Genotyping tools for forensic DNA phenotyping: From low- to high-throughput

Bhavik M Mehta

A Thesis submitted for the degree of
Doctor of Philosophy (Applied Science)

FEBRUARY 2019
UNIVERSITY OF CANBERRA

I dedicate this work to my spiritual master, His Divine Holiness Hariprasad Swamiji; my wife, Purvi Patel; and my parents.

Abstract

Forensic analysis of deoxyribonucleic acid (DNA) evidence is a powerful tool for law enforcement that can provide a link between a suspect and a crime or eliminate a suspect from suspicion. Forensic analysis of DNA has two broad purposes: forensic identification and forensic DNA phenotyping (FDP). The purpose of identification is to associate a suspect with a crime based on DNA evidence, a feat that is achieved using genetic markers, such as short tandem repeats and single nucleotide polymorphisms (SNPs). Effective DNA identification relies on comparison between a reference DNA profile and an evidentiