Y chromosome analyses for crime scene investigation with decision network. Different types of Y chromosome analyses on traces, as applied in cases with male trace donors and no autosomal short tandem repeat (STR) profiles being available, for the identification of the paternal lineage of a male suspect, that is, person of interest (PoI), for the differentiation of a male suspect from his paternal male relatives, for the individual identification of a male suspect being the trace donor, and for providing investigative leads to help find potential case suspects in cases without known suspects. Solid blue arrows lead towards individual male identification. Dotted blue arrows lead towards paternal lineage identification and investigative analyses. Solid black arrows lead towards individual male identification after successful investigative analyses. RM Y-STR, rapidly-mutating Y chromosome STR; SNV, single-nucleotide variant; WGS, whole-genome sequencing.

the reference DNA may not be present in the trace DNA[85]. Choosing the tissue type of the reference samples based on the trace tissue type avoids this problem, but it can provide challenges if the reference material is regulated by law. If the twin-differentiating mutation created an extra SNV allele — that is, one twin being homozygous, the other heterozygous — its detection in the reference and trace DNA is straightforward. If the mutation created imbalanced heterozygotes, that is, both twins are heterozygous but with opposite major alleles, identification requires specialized bioinformatic analysis of the WGS data, and detection of the mutant allele in trace DNA needs a targeted method for quantitative SNP analysis, such as targeted MPS. A recently reported twin case demonstrated the success of the genomic approach despite various challenges provided by the case circumstances and the trace material[86].

In the epigenomics approach, monozygotic twin differentiation is achieved by quantifying DNA methylation differences between the two twins[87,88]. Reference DNA samples of both twins are analysed by DNA methylation microarrays to identify twin-differentiating CpGs, which are then investigated in the trace DNA using targeted methods, such as targeted MPS. Due to method-to-method bias, the number of twin-differentiating CpGs usable for trace DNA analysis is typically smaller than those identified in the reference DNA samples, as was demonstrated in different tissue types[89,90]. As DNA methylation patterns can differ between tissues, it is important to match the tissue type of the trace and reference samples. Theoretically, the epigenomics approach can identify twin-differentiating markers in every monozygotic twin pair, in contrast to the genomics approach. However, in practice, several factors make the epigenomics approach less appealing for forensic casework than the genomics one, provided that the high expenses of ultra-deep WGS can be covered. These factors include the technical noise in quantitative CpG analysis, the need for more trace DNA for bisulfite conversion and the resulting DNA degradation, as well as the between-tissue differences of DNA methylation variation. Universal twin-differentiating CpGs applicable to most, if not all, monozygotic twin pairs, together with a targeted method suitable for trace analysis, could change the potential of the epigenomic approach in the future. That such CpGs exist is indicated by the discovery of hundreds of CpGs with equally large variation in unrelated individuals as in monozygotic twins[91].

Identification of externally visible characteristics

Predicting a trace DNA donor's physical appearance, biogeographic ancestry, age and certain behavioural traits — a process known as forensic DNA phenotyping (FDP) — is relevant in criminal cases without suspects, in which, by providing investigative leads, FDP can help find potential suspects through focused police investigation[92]. FDP generally requires two parts, a forensically validated DNA tool for analysing predictive DNA markers based on a technology suitable for trace analysis, and a prediction tool based on a validated statistical prediction model. To increase its effectiveness, FDP can be combined with other investigative DNA analyses. For instance, depending on the rarity of its outcome in the population, FDP can reduce the number of persons invited for a voluntary DNA mass screening for familial searching. This combined approach has been used to solve cold cases of rape and murder, for example, in the Netherlands[93]. FDP is an investigative approach; final identification of the trace donor is always done via forensic STR profile matching.

Appearance prediction

Over the past 15 years, the science and technology of predicting eye, hair and skin colour from trace DNA has been fully established, and it is applied in forensic casework in many countries[92]. Although all appearance traits are genetically complex with hundreds or more underlying genes, pigmentation traits have been associated with a few genes with enlarged effect sizes, which harbour SNPs with increased prediction power. Therefore, DNA prediction for eye, hair and skin colour is achievable with small sets of SNPs, most of them identified in genome-wide association studies (GWAS), for instance, six SNPs for eye colour[94,95], 22 for hair colour[96,97] and 36 for skin colour[98,99]. Statistical prediction models based on these sets of SNPs in thousands of individuals achieved categorical prediction accuracies expressed as area under the receiver operating characteristic curve (AUC) of 0.74–0.94 for eye[95,99], 0.72–0.92 for hair[97,99] and 0.72–0.96 for skin colour[99]. Initial DNA tools for eye, hair and skin colour prediction were based on primer extension and capillary electrophoresis[95,97,100–102], whereas in recent years, targeted MPS assays have been developed and are being applied[92,103]. Two web tools for DNA-based eye, hair and skin colour prediction are publicly available and used in forensic casework (and anthropological research): HIrisPlex-S and Snipper. HIrisPlex-S allows the upload of partial SNP profiles encountered when the trace DNA is degraded.

Recently, enhanced hair and eye colour prediction models based on increasing numbers of SNPs identified by non-targeted genomics approaches have been introduced. For hair colour, more than 250 SNPs discovered in a large GWAS[104] improved prediction, albeit with a minor AUC increase of 0.01–0.08 compared with 18 HIrisPlex SNPs[97,104]. For eye colour, more than 100 SNPs partly identified via whole-exome sequencing improved the prediction of intermediate eye colour with an AUC of 0.85 (ref. 105), which compared with 0.74 achieved with the 6 IrisPlex SNPs[95,99]. Prediction models were also developed for additional pigmentation traits such as freckles, with AUCs of 0.66–0.79 (ref. 106), eyebrow colour 0.62–0.7 (ref. 107) and for hair greying 0.79–0.87 (ref. 108), where, however, most prediction was provided by age alone. For these markers and models, DNA tools suitable for trace analysis and publicly accessible prediction tools have not been made available to date.

The genetic prediction of non-pigmentation appearance traits requires more SNPs for accurate prediction. However, large GWAS are currently only available for stature. Therefore, the prediction accuracy for non-pigmentation appearance traits is currently lower than those for pigmentation traits. For example, models for hair shape, based on >30 SNPs, and male pattern baldness, based on >100 SNPs, provided AUCs of 0.59–0.68 (albeit higher in non-Europeans) and 0.6–0.73, respectively[109,110], with similarly low AUCs reported for other hair traits[111]. For extremely tall stature, a model based on 689 SNPs

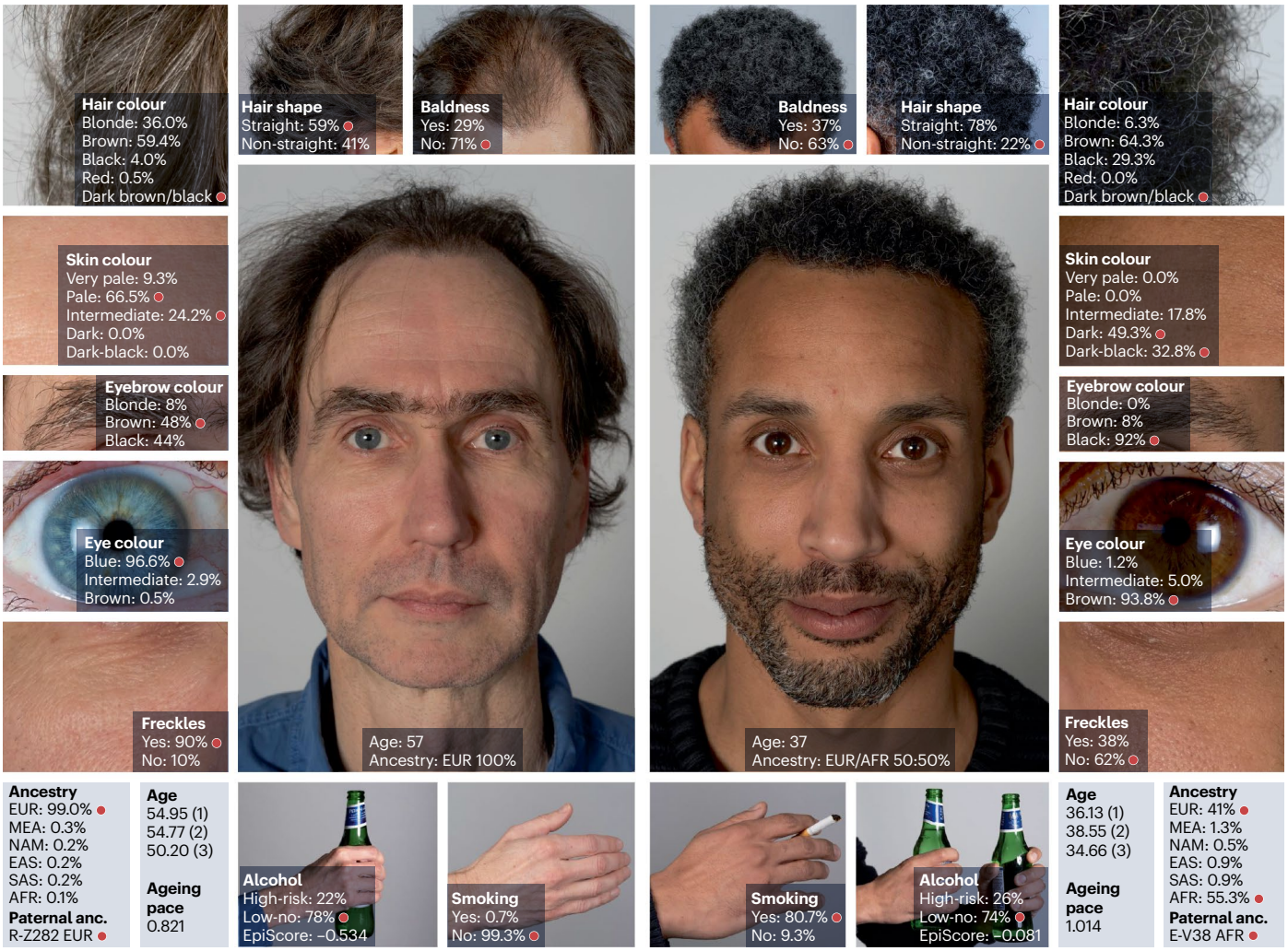

Fig. 3 | Genetic and epigenetic prediction of externally visible characteristics in two example persons. Genetic predictions of seven appearance traits and biogeographic ancestry were achieved with the VISAGE Enhanced Tool for Appearance and Ancestry[113] and the VISAGE software based on previously published prediction models for eye, hair and skin colour[99], eyebrow colour[107], freckles[106], hair shape[109], and hair loss in men[110]. Bi-parental biogeographic ancestry was inferred with STRUCTURE[218] using a reference dataset with population samples drawn from six geographic regions (EUR, Europe; MEA, Middle East; NAM, Native Americans; EAS, East Asia; SAS, South Asia; AFR, Sub-Saharan Africa (n = 2,523))[219]. Paternal ancestry was inferred from the geographic distributions of Y chromosome haplogroups derived from Y chromosomal SNPs with the Yleaf software[126]. Epigenetic predictions of age, tobacco smoking and alcohol drinking were achieved from EPIC2 microarray (Illumina) DNA methylation data using previously published prediction models, that is, for age with three first-generation epigenetic clocks (DNAm Age Horvath[141], Hannum Age[143] and DNAm Age Skin & Blood[142]), for pace of ageing with DunedinPACE[165], for tobacco smoking[167] and for alcohol drinking[175,220]. Prediction outcomes are given as categorical probabilities for appearance traits, smoking and drinking. For drinking, an alcohol consumption EpiScore (values on a relative scale, the lower the value, the lower the consumption) is additionally provided. For bi-parental ancestry, predicted regional ancestry proportions are reported. For paternal ancestry (paternal anc.), the geographic origin of Y haplogroups are provided. Predicted age is given in years and predicted pace of ageing as PACE parameter (<1 decelerated and >1 accelerated pace). Prediction conclusions are marked by red dots. Both persons provided informed consent for their photographs and predicted traits to be shown. Copyright: Erasmus MC.

achieved an AUC of 0.79 (ref. 112). For these additional traits, DNA tools suitable for trace analysis and publicly accessible prediction tools were initially not made available. Currently, the most comprehensive and forensically validated system for appearance (and ancestry) prediction from trace DNA is the MPS-based VISAGE Enhanced Tool for Appearance and Ancestry[113] for genotyping 524 SNPs, and its corresponding prediction tool, the VISAGE software, which together enable the prediction of seven appearance traits based on previously established prediction models for eye colour[99], hair colour[99], skin colour[99], eyebrow colour[107], freckles[106], hair shape[109] and hair loss in men[110], together with ancestry (Fig. 3). Most genetic appearance prediction models are based on logistic regression; machine learning methods have been explored, but they have not improved prediction accuracies thus far[105,114], presumably because of small numbers of SNP predictors.

For many appearance traits, associated SNPs identified in GWAS explain too little of the phenotypic variation, such as for ear

morphology[115] or facial shape morphology[116]. The currently largest GWAS on facial shape identified 253 independently associated SNPs that explained 2.35–7.90% of the phenotypic variation and provides the first preliminary prediction model for nose shape using a polygenic risk score based on 382 face-associated SNPs[117]. Larger GWAS for more appearance traits are needed, followed by improved prediction studies, to further increase the number of DNA-predictable appearance traits in forensic practice. Future advances in the genetic understanding of human appearance may eventually lead to a move from current categorical trait prediction to continuous trait prediction as well as from current genetic prediction with selected SNPs analysed by targeted MPS to genomic prediction with genome-wide SNPs analysed by microarrays or WGS. Genomic prediction has been demonstrated for normal stature based on 20,000–100,000 SNPs from microarray analysis, with a correlation between measured and predicted height with a value above 0.6 (ref. 118). An intermediate approach is to use polygenetic risk scores for trait prediction based on large numbers of significantly associated SNPs, which is widely applied for clinically relevant complex traits[119] and has started to be used for appearance traits such as hair colour[120] and facial shape[117]. However, polygenetic risk score-based trait prediction comes with limitations caused by the relationship between minor allele frequency and effect size of SNPs identified by GWAS[121].

Ancestry prediction

Inferring biogeographic ancestry, that is, the geographic region from which biological (genetic) ancestors originate, is a relevant part of FDP, because biogeographic ancestry has investigative value. It is relevant further because some DNA-predictable appearance traits (for instance, brown eyes with black hair and light skin) are found across large geographic regions passing several continents, which can be subdivided with genetic ancestry testing, thereby specifying and thus improving investigative information.

Traditionally, biogeographic ancestry is genetically inferred separately for the paternal and maternal sides using Y chromosome and mitochondrial DNA SNPs, respectively. During the past decade, targeted MPS assays for simultaneous genotyping of many hundreds of Y chromosomal SNPs[122,123] as well as for the entire mitochondrial genome[124,125] were developed. These genotyping tools and respective software allow high-resolution inference of Y chromosome and mitochondrial DNA haplogroups for paternal and maternal linage identification[126]. Many haplogroups have distinct spatial frequency distributions[127], providing ancestry information on various geographic levels[9].

In contrast to their Y chromosomal and mitochondrial counterparts, autosomal SNPs can differentiate between recent and distant ancestry, and they can quantify ancestry proportions in persons of mixed ancestry[15]. Initially based on small SNP numbers, recent years have seen forensic ancestry tools based on hundreds of autosomal SNPs using targeted MPS – some in combination with Y chromosomal SNPs, appearance-predictive SNPs, microhaplotypes and X-chromosomal SNPs. These tools enable bi-parental ancestry inference on the level of five to seven worldwide geographic regions[92,113,128].

The past decade has also seen progress in methods and software for detecting genetic population substructure[129]. When these tools are used for inferring a person's biogeographic ancestry, population reference data are needed, for example, through web applications such as Snipper[130]. Limitations in sample size and worldwide coverage of population reference datasets available to forensic practitioners mark a limitation of forensic ancestry testing. Another limitation comes from the relatively small number of SNPs in forensic ancestry tools, which is caused by the limited multiplexing capacity of genotyping technologies suitable for forensic trace analysis. Because of the positive relationship between the number of autosomal SNPs and the revealed ancestry information, forensic ancestry tools provide less-detailed ancestry information than is available with SNP microarrays or WGS[23,131].

Moreover, population specificity of human DNA methylation has been reported[132,133]. Although a strong correspondence between population-specific variation in DNA methylation and SNP variation was noted[133], DNA methylation differences were also found in populations with common genetic background living in different habitats[134], indicative of the influence of environmental factors on variation in DNA methylation. Furthermore, several studies indicate the suitability of microbiome analysis for inferring human geolocation, that is, region of living or upbringing[135]. Significant differences in gut microbiome diversity and composition between populations from different continents and countries have been reported[136], and the first studies on forensically relevant hair, skin and oral microbiomes investigated microbiome-based human differentiation between countries[137] and between regions within countries[138]. Viral DNA has started being used to infer geolocation and being combined with genetic biogeographic ancestry testing[139]. The future may see more combined use of genetic ancestry testing and epigenetic and microbiome-based testing for geolocation inference. This will add geographic information regarding the region where a person lived to that regarding the region where their ancestors had lived.

Age prediction

Estimating the age of a person is another relevant part of FDP because age describes a person and thus has investigative value. Furthermore, some appearance traits are strongly influenced by age so that age estimates are required as input for appearance prediction models. Over the past decade, machine learning tools, known as epigenetic clocks, have been developed that can estimate a person's age based on the methylation status at hundreds to thousands of CpGs[140]. First-generation epigenetic clocks[141–143], trained by regressing chronological age on DNA methylation across selected CpG sites, provide more accurate chronological age predictions, which is relevant in the forensic context (Fig. 3), than second-generation and third-generation epigenetic clocks, which are trained with clinical measures of ageing and are more indicative of biological age[144].

However, many crime scene traces do not contain DNA in sufficient quality and quantity for reliable DNA methylation microarray analysis used for epigenetic clocks. Hence, tools for chronological age prediction have been developed based on technologies suitable for trace DNA, including targeted MPS, and based on CpG markers mostly identified previously with DNA methylation microarrays[17,92]. Given their multiplex limitations and as result of reduced DNA complexity after bisulfite conversion, forensic age estimation tools are typically based on small numbers of CpGs (dozens or less)[16,17,145]. They have been developed for various forensically relevant tissues, such as blood, saliva, hair, bone, teeth and semen, and most of them provide age estimates with mean errors of 3–5 years[146]. Cross-tissue tools can be developed but require the use of tissue-specific prediction models[147]. Due to bisulfite conversion and the quantitative nature of DNA methylation variation, forensic age tools based on DNA methylation require more input DNA (tens of nanograms) than it is needed for forensic STR or SNP analysis (picograms). Using input DNA below the method's sensitivity threshold will reduce the accuracy of DNA methylation measurements, with a negative impact on age prediction[148].

Box 2 | Ethical issues of omics applications in forensics

The forensic use of non-targeted omics technologies generates much more data than needed to answer a given forensic question. Moreover, if the (epi)genetic information of a trace donor generated from a trace is restricted by law, applying non-targeted omics to traces means generating more information than is legally allowed. However, there are bioinformatic solutions to ensure that only targeted parts of non-targeted data are being extracted and used, while the rest of the data are left untouched. Targeted bioinformatic analysis of non-targeted omics data is widely established outside the forensic field. Applying targeted bioinformatic solutions on non-targeted omics data obtained from traces to ensure that only those parts of the data are used that are necessary to answer the given forensic question within existing legal frameworks helps mitigate ethical or legal issues arising from the forensic application of non-targeted omics technologies. However, open questions remain, such as what to do with the untargeted raw data after the forensically relevant targeted data are extracted. For example, should they be destroyed or not, and should they be treated with the same care regarding data protection and data security as targeted DNA data are dealt with in forensic genetics?

In principle, externally visible characteristics cannot be considered private. External visibility is mostly the case for physical appearance traits, often for continental biogeographic ancestry and largely for age range. However, persons may artificially alter their natural appearance (for instance by using coloured contact lenses, dying their hair or undergoing cosmetic surgery, which can also alter visible age) to purposefully conceal their natural appearance, in which case such traits could be considered private. However, predicting natural traits would be useless in such cases, as the person cannot be traced using DNA-predicted natural trait outcomes. Furthermore, admixed continental ancestry can be externally invisible depending on the time and complexity of the admixture event(s) in the family, which would then be considered private. For behavioural habits based on environmental interaction, external visibility is not a given. Depending on the habit, societal conventions or legislations, there can be individual reasons to practise such a habit in privacy and keep it hidden from the public, in which case the habit would be considered private. For instance, for most tobacco smokers, their smoking habit is visible as they smoke in public; hence, the epigenetic prediction of tobacco smoking in the context of FDP may not be seen as problematic when it comes to privacy protection. By contrast, consumers of illicit drugs may only consume in privacy; hence, the epigenetic prediction of illicit drug consumption in the context of FDP, once possible based on validated science and technology, may cause an ethical dilemma regarding privacy protection. A broad discussion is needed on societally responsible (epi)genetic prediction of phenotypic traits, ancestry, age and habits based on environmental interactions in particular in forensic casework, as well as on the forensic use of non-targeted omics in general. This discussion will need to consider the societal interest in fighting and preventing crime and the individual's interest in reducing trauma and uncertainty and bringing resolution to survivors, victims, families and communities, versus individual and societal interests in protecting privacy and human rights through governance and systematic sample and data security.

Method-to-method bias provides challenges for developing statistical models for age prediction that are suitable for the data produced with forensic DNA tools. This can be overcome by applying data normalization methods[149] or by generating the model data with the forensic DNA tool[147,150]. Due to limited available resources, prediction models generated with forensic DNA tools typically suffer from small sample size, leading to unknown reliability of the obtained age predictions. Other uncertainties come from external and internal factors that can affect DNA methylation-based age prediction, such as certain diseases[151] and intake of substances such as alcohol[152], among other factors[146]. Age prediction models are often based on regression analysis but also other methods are used[17].

Recent developments have aimed at increasing the number of age-informative CpGs analysable with targeted technologies. Targeted MPS with amplicon sequencing was shown for 161 CpGs covering markers of four epigenetic age models from input DNA as low as 25 ng[149]. Capture-based targeted MPS using single-molecule molecular inversion probes was applied to over 500 age-associated CpGs, but only half of them gave sufficient results down to 100 ng input DNA[153]. Nanopore adaptive sequencing for DNA methylation analysis was applied to hundreds of age-related genes (and others for body fluid identification), but it required large amounts of input DNA[154]. Moreover, enzymatic conversion was tested as an alternative to bisulfite conversion, and it improved multiplexing but did not outperform bisulfite conversion in sensitivity[155].

Other types of nucleic acid markers were also explored for age prediction. A quantitative PCR test based on a specific T cell DNA rearrangement (sjTREC) revealed a high correlation with chronological age with a standard error of 9 years[156]. Transcription-based ageing clocks from non-targeted RNA sequencing (RNA-seq) or expression microarray data provided age prediction with mean errors of 6–8 years[157,158]. Similar age prediction errors were achieved with hundreds of mRNA markers via targeted MPS[159] or with small numbers of microRNA or circular RNA markers[160]. These errors are about twice as high as those of most forensic age prediction systems based on DNA methylation. A comparative study of sjTREC, telomere length, mRNA and DNA methylation revealed that DNA methylation markers were most informative for age[161].

Combining prediction of chronological age with facial age prediction would be desirable in the context of FDP, as some individuals look considerably younger or older in their face than their chronological age. A GWAS on perceived facial age (how old one looks) identified significantly associated SNPs, albeit with small effects (looking 2 years older)[162]. An epigenome-wide Mendelian randomization study on perceived age highlighted over 1,000 potentially causal CpGs[163], providing targets for developing prediction tools for facial age. The second-generation epigenetic clock VisAgeX achieved facial age prediction with an average error of 6.5 years based on microarray data[164]. The third-generation epigenetic clock DunedinPACE quantifies pace of ageing from microarray data[165]; this model predicts if someone ages faster or slower than expected, which may or may not be visible in the face (Fig. 3). Future progress will show whether accurate and reliable (epi)genetic prediction of facial age from trace DNA is achievable.

Environmental habit prediction

Externally visible behavioural traits can potentially be informative in the context of FDP. Various visible habits based on a person's interaction with the environment leave molecular footprints, particularly in the methylome[166]. Epigenetic prediction studies of tobacco smoking in blood, based on small sets of CpGs identified in epigenome-wide association studies, revealed high AUCs of close to and above 0.9 (refs. 167–169) (Fig. 3). Some, but not all, DNA methylation marks of smoking reverse when smoking stops, so that former smoking is predictable, albeit with lower accuracy than the prediction of current smokers and non-smokers[167,169]. Forensic DNA tests for smoking prediction were developed based on forensically suitable technologies, including targeted MPS[170,171]. Smoking also affects the oral microbiota, with several bacteria species showing significant abundance differences in saliva depending on smoking status[172]. A taxonomy-free microbiome approach for smoking prediction yielded an AUC of 0.81 (ref. 173).

For alcohol drinking, extremely high AUC values, 0.90–0.99, were initially reported[174], but they turned out to be overestimates[175]. Models based on 5–144 CpGs achieved AUCs around 0.65 for heavy and at-risk versus light and non-drinkers[175] (Fig. 3). Recently reported models with only three CpGs yielded higher AUCs of 0.77–0.80, but they are based on small sample size[176]. DNA methylation changes resulting from intake of other substances have also been reported[177]. Epigenome-wide association studies on the use of cannabis[178,179] or injectable drugs[180] identified small numbers of significantly associated CpGs explaining small proportions of phenotypic variance, for example, 3.8% for lifetime cannabis use[178]. Considering epigenetic prediction of illicit drugs for FDP applications is therefore currently premature. Other visible habits such as physical exercise and diet also alter the methylome[181,182]. In the future, more visible habits involving environmental interactions will likely become predictable from DNA methylation, which will require discussions on societally responsible forensic applications (Box 2).

Identification of trace characteristics

Having identified the most likely trace donor by forensic STR profiling may not necessarily answer questions of culpability in forensic cases. Additional trace analysis can help the courts decide whether a DNA-identified trace donor is the perpetrator. Determining the tissue type of a trace can reveal the activity that led to the trace deposition. Estimating the trace deposition time may serve as molecular alibi.

Trace tissue type determination

Traditionally, chemical, chemiluminescent, immunological, spectroscopic and microscopic tests are used for the forensic identification of body fluids such as saliva, semen or blood[183]. However, most of these tests are presumptive, that is, they are able to indicate the possible presence of a body fluid, although other substances can also produce a positive test outcome. Therefore, the forensic genetics community developed nucleic acid-based markers and tools for determining the tissue type of traces for blood, menstrual blood, saliva, semen, vaginal secretion and skin, which are routinely applied to forensic casework by specialized laboratories[19] (Fig. 4). For this purpose, several types of nucleic acid markers are used, including mRNA[18,161], microRNA[184,185], DNA methylation[186,187] and DNA copy number variation[188] (Fig. 4); among them, mRNA-based tools are currently the most established[189,190]. A critical issue when using RNA, particularly mRNA, is the impact of trace degradation on RNA integrity, which requires empirical demonstration that the RNA marker considered for any forensic application is not affected by RNA degradation[191,192].

Tissues comprised of overlapping human cell types provide challenges for identification with human biomarkers, but can be differentiated based on microbiome analysis. Microbial 16S ribosomal RNA (rRNA) gene sequencing, increasingly done by targeted MPS, has been used to differentiate and classify epithelial cells from different body locations[193] and blood from different body sources[194] with high prediction accuracies (AUCs over 0.97). Inferring tissue type from marker analysis readouts can be done based on marker presence or absence, abundance quantification with likelihood ratios or with Bayesian networks[19,195,196]. Based on the identified tissue type, activity-level evaluation may identify the most likely activity that placed the trace at the crime scene, by estimating likelihood ratios to differentiate between activities proposed by different court actors, for example, prosecution and defence in an adversarial judicial system[19].

More recent developments aimed at cases with multi-person mixtures trying to link a tissue type with a trace donor (Fig. 4), for which SNPs located in (copy)DNA amplicons of the tissue-indicating mRNA or CpG markers are used. Because these fragments are short (few hundred base pairs), the SNP number is small (for instance, one to four SNPs per mRNA marker for different body fluids[197], or one to two SNPs per CpG marker for semen[198,199]). If several SNPs per amplicon are identified, they are in high linkage disequilibrium, which reduces their strength of evidence for individual identification. Extending the sequence around the tissue-indicating mRNA or CpG marker by targeted long-read sequencing may provide solutions, but it will be conditional on the RNA or DNA quality and quantity of the trace.

Trace deposition time estimation

One approach to estimate the time a trace was placed at a crime scene requires molecular solutions based on the ex vivo stage of the trace. This has been investigated via differential RNA degradation, either by RNA markers that degrade differently over time – for example, rRNA is more stable, whereas certain mRNA markers are less stable[200,201] – or via different amplicons within RNA markers that degrade differently depending on their location within the transcript or their fragment length[20,202]. For instance, based on differently sized mRNA amplicons, three age groups of bloodstains (fresh, up to 6 days old, older than 6 days) could be distinguished[202]; moreover, with differently spaced PCR primers, the accuracy of blood stain timing was reported to be within 2–4 weeks for stains stored for less than 6 months and 4–6 weeks for older samples[203]. Recently, this question was re-investigated with transcriptome sequencing, which identified numerous stable and unstable transcripts[204], and a high AUC of 0.99 for classifying samples into four age groups was reported[205]. As environmental factors other than time, such as temperature and humidity, affect RNA degradation too, applying such tests to traces with unknown or fluctuating environmental conditions, as with crimes occurring outdoors, is challenging. Furthermore, changes in microbial diversity, abundance and succession over sample storage time have been investigated using 16S rRNA-targeted MPS or non-targeted RNA-seq. Many bacterial orders were found to change in abundance or appeared/disappeared with increased sample storage time[206–209]. Genetic models to predict sample age based on a few bacteria achieved average errors of a few days[207,208]. None of these methods have reached forensic casework applications yet.

An alternative way is to estimates the time of day or night when a trace was placed based on the in vivo stage of the body at the time the sample was left behind at the crime scene. This can be investigated

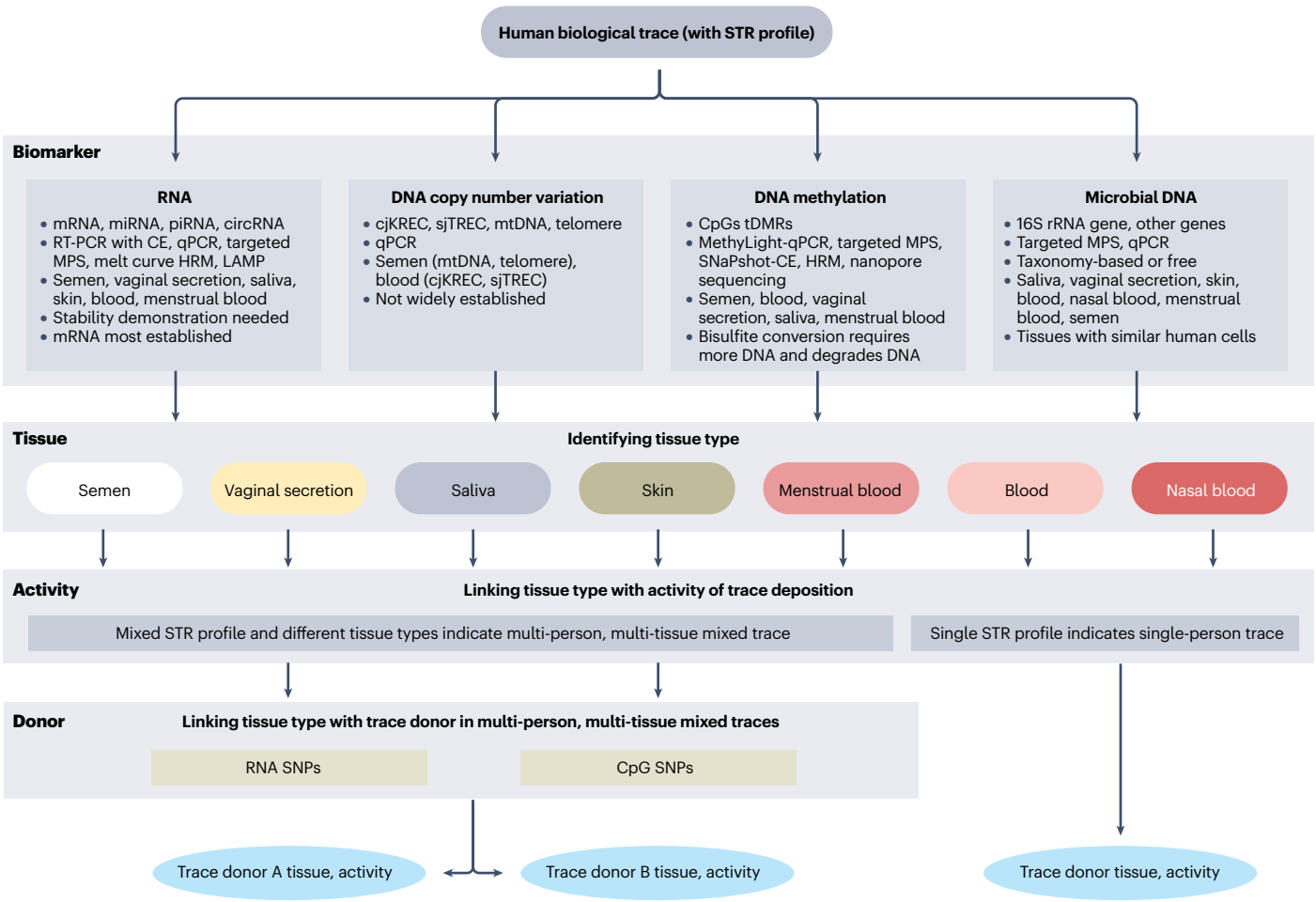

Fig. 4 | Overview of forensic tissue identification with nucleic acid markers. Trace analyses involving different types of nucleic acid markers and analysis techniques for the identification of forensically relevant tissue types, to link identified tissue type(s) with activity of the identified trace donor(s) for depositing the trace at the crime scene, for single-person (single source) traces and multi-person mixed traces. CE, capillary electrophoresis; circRNA, circular RNA; cjKREC, coding joint kappa-deleting recombination excision circle; sjTREC, signal joint T cell receptor excision circle; HRM, high-resolution melt analysis; LAMP, loop-mediated isothermal amplification; miRNA, microRNA; MPS, massively parallel sequencing; mtDNA, mitochondrial DNA; piRNA, piwi-interacting RNA; qPCR, quantitative PCR; rRNA, ribosomal RNA; RT-PCR, PCR with reverse transcription; STR, short tandem repeat; tDMRs, tissue-specific differentially methylated regions.

via biomarkers with biologically controlled concentration changes over the day and night cycle, that is, circadian or at least diurnal biomarkers. Applying circadian biomarkers for forensic trace deposition timing was first demonstrated with two circadian hormones, melatonin and cortisol[210]. mRNA markers from clock genes have also been tested, and a full metabolite survey was done, in blood samples collected across the 24 h cycle[21,211]. Based on the most time-informative biomarkers of all three types (hormones, mRNA, metabolites), prediction models yielded AUCs between 0.8 and 0.96 for three day/night time periods[211]. An mRNA-based model gave an average error of 4 h[212]. Recently, this question was re-investigated with non-targeted and targeted RNA-seq on blood samples collected across the 24 h cycle, which revealed more than 80 mRNA markers with significantly different expression[213]. Targeted MPS of almost 70 mRNA markers in blood samples collected over a day and model-based time estimation achieved average prediction errors of about 4 h or more[214]. DNA methylation markers that oscillate during the day and night may also become useful for this purpose[215]. However, all currently available models are heavily underpowered, given the low number of individuals included and the sample dependency over time points within the same individuals. Moreover, it needs to be demonstrated that the biomarkers used, particularly mRNA, are not affected by trace degradation, as shown for melatonin[210] but not yet for other markers. None of these methods have reached forensic casework applications yet.

Conclusions

Over the past two decades, the field of forensic genetics has considerably advanced in all its subfields, largely due to benefits provided by nucleic acid-based omics approaches and technologies, a trajectory that is expected to continue in the future. Scientific developments are needed to identify more biomarkers for existing and novel forensic

applications to further increase the content, reliability and accuracy of forensic information obtainable from crime scene traces. Technical developments are required to improve targeted and non-targeted omics technologies to cope with both the increasing types and numbers of forensic biomarkers and the limited quality and quantity of biological traces. Future developments of multi-omics tools may allow combining different omics technologies and applying them to limited trace samples, preferentially single cells, to obtain answers to different forensic questions in parallel. As we move forward from forensic genetics to forensic multi-omics, combining nucleic acid-based omics approaches with omics approaches based on other types of molecular biomarkers, such as proteomics (which has started to be used for forensic tissue identification[216]), metabolomics (which has started to be applied for trace deposition time estimation[211]) and glycomics (which has started to be used for age prediction[217]) may lead to further improvements in answering the Who?, What? and When? in crime scene investigations.

Despite the direct and considerable societal benefits of research outcomes in forensic genetics, progress is substantially impeded by serious underfunding. In many countries, there are no or very limited possibilities to apply for forensic genetic research funding, because suitable funding programmes no longer exist or have never existed. Moreover, over the past years, some universities, particularly in Europe, have wound down their forensic genetic research activities, which is

Glossary

Area under the receiver operating characteristic curve
(AUC). A metric that classifies a prediction model's performance, based on graphical representation, the receiver operating characteristic curve, which plots the true positive rate (sensitivity) against the false positive rate (specificity) at various threshold settings; AUC of 0.5 means random prediction and 1.0 completely accurate prediction.

Bayesian networks
Type of probabilistic graphical model to visualize variables and their relationships; in forensics, used to reason about the uncertain conclusions inferred from evidence information in a case.

Biogeographic ancestry
Geographic region(s) from which the biological (genetic) ancestors of a person originate.

Biological trace
Any, including small or residual, amount of (human) biological material of any type left behind at a crime scene and collected for analysis in crime scene investigation.

Bisulfite conversion
Treatment of DNA with bisulfite leading to deamination of unmethylated cytosines to become uracils, detected in subsequent analysis as thymidines. By contrast, methylated cytosines are resistant to deamination and remain cytosines, which allows quantitative analysis of DNA methylation variation at CpGs.

Capture enrichment
Procedure to enrich for pre-selected DNA targets, for instance via probe hybridization, with subsequent sequence analysis via massively parallel sequencing technologies.

DNA fingerprints
Extraordinary variable DNA patterns based on minisatellite repeats at multiple loci that are found with extremely low probability in more than one person (except for monozygotic twins).

DNA match
Complete resemblance at each and every genotype at all DNA markers analysed of two forensic DNA profiles, one obtained from a trace and the other from a known person's reference DNA, thereby allowing to identify the trace donor, or both obtained from different traces, thereby allowing to link traces from the same donor within the same or between different criminal cases.

Epigenetic clocks
Machine learning algorithms trained from DNA methylation data and ageing parameters to estimate a person's biological age, some of which can approximate chronological age.

Forensic DNA phenotyping
(FDP). DNA analysis of traces to predict physical appearance, biogeographic ancestry and age of the trace donor aimed at providing investigative leads to help find unknown perpetrators in cases without known suspects.

Forensic DNA profiling
Generating a DNA pattern, most commonly based on multi-locus short tandem repeats (STR profiling), from a trace and comparing it with that of a person's reference DNA. If enough STRs are analysed, as done in current forensic practice, a match (see below) of the complete STR profile indicates an extremely low probability that it is found in more than one person (except for monozygotic twins).

Genetic population substructure
Systematic differences in allele frequencies within a population between subpopulations such as those defined by geography.

Mendelian randomization
Method to analyse (epi)genetic data for studying causal effects of modifiable exposures on influencing different outcomes.

PCR stutter
Artificial STR alleles, differing in repeat lengths from the true allele, produced during the PCR (polymerase chain reaction) because of DNA strand slippage of the DNA polymerase during the amplification of repetitive sequences (also called slippage artefacts).

Product rule
Multiplying frequencies of each and every genotype of all analysed DNA markers for obtaining an estimate for the DNA profile frequency, also referred to as random match probability (or the chance for a random match), which is only applicable to genetically unlinked autosomal DNA markers with completely independent inheritance.

Single-nucleotide polymorphisms
(SNPs). Variant at a single DNA position, mostly single-base substitutions, but also single-base deletions or insertions, sometimes used for variants with population frequency of over 1%.

Single-nucleotide variants
(SNVs). Variants at a single DNA position, mostly single-base substitutions but also single-base deletions/insertions, often used for variants without a certain population frequency and typically used for rare variants.

Touch DNA
DNA analysis, most commonly via forensic STR profiling, of small amounts of biological material, usually skin cells, transferred by a person to an object via physical contact.

especially critical as governmental forensic institutions, with very few exceptions, cannot perform the type of forensic genetic research that leads to breakthrough innovations. To ensure the continuation of advances in forensic genetics and its societal benefits, governments must provide better funding opportunities for forensic genetic research, and universities must resume their societal responsibility in supporting forensic genetic research capabilities.

Published online: 08 October 2025

References

1. Jeffreys, A. J., Wilson, V. & Thein, S. L. Individual-specific 'fingerprints' of human DNA. *Nature* **316**, 76–79 (1985).
2. Jeffreys, A. J., Brookfield, J. F. & Semeonoff, R. Positive identification of an immigration test-case using human DNA fingerprints. *Nature* **317**, 818–819 (1985).
3. Edwards, A., Civitello, A., Hammond, H. A. & Caskey, C. T. DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. *Am. J. Hum. Genet.* **49**, 746–756 (1991).
4. Budowle, B. & Sajantila, A. Short tandem repeats — how microsatellites became the currency of forensic genetics. *Nat. Rev. Genet.* **25**, 450–451 (2024).
5. Higuchi, R., von Beroldingen, C. H., Sensabaugh, G. F. & Erlich, H. A. DNA typing from single hairs. *Nature* **332**, 543–546 (1988).
6. Stoneking, M., Hedgecock, D., Higuchi, R. G., Vigilant, L. & Erlich, H. A. Population variation of human mtDNA control region sequences detected by enzymatic amplification and sequence-specific oligonucleotide probes. *Am. J. Hum. Genet.* **48**, 370–382 (1991).
7. Cavalcanti, P., Nogueira, T. L. S., Carvalho, E. F. & Silva, D. A. D. Forensic use of human mitochondrial DNA: a review. *An. Acad. Bras. Cienc.* **96**, e20231179 (2024).
8. Roewer, L. & Epplen, J. T. Rapid and sensitive typing of forensic stains by PCR amplification of polymorphic simple repeat sequences in case work. *Forensic Sci. Int.* **53**, 163–171 (1992).
9. Kayser, M. Forensic use of Y-chromosome DNA: a general overview. *Hum. Genet.* **136**, 621–635 (2017).
10. Gill, P. An assessment of the utility of single nucleotide polymorphisms (SNPs) for forensic purposes. *Int. J. Leg. Med.* **114**, 204–210 (2001).
11. Sobrino, B., Brión, M. & Carracedo, A. SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Sci. Int.* **154**, 181–194 (2005).
12. Grimes, E. A., Noake, P. J., Dixon, L. & Urquhart, A. Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. *Forensic Sci. Int.* **122**, 124–129 (2001).
13. Kayser, M. Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. *Forensic Sci. Int. Genet.* **18**, 33–48 (2015).
14. Phillips, C. et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci. Int. Genet.* **1**, 273–280 (2007).
15. Phillips, C. Forensic genetic analysis of bio-geographical ancestry. *Forensic Sci. Int. Genet.* **18**, 49–65 (2015).
16. Zbiec-Piekarska, R. et al. Development of a forensically useful age prediction method based on DNA methylation analysis. *Forensic Sci. Int. Genet.* **17**, 173–179 (2015).
17. Naue, J. Getting the chronological age out of DNA: using insights of age-dependent DNA methylation for forensic DNA applications. *Genes Genom.* **45**, 1239–1261 (2023).
18. Bauer, M. & Patzelt, D. Evaluation of mRNA markers for the identification of menstrual blood. *J. Forensic Sci.* **47**, 1278–1282 (2002).
19. Sijen, T. & Harbison, S. On the identification of body fluids and tissues: a crucial link in the investigation and solution of crime. *Genes* **12**, 1728 (2021).
20. Bauer, M., Polzin, S. & Patzelt, D. Quantification of RNA degradation by semi-quantitative duplex and competitive RT-PCR: a possible indicator of the age of bloodstains? *Forensic Sci. Int.* **138**, 94–103 (2003).
21. Lech, K. et al. Evaluation of mRNA markers for estimating blood deposition time: towards alibi testing from human forensic stains with rhythmic biomarkers. *Forensic Sci. Int. Genet.* **21**, 119–125 (2016).
22. Jobling, M. A. & Gill, P. Encoded evidence: DNA in forensic analysis. *Nat. Rev. Genet.* **5**, 739–751 (2004).
23. Kayser, M. & de Knijff, P. Improving human forensics through advances in genetics, genomics and molecular biology. *Nat. Rev. Genet.* **12**, 179–192 (2011).
24. Butler, J. M. & Willis, S. Interpol review of forensic biology and forensic DNA typing 2016-2019. *Forensic Sci. Int. Synergy* **2**, 352–367 (2020).
25. Sullivan, K. M., Mannucci, A., Kimpton, C. P. & Gill, P. A rapid and quantitative DNA sex test: fluorescence-based PCR analysis of X-Y homologous gene amelogenin. *BioTechniques* **15**, 636–638, 640–641 (1993).
26. Guideline for DNA database management review and recommendation. *European Network of Forensic Science Institutes* https://enfsi.eu/wp-content/uploads/2024/07/DNA-GDL-004-GUIDELINE-FOR-DNA-DATABASE-MANAGEMENT-REVIEW-AND-RECOMMENDATIONS.pdf (2023).
27. Wickenheiser, R. A. Expanding DNA database effectiveness. *Forensic Sci. Int. Synergy* **4**, 100226 (2022).
28. Burrill, J., Daniel, B. & Frascione, N. A review of trace "Touch DNA" deposits: variability factors and an exploration of cellular composition. *Forensic Sci. Int. Genet.* **39**, 8–18 (2019).
29. van Oorschot, R. A. H. & Jones, M. K. DNA fingerprints from fingerprints. *Nature* **387**, 767 (1997).
30. van Oorschot, R. A. H., Szkuta, B., Meakin, G. E., Kokshoorn, B. & Goray, M. DNA transfer in forensic science: a review. *Forensic Sci. Int. Genet.* **38**, 140–166 (2019).
31. van Oorschot, R. A. H., Meakin, G. E., Kokshoorn, B., Goray, M. & Szkuta, B. DNA transfer in forensic science: recent progress towards meeting challenges. *Genes* **12**, 1766 (2021).
32. Taylor, D., Kokshoorn, B. & Biedermann, A. Evaluation of forensic genetics findings given activity level propositions: a review. *Forensic Sci. Int. Genet.* **36**, 34–49 (2018).
33. Gill, P. et al. DNA commission of the international society for forensic genetics: assessing the value of forensic biological evidence - guidelines highlighting the importance of propositions. Part II: evaluation of biological traces considering activity level propositions. *Forensic Sci. Int. Genet.* **44**, 102186 (2020).
34. Gill, P. Application of low copy number DNA profiling. *Croat. Med. J.* **42**, 229–232 (2001).
35. Buckleton, J. Validation issues around DNA typing of low level DNA. *Forensic Sci. Int. Genet.* **3**, 255–260 (2009).
36. Taylor, D., Bright, J. A. & Buckleton, J. The interpretation of single source and mixed DNA profiles. *Forensic Sci. Int. Genet.* **7**, 516–528 (2013).
37. Bille, T. W., Weitz, S. M., Coble, M. D., Buckleton, J. & Bright, J. A. Comparison of the performance of different models for the interpretation of low level mixed DNA profiles. *Electrophoresis* **35**, 3125–3133 (2014).
38. Bleka, O., Storvik, G. & Gill, P. EuroForMix: an open source software based on a continuous model to evaluate STR DNA profiles from a mixture of contributors with artefacts. *Forensic Sci. Int. Genet.* **21**, 35–44 (2016).
39. Coble, M. D. & Bright, J. A. Probabilistic genotyping software: an overview. *Forensic Sci. Int. Genet.* **38**, 219–224 (2019).
40. Gill, P., Benschop, C., Buckleton, J., Bleka, O. & Taylor, D. A review of probabilistic genotyping systems: EuroForMix, DNAStatistX and STRmix. *Genes* **12**, 1559 (2021).
41. Oldoni, F. & Podini, D. Forensic molecular biomarkers for mixture analysis. *Forensic Sci. Int. Genet.* **41**, 107–119 (2019).
42. Bennett, L. et al. Mixture deconvolution by massively parallel sequencing of microhaplotypes. *Int. J. Leg. Med.* **133**, 719–729 (2019).
43. Castella, V., Gervaix, J. & Hall, D. DIP–STR: highly sensitive markers for the analysis of unbalanced genomic mixtures. *Hum. Mutat.* **34**, 644–654 (2013).
44. Elliott, K., Hill, D. S., Lambert, C., Burroughes, T. R. & Gill, P. Use of laser microdissection greatly improves the recovery of DNA from sperm on microscope slides. *Forensic Sci. Int.* **137**, 28–36 (2003).
45. Anslinger, K. & Bayer, B. Whose blood is it? Application of DEParray™ technology for the identification of individual/s who contributed blood to a mixed stain. *Int. J. Leg. Med.* **133**, 419–426 (2019).
46. Huffman, K. & Ballantyne, J. Single cell genomics applications in forensic science: current state and future directions. *iScience* **26**, 107961 (2023).
47. Kulhankova, L. et al. Single-cell transcriptome sequencing allows genetic separation, characterization and identification of individuals in multi-person biological mixtures. *Commun. Biol.* **6**, 201 (2023).
48. Kulhankova, L., Bindels, E., Kayser, M. & Mulugeta, E. Deconvoluting multi-person biological mixtures and accurate characterization and identification of separated contributors using non-targeted single-cell DNA sequencing. *Forensic Sci. Int. Genet.* **71**, 103030 (2024).
49. van der Gaag, K. J. et al. Massively parallel sequencing of short tandem repeats-population data and mixture analysis results for the PowerSeq system. *Forensic Sci. Int. Genet.* **24**, 86–96 (2016).
50. Carratto, T. M. T., Moraes, V. M. S., Recalde, T. S. F., Oliveira, M. L. G. & Teixeira Mendes-Junior, C. Applications of massively parallel sequencing in forensic genetics. *Genet. Mol. Biol.* **45**, e20220077 (2022).
51. Ge, J., King, J., Mandape, S. & Budowle, B. Enhanced mixture interpretation with macrohaplotypes based on long-read DNA sequencing. *Int. J. Leg. Med.* **135**, 2189–2198 (2021).
52. de Bruin, D. et al. Exploring nanopore direct sequencing performance of forensic STRs, SNPs, InDels, and DNA methylation markers in a single assay. *Forensic Sci. Int. Genet.* **74**, 103154 (2025).
53. Sobral, A. F., Dinis-Oliveira, R. J. & Barbosa, D. J. CRISPR-Cas technology in forensic investigations: principles, applications, and ethical considerations. *Forensic Sci. Int. Genet.* **74**, 103163 (2025).
54. Shin, G. et al. CRISPR–Cas9-targeted fragmentation and selective sequencing enable massively parallel microsatellite analysis. *Nat. Commun.* **8**, 14291 (2017).
55. Butler, J. M. Recent advances in forensic biology and forensic DNA typing: INTERPOL review 2019-2022. *Forensic Sci. Int. Synergy* **6**, 100311 (2023).
56. Goedbloed, M. et al. Comprehensive mutation analysis of 17 Y-chromosomal short tandem repeat polymorphisms included in the AmpFlSTR Yfiler PCR amplification kit. *Int. J. Leg. Med.* **123**, 471–482 (2009).
57. Roewer, L. et al. DNA commission of the international society of forensic genetics (ISFG): recommendations on the interpretation of Y-STR results in forensic analysis. *Forensic Sci. Int. Genet.* **48**, 102308 (2020).
58. Willuweit, S. & Roewer, L. The new Y chromosome haplotype reference database. *Forensic Sci. Int. Genet.* **15**, 43–48 (2015).

59. Caliebe, A., Zandstra, D., Ralf, A., Kayser, M. & Krawczak, M. A novel mathematical framework for pedigree-based calculation of Y-STR match probabilities. *Sci. Rep.* **15**, 14651 (2025).
60. Ballantyne, K. N. et al. Mutability of Y-chromosomal microsatellites: rates, characteristics, molecular bases, and forensic implications. *Am. J. Hum. Genet.* **87**, 341–353 (2010).
61. Ralf, A. et al. Identification and characterization of novel rapidly mutating Y-chromosomal short tandem repeat markers. *Hum. Mutat.* **41**, 1680–1696 (2020).
62. Ballantyne, K. N. et al. Toward male individualization with rapidly mutating Y-chromosomal short tandem repeats. *Hum. Mutat.* **35**, 1021–1032 (2014).
63. Neuhuber, F. et al. Improving the differentiation of closely related males by RMplex analysis of 30 Y-STRs with high mutation rates. *Forensic Sci. Int. Genet.* **58**, 102682 (2022).
64. Ralf, A. et al. RMplex: an efficient method for analyzing 30 Y-STRs with high mutation rates. *Forensic Sci. Int. Genet.* **55**, 102595 (2021).
65. Ralf, A. et al. Large-scale pedigree analysis highlights rapidly mutating Y-chromosomal short tandem repeats for differentiating patrilineal relatives and predicting their degrees of consanguinity. *Hum. Genet.* **142**, 145–160 (2023).
66. Kayser, M. et al. A comprehensive survey of human Y-chromosomal microsatellites. *Am. J. Hum. Genet.* **74**, 1183–1197 (2004).
67. Zandstra, D. et al. Unprecedented male relative differentiation with Y-SNVs from whole genome sequencing. *Forensic Sci. Int. Genet.* **78**, 103265 (2025).
68. Gershaw, C. J., Schweighardt, A. J., Rourke, L. C. & Wallace, M. M. Forensic utilization of familial searches in DNA databases. *Forensic Sci. Int. Genet.* **5**, 16–20 (2011).
69. Ralf, A., Zieger, M. & Kayser, M. Considerations on expanding criminal offender DNA databases with Y-STR profiles. *J. Law Biosci.* **11**, lsae017 (2024).
70. Tillmar, A., Sturk-Andreaggi, K., Daniels-Higginbotham, J., Thomas, J. T. & Marshall, C. The FORCE panel: an all-in-one SNP marker set for confirming investigative genetic genealogy leads and for general forensic applications. *Genes* **12**, 1968 (2021).
71. Watson, J. L., Grisedale, K., McNevin, D. & Ward, J. Evaluation of the ForenSeq® Kintelligence Kit and the FORensic Capture Enrichment panel for unidentified and missing persons casework. *Int. J. Leg. Med.* **139**, 2047–2062 (2025).
72. Kling, D. & Tillmar, A. Forensic genealogy-a comparison of methods to infer distant relationships based on dense SNP data. *Forensic Sci. Int. Genet.* **42**, 113–124 (2019).
73. Budowle, B., Ge, J., Baker, L., Mittelman, K. & Mittelman, D. Analytical validation of the IBD segment-based tool KinSNP® for human identification applications. *BioTechniques* **77**, 9–22 (2025).
74. Kling, D., Phillips, C., Kennett, D. & Tillmar, A. Investigative genetic genealogy: current methods, knowledge and practice. *Forensic Sci. Int. Genet.* **52**, 102474 (2021).
75. Kennett, D. Using genetic genealogy databases in missing persons cases and to develop suspect leads in violent crimes. *Forensic Sci. Int.* **301**, 107–117 (2019).
76. Phillips, C. The Golden State Killer investigation and the nascent field of forensic genealogy. *Forensic Sci. Int. Genet.* **36**, 186–188 (2018).
77. Dowdeswell, T. L. Forensic genetic genealogy: a profile of cases solved. *Forensic Sci. Int. Genet.* **58**, 102679 (2022).
78. Tillmar, A., Fagerholm, S. A., Staaf, J., Sjolund, P. & Ansell, R. Getting the conclusive lead with investigative genetic genealogy - a successful case study of a 16 year old double murder in Sweden. *Forensic Sci. Int. Genet.* **53**, 102525 (2021).
79. Aanes, H. et al. Heating up three cold cases in Norway using investigative genetic genealogy. *Forensic Sci. Int. Genet.* **76**, 103217 (2025).
80. de Vries, J. H. et al. Impact of SNP microarray analysis of compromised DNA on kinship classification success in the context of investigative genetic genealogy. *Forensic Sci. Int. Genet.* **56**, 102625 (2022).
81. Liu, M. H. et al. Current progress and future perspectives in personal identification of monozygotic twins in forensic medicine. *Forensic Sci. Int. Genet.* **76**, 103231 (2025).
82. Weber-Lehmann, J. et al. Finding the needle in the haystack: differentiating "identical" twins in paternity testing and forensics by ultra-deep next generation sequencing. *Forensic Sci. Int. Genet.* **9**, 42–46 (2014).
83. Rolf, B. & Krawczak, M. The germlines of male monozygotic (MZ) twins: very similar, but not identical. *Forensic Sci. Int. Genet.* **50**, 102408 (2021).
84. Jonsson, H. et al. Differences between germline genomes of monozygotic twins. *Nat. Genet.* **53**, 27–34 (2021).
85. Krawczak, M., Budowle, B., Weber-Lehmann, J. & Rolf, B. Distinguishing genetically between the germlines of male monozygotic twins. *PLoS Genet.* **14**, e1007756 (2018).
86. van der Gaag, K. J. et al. Identifying a monozygotic twin brother as a donor of DNA in minimal, mixed forensic stains - a case example. *Forensic Sci. Int. Genet.* **78**, 103292 (2025).
87. Kaminsky, Z. A. et al. DNA methylation profiles in monozygotic and dizygotic twins. *Nat. Genet.* **41**, 240–245 (2009).
88. Kim, J. Y. et al. DNA methylome profiling of blood to identify individuals in a pair of monozygotic twins. *Genes Genom.* **45**, 1273–1279 (2023).
89. Vidaki, A. et al. Epigenetic discrimination of identical twins from blood under the forensic scenario. *Forensic Sci. Int. Genet.* **31**, 67–80 (2017).
90. Vidaki, A. et al. Investigating the epigenetic discrimination of identical twins using buccal swabs, saliva, and cigarette butts in the forensic setting. *Genes* **9**, 252 (2018).
91. Planterose Jiménez, B. et al. Equivalent DNA methylation variation between monozygotic co-twins and unrelated individuals reveals universal epigenetic inter-individual dissimilarity. *Genome Biol.* **22**, 18 (2021).
92. Kayser, M., Branicki, W., Parson, W. & Phillips, C. Recent advances in forensic DNA phenotyping of appearance, ancestry and age. *Forensic Sci. Int. Genet.* **65**, 102870 (2023).
93. DNA match in Milica van Doorn cold case. *Netherlands Forensic Institute* https://www.forensicinstitute.nl/news/news/2018/01/29/dna-match-in-milica-van-doorn-cold-case (2017).
94. Liu, F. et al. Eye color and the prediction of complex phenotypes from genotypes. *Curr. Biol.* **19**, R192–R193 (2009).
95. Walsh, S. et al. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. *Forensic Sci. Int. Genet.* **5**, 170–180 (2011).
96. Branicki, W. et al. Model-based prediction of human hair color using DNA variants. *Hum. Genet.* **129**, 443–454 (2011).
97. Walsh, S. et al. The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci. Int. Genet.* **7**, 98–115 (2013).
98. Walsh, S. et al. Global skin colour prediction from DNA. *Hum. Genet.* **136**, 847–863 (2017).
99. Chaitanya, L. et al. The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: introduction and forensic developmental validation. *Forensic Sci. Int. Genet.* **35**, 123–135 (2018).
100. Ruiz, Y. et al. Further development of forensic eye color predictive tests. *Forensic Sci. Int. Genet.* **7**, 28–40 (2013).
101. Pośpiech, E. et al. The common occurrence of epistasis in the determination of human pigmentation and its impact on DNA-based pigmentation phenotype prediction. *Forensic Sci. Int. Genet.* **11**, 64–72 (2014).
102. Sochtig, J. et al. Exploration of SNP variants affecting hair colour prediction in Europeans. *Int. J. Leg. Med.* **129**, 963–975 (2015).
103. Breslin, K. et al. HIrisPlex-S system for eye, hair, and skin color prediction from DNA: massively parallel sequencing solutions for two common forensically used platforms. *Forensic Sci. Int. Genet.* **43**, 102152 (2019).
104. Hysi, P. G. et al. Genome-wide association meta-analysis of individuals of European ancestry identifies new loci explaining a substantial fraction of hair color variation and heritability. *Nat. Genet.* **50**, 652–656 (2018).
105. Kukla-Bartoszek, M. et al. Searching for improvements in predicting human eye colour from DNA. *Int. J. Leg. Med.* **135**, 2175–2187 (2021).
106. Kukla-Bartoszek, M. et al. DNA-based predictive models for the presence of freckles. *Forensic Sci. Int. Genet.* **42**, 252–259 (2019).
107. Peng, F. et al. Genome-wide association studies identify multiple genetic loci influencing eyebrow color variation in Europeans. *J. Investig. Dermatol.* **139**, 1601–1605 (2019).
108. Pośpiech, E. et al. Exploring the possibility of predicting human head hair greying from DNA using whole-exome and targeted NGS data. *BMC Genom.* **21**, 538 (2020).
109. Pośpiech, E. et al. Towards broadening Forensic DNA Phenotyping beyond pigmentation: Improving the prediction of head hair shape from DNA. *Forensic Sci. Int. Genet.* **37**, 241–251 (2018).
110. Chen, Y. et al. Genetic prediction of male pattern baldness based on large independent datasets. *Eur. J. Hum. Genet.* **31**, 321–328 (2023).
111. Pośpiech, E. et al. Overlapping association signals in the genetics of hair-related phenotypes in humans and their relevance to predictive DNA analysis. *Forensic Sci. Int. Genet.* **59**, 102693 (2022).
112. Liu, F. et al. Common DNA variants predict tall stature in Europeans. *Hum. Genet.* **133**, 587–597 (2014).
113. Xavier, C. et al. Development and inter-laboratory evaluation of the VISAGE enhanced tool for appearance and ancestry inference from DNA. *Forensic Sci. Int. Genet.* **61**, 102779 (2022).
114. Katsara, M. A. et al. Evaluation of supervised machine-learning methods for predicting appearance traits from DNA. *Forensic Sci. Int. Genet.* **53**, 102507 (2021).
115. Li, Y. et al. Combined genome-wide association study of 136 quantitative ear morphology traits in multiple populations reveal 8 novel loci. *PLoS Genet.* **19**, e1010786 (2023).
116. White, J. D. et al. Insights into the genetic architecture of the human face. *Nat. Genet.* **53**, 45–53 (2021).
117. Xiong, Z., et al. Combined genome-wide association study of facial traits in Europeans increases explained variance and improves prediction. *Nat. Commun.* **16**, 6562 (2025).
118. Lello, L. et al. Accurate genomic prediction of human height. *Genetics* **210**, 477–497 (2018).
119. Ndong Sima, C. A. A. et al. Methodologies underpinning polygenic risk scores estimation: a comprehensive overview. *Hum. Genet.* **143**, 1265–1280 (2024).
120. Cabrejas-Olalla, A. et al. Genetic predictions of eye and hair colour in the Danish population. *Forensic Sci. Int. Genet.* **78**, 103267 (2025).
121. Watanabe, K. et al. A global overview of pleiotropy and genetic architecture in complex traits. *Nat. Genet.* **51**, 1339–1348 (2019).
122. Ralf, A., et al. Forensic Y-SNP analysis beyond SNaPshot: high-resolution Y-chromosomal haplogrouping from low quality and quantity DNA using Ion AmpliSeq and targeted massively parallel sequencing. *Forensic Sci. Int. Genet.* **41**, 93–106 (2019).
123. Claerhout, S. et al. CSYseq: the first Y-chromosome sequencing tool typing a large number of Y-SNPs and Y-STRs to unravel worldwide human population genetics. *PLoS Genet.* **17**, e1009758 (2021).
124. McElhoe, J. A. et al. Development and assessment of an optimized next-generation DNA sequencing approach for the mtgenome using the Illumina MiSeq. *Forensic Sci. Int. Genet.* **13**, 20–29 (2014).
125. Chaitanya, L. et al. Simultaneous whole mitochondrial genome sequencing with short overlapping amplicons suitable for degraded DNA using the ion torrent personal genome machine. *Hum. Mutat.* **36**, 1236–1247 (2015).

126. Ralf, A., Montiel González, D., Zhong, K. & Kayser, M. Yleaf: software for human Y-chromosomal haplogroup inference from next-generation sequencing data. *Mol. Biol. Evol.* **35**, 1291–1294 (2018).
127. Ralf, A. et al. UYSD: a novel data repository accessible via public website for worldwide population frequencies of Y-SNP haplogroups. *Eur. J. Hum. Genet.* **33**, 904–912 (2025).
128. Phillips, C. et al. MAPlex - a massively parallel sequencing ancestry analysis multiplex for Asia-Pacific populations. *Forensic Sci. Int. Genet.* **42**, 213–226 (2019).
129. Moorjani, P. & Hellenthal, G. Methods for assessing population relationships and history using genomic data. *Annu. Rev. Genom. Hum. Genet.* **24**, 305–332 (2023).
130. Santos, C. et al. Inference of ancestry in forensic analysis II: analysis of genetic data. *Methods Mol. Biol.* **1420**, 255–285 (2016).
131. Mandape, S. N. et al. Dense SNP-based analyses complement forensic anthropology biogeographical ancestry assessments. *Forensic Sci. Int. Genet.* **74**, 103147 (2025).
132. Fraser, H. B., Lam, L. L., Neumann, S. M. & Kobor, M. S. Population-specificity of human DNA methylation. *Genome Biol.* **13**, R8 (2012).
133. Carja, O. et al. Worldwide patterns of human epigenetic variation. *Nat. Ecol. Evol.* **1**, 1577–1583 (2017).
134. Fagny, M. et al. The epigenomic landscape of African rainforest hunter-gatherers and farmers. *Nat. Commun.* **6**, 10047 (2015).
135. Franceschetti, L., Lodetti, G., Blandino, A., Amadasi, A. & Bugelli, V. Exploring the role of the human microbiome in forensic identification: opportunities and challenges. *Int. J. Leg. Med.* **138**, 1891–1905 (2024).
136. Abdill, R. J. et al. Integration of 168,000 samples reveals global patterns of the human gut microbiome. *Cell* **188**, 1100–1118.e17 (2025).
137. Skonieczna, K. et al. Salivary microbiome signatures of Poles and Serbians and its potential for prediction of biogeographic ancestry. *Forensic Sci. Int. Genet.* **74**, 103173 (2025).
138. Lei, Y. et al. Comparative analysis of the human microbiome from four different regions of China and machine learning-based geographical inference. *mSphere* **10**, e00672-24 (2025).
139. Toppinen, M. et al. Bones hold the key to DNA virus history and epidemiology. *Sci. Rep.* **5**, 17226 (2015).
140. Teschendorff, A. E. & Horvath, S. Epigenetic ageing clocks: statistical methods and emerging computational challenges. *Nat. Rev. Genet.* **26**, 350–368 (2025).
141. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, R115 (2013).
142. Horvath, S. et al. Epigenetic clock for skin and blood cells applied to Hutchinson Gilford progeria syndrome and ex vivo studies. *Aging* **10**, 1758–1775 (2018).
143. Hannum, G. et al. Genome-wide methylation profiles reveal quantitative views of human aging rates. *Mol. Cell* **49**, 359–367 (2013).
144. Min, M., Egli, C., Dulai, A. S. & Sivamani, R. K. Critical review of aging clocks and factors that may influence the pace of aging. *Front. Aging* **5**, 1487260 (2024).
145. Lee, H. Y. et al. Epigenetic age signatures in the forensically relevant body fluid of semen: a preliminary study. *Forensic Sci. Int. Genet.* **19**, 28–34 (2015).
146. Castagnola, M. J., Medina-Paz, F. & Zapico, S. C. Uncovering forensic evidence: a path to age estimation through DNA methylation. *Int. J. Mol. Sci.* **25**, 4917 (2024).
147. Wozniak, A. et al. Development of the VISAGE enhanced tool and statistical models for epigenetic age estimation in blood, buccal cells and bones. *Aging* **13**, 6459–6484 (2021).
148. Naue, J., Hoefsloot, H. C. J., Kloosterman, A. D. & Verschure, P. J. Forensic DNA methylation profiling from minimal traces: how low can we go? *Forensic Sci. Int. Genet.* **33**, 17–23 (2018).
149. Pośpiech, E. et al. Introduction of a multiplex amplicon sequencing assay to quantify DNA methylation in target cytosine markers underlying four selected epigenetic clocks. *Clin. Epigenetics* **15**, 128 (2023).
150. Pisarek, A. et al. Epigenetic age prediction in semen – marker selection and model development. *Aging* **13**, 19145–19164 (2021).
151. Spolnicka, M. et al. DNA methylation signature in blood does not predict calendar age in patients with chronic lymphocytic leukemia but may alert to the presence of disease. *Forensic Sci. Int. Genet.* **34**, e15–e17 (2018).
152. Piniewska-Rog, D. et al. Impact of excessive alcohol abuse on age prediction using the VISAGE enhanced tool for epigenetic age estimation in blood. *Int. J. Leg. Med.* **135**, 2209–2219 (2021).
153. Simons, R. B., Adams, H. H. H., Kayser, M. & Vidaki, A. Investigating single-molecule molecular inversion probes for medium-scale targeted DNA methylation analysis. *Epigenomes* **9**, 8 (2025).
154. Yuen, Z. W. et al. Profiling age and body fluid DNA methylation markers using nanopore adaptive sampling. *Forensic Sci. Int. Genet.* **71**, 103048 (2024).
155. Simons, R. B. et al. Comparative performance evaluation of bisulfite- and enzyme-based DNA conversion methods. *Clin. Epigenetics* **17**, 56 (2025).
156. Zubakov, D. et al. Estimating human age from T-cell DNA rearrangements. *Curr. Biol.* **20**, R970–R971 (2010).
157. Fleischer, J. G. et al. Predicting age from the transcriptome of human dermal fibroblasts. *Genome Biol.* **19**, 221 (2018).
158. Peters, M. J. et al. The transcriptional landscape of age in human peripheral blood. *Nat. Commun.* **6**, 8570 (2015).
159. Hanggi, N. V. et al. Assessing transcriptomic signatures of aging: testing an mRNA marker panel for forensic age estimation of blood samples. *Forensic Sci. Int. Genet.* **78**, 103282 (2025).
160. Wang, J. et al. Forensic age estimation from human blood using age-related microRNAs and circular RNAs markers. *Front. Genet.* **13**, 1031806 (2022).
161. Zubakov, D. et al. Human age estimation from blood using mRNA, DNA methylation, DNA rearrangement, and telomere length. *Forensic Sci. Int. Genet.* **24**, 33–43 (2016).
162. Liu, F. et al. The MC1R gene and youthful looks. *Curr. Biol.* **26**, 1213–1220 (2016).
163. Vladimir, K., Perisic, M. M., Storga, M., Mostashari, A. & Khanin, R. Epigenetics insights from perceived facial aging. *Clin. Epigenetics* **15**, 176 (2023).
164. Bienkowska, A. et al. Development of an epigenetic clock to predict visual age progression of human skin. *Front. Aging* **4**, 1258183 (2023).
165. Belsky, D. W. et al. DunedinPACE, a DNA methylation biomarker of the pace of aging. *eLife* **11**, e73420 (2022).
166. Vidaki, A. & Kayser, M. From forensic epigenetics to forensic epigenomics: broadening DNA investigative intelligence. *Genome Biol.* **18**, 238 (2017).
167. Maas, S. C. E. et al. Validated inference of smoking habits from blood with a finite DNA methylation marker set. *Eur. J. Epidemiol.* **34**, 1055–1074 (2019).
168. Philibert, R., Dogan, M., Beach, S. R. H., Mills, J. A. & Long, J. D. AHRR methylation predicts smoking status and smoking intensity in both saliva and blood DNA. *Am. J. Med. Genet. B Neuropsychiatr. Genet.* **183**, 51–60 (2020).
169. Pośpiech, E. et al. DNA methylation at AHRR as a master predictor of smoke exposure and a biomarker for sleep and exercise. *Clin. Epigenetics* **16**, 147 (2024).
170. Alghanim, H., Wu, W. & McCord, B. DNA methylation assay based on pyrosequencing for determination of smoking status. *Electrophoresis* **39**, 2806–2814 (2018).
171. Vidaki, A. et al. Targeted DNA methylation analysis and prediction of smoking habits in blood based on massively parallel sequencing. *Forensic Sci. Int. Genet.* **65**, 102878 (2023).
172. Wu, J. et al. Cigarette smoking and the oral microbiome in a large study of American adults. *ISME J.* **10**, 2435–2446 (2016).
173. Díez López, C., Montiel González, D., Vidaki, A. & Kayser, M. Prediction of smoking habits from class-imbalanced saliva microbiome data using data augmentation and machine learning. *Front. Microbiol.* **13**, 886201 (2022).
174. Liu, C. et al. A DNA methylation biomarker of alcohol consumption. *Mol. Psychiatry* **23**, 422–433 (2018).
175. Maas, S. C. E. et al. Validating biomarkers and models for epigenetic inference of alcohol consumption from blood. *Clin. Epigenetics* **13**, 198 (2021).
176. Ambroa-Conde, A. et al. Inference of tobacco and alcohol consumption habits from DNA methylation analysis of blood. *Forensic Sci. Int. Genet.* **70**, 103022 (2024).
177. Mahna, D., Puri, S. & Sharma, S. DNA methylation signatures: biomarkers of drug and alcohol abuse. *Mutat. Res. Rev. Mutat. Res.* **777**, 19–28 (2018).
178. Fang, F. et al. Trans-ancestry epigenome-wide association meta-analysis of DNA methylation with lifetime cannabis use. *Mol. Psychiatry* **29**, 124–133 (2024).
179. Garrett, M. E. et al. Genome-wide DNA methylation analysis of cannabis use disorder in a veteran cohort enriched for posttraumatic stress disorder. *Psychiatry Res.* **333**, 115757 (2024).
180. Shu, C. et al. Epigenome-wide association analyses of active injection drug use. *Drug Alcohol Depend.* **235**, 109431 (2022).
181. Sharples, A. P. A multi-epigenomic map of endurance exercise training. *Trends Genet.* **40**, 736–738 (2024).
182. Mirisola, M. G. The nutriepigenome. *Genes* **14**, 1997 (2023).
183. Virkler, K. & Lednev, I. K. Analysis of body fluids for forensic purposes: from laboratory testing to non-destructive rapid confirmatory identification at a crime scene. *Forensic Sci. Int.* **188**, 1–17 (2009).
184. Hanson, E. K., Lubenow, H. & Ballantyne, J. Identification of forensically relevant body fluids using a panel of differentially expressed microRNAs. *Anal. Biochem.* **387**, 303–314 (2009).
185. Zubakov, D. et al. MicroRNA markers for forensic body fluid identification obtained from microarray screening and quantitative RT-PCR confirmation. *Int. J. Leg. Med.* **124**, 217–226 (2010).
186. Frumkin, D., Wasserstrom, A., Budowle, B. & Davidson, A. DNA methylation-based forensic tissue identification. *Forensic Sci. Int. Genet.* **5**, 517–524 (2010).
187. An, J. H., Choi, A., Shin, K. J., Yang, W. I. & Lee, H. Y. DNA methylation-specific multiplex assays for body fluid identification. *Int. J. Leg. Med.* **127**, 35–43 (2013).
188. Zubakov, D. et al. Introducing novel type of human DNA markers for forensic tissue identification: DNA copy number variation allows the detection of blood and semen. *Forensic Sci. Int. Genet.* **36**, 112–118 (2018).
189. Lindenbergh, A. et al. A multiplex (m)RNA-profiling system for the forensic identification of body fluids and contact traces. *Forensic Sci. Int. Genet.* **6**, 565–577 (2012).
190. Haas, C., Neubauer, J., Salzmann, A. P., Hanson, E. & Ballantyne, J. Forensic transcriptome analysis using massively parallel sequencing. *Forensic Sci. Int. Genet.* **52**, 102486 (2021).
191. Zubakov, D., Hanekamp, E., Kokshoorn, M., van Ijcken, W. & Kayser, M. Stable RNA markers for identification of blood and saliva stains revealed from whole genome expression analysis of time-wise degraded samples. *Int. J. Leg. Med.* **122**, 135–142 (2008).
192. Zubakov, D., Kokshoorn, M., Kloosterman, A. & Kayser, M. New markers for old stains: stable mRNA markers for blood and saliva identification from up to 16-year-old stains. *Int. J. Leg. Med.* **123**, 71–74 (2009).
193. Díez López, C. et al. Novel taxonomy-independent deep learning microbiome approach allows for accurate classification of different forensically relevant human epithelial materials. *Forensic Sci. Int. Genet.* **41**, 72–82 (2019).
194. Díez López, C., Montiel González, D., Haas, C., Vidaki, A. & Kayser, M. Microbiome-based body site of origin classification of forensically relevant blood traces. *Forensic Sci. Int. Genet.* **47**, 102280 (2020).

195. Taylor, D. Probabilistically determining the cellular source of DNA derived from differential extractions in sexual assault scenarios. *Forensic Sci. Int. Genet.* **24**, 124–135 (2016).
196. Samie, L. et al. Use of Bayesian networks for the investigation of the nature of biological material in casework. *Forensic Sci. Int.* **331**, 111174 (2022).
197. Ingold, S., Dorum, G., Hanson, E., Ballantyne, J. & Haas, C. Assigning forensic body fluids to donors in mixed body fluids by targeted RNA/DNA deep sequencing of coding region SNPs. *Int. J. Leg. Med.* **134**, 473–485 (2020).
198. Watanabe, K., Taniguchi, K. & Akutsu, T. Development of a DNA methylation-based semen-specific SNP typing method: a new approach for genotyping from a mixture of body fluids. *Forensic Sci. Int. Genet.* **37**, 227–234 (2018).
199. Li, Z. et al. Development of a multiplex methylation-sensitive restriction enzyme-based SNP typing system for deconvolution of semen-containing mixtures. *Int. J. Leg. Med.* **135**, 1281–1294 (2021).
200. Anderson, S., Howard, B., Hobbs, G. R. & Bishop, C. P. A method for determining the age of a bloodstain. *Forensic Sci. Int.* **148**, 37–45 (2005).
201. Hampson, C., Louhelainen, J. & McColl, S. An RNA expression method for aging forensic hair samples. *J. Forensic Sci.* **56**, 359–365 (2011).
202. Anderson, S. E., Hobbs, G. R. & Bishop, C. P. Multivariate analysis for estimating the age of a bloodstain. *J. Forensic Sci.* **56**, 186–193 (2011).
203. Fu, J. & Allen, R. W. A method to estimate the age of bloodstains using quantitative PCR. *Forensic Sci. Int. Genet.* **39**, 103–108 (2019).
204. Salzmann, A. P., Russo, G., Kreutzer, S. & Haas, C. Degradation of human mRNA transcripts over time as an indicator of the time since deposition (TsD) in biological crime scene traces. *Forensic Sci. Int. Genet.* **53**, 102524 (2021).
205. Zhang, J. et al. Transcriptomic changes and prediction of time since deposition of blood stains. *Forensic Sci. Int.* **355**, 111930 (2024).
206. Salzmann, A. P. et al. Assessing time dependent changes in microbial composition of biological crime scene traces using microbial RNA markers. *Forensic Sci. Int. Genet.* **53**, 102537 (2021).
207. Díez López, C., Kayser, M. & Vidaki, A. Estimating the time since deposition of saliva stains with a targeted bacterial DNA approach: a proof-of-principle study. *Front. Microbiol.* **12**, 647933 (2021).
208. Wang, J. et al. Estimating the time since deposition (TsD) in saliva stains using temporal changes in microbial markers. *Forensic Sci. Int. Genet.* **60**, 102747 (2022).
209. Gursoy, N., Karadayi, S., Akmayan, I., Karadayi, B. & Ozbek, T. Time-dependent change in the microbiota structure of seminal stains exposed to indoor environmental. *Int. J. Leg. Med.* **138**, 591–602 (2024).
210. Ackermann, K., Ballantyne, K. N. & Kayser, M. Estimating trace deposition time with circadian biomarkers: a prospective and versatile tool for crime scene reconstruction. *Int. J. Leg. Med.* **124**, 387–395 (2010).
211. Lech, K. et al. Investigation of metabolites for estimating blood deposition time. *Int. J. Leg. Med.* **132**, 25–32 (2018).
212. Cheng, F. et al. Estimation of bloodstain deposition time within a 24-h day-night cycle with rhythmic mRNA based on a machine learning algorithm. *Forensic Sci. Int. Genet.* **66**, 102910 (2023).
213. Gosch, A., Bhardwaj, A. & Courts, C. TrACES of time: transcriptomic analyses for the contextualization of evidential stains - identification of RNA markers for estimating time-of-day of bloodstain deposition. *Forensic Sci. Int. Genet.* **67**, 102915 (2023).
214. Gosch, A., Sendel, S., Caliebe, A. & Courts, C. TrACES of time: towards estimating time-of-day of bloodstain deposition by targeted RNA sequencing. *Forensic Sci. Int. Genet.* **78**, 103287 (2025).
215. Koncevičius, K. et al. Epigenetic age oscillates during the day. *Aging Cell* **23**, e14170 (2024).
216. Van Steendam, K., De Ceuleneer, M., Dhaenens, M., Van Hoofstat, D. & Deforce, D. Mass spectrometry-based proteomics as a tool to identify biological matrices in forensic science. *Int. J. Leg. Med.* **127**, 287–298 (2013).
217. Gudelj, I. et al. Estimation of human age using *N*-glycan profiles from bloodstains. *Int. J. Leg. Med.* **129**, 955–961 (2015).
218. Pritchard, J. K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
219. Ruiz-Ramírez, J. et al. Development and evaluations of the ancestry informative markers of the VISAGE enhanced tool for appearance and ancestry. *Forensic Sci. Int. Genet.* **64**, 102853 (2023).
220. Bernabeu, E. et al. Blood-based epigenome-wide association study and prediction of alcohol consumption. *Clin. Epigenetics* **17**, 14 (2025).

Acknowledgements

The author thanks A. Ralf for having volunteered for, and carried out, the genetic analyses and genetic prediction analyses shown in Fig. 3. He thanks E. Pośpiech for having carried out epigenetic prediction analyses shown in Fig. 3 based on data generated by the Genomics Core Facility of Erasmus MC. He thanks R. Koppenol for the photographs shown in Fig. 3 and D. Zandstra for his valuable contributions to the design and preparation of all figures. He acknowledges B. Budowle, W. Parson, E. Pośpiech and A. Sajantila for their comments, which all improved the final manuscript. Finally, he thanks all colleagues who contributed with their work to the advances in forensic genetics summarized in this Review.

Competing interests

The author is inventor on three patents EP2195448 A1 "Method to predict iris colour", US20110263437-A1 "Analysis of Y-STR markers" and EP20158807 "Novel Y-chromosomal short tandem repeat markers for typing male individuals". He declares no other competing interests.

Additional information

Peer review information *Nature Reviews Genetics* thanks Jack Ballantyne, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Related links

HIrisPlex-S: https://hirisplex.erasmusmc.nl/
Snipper: https://mathgene.usc.es/snipper/
VISAGE: https://www.visage-h2020.eu
Y-chromosome STR Haplotype Reference Database: https://yhrd.org/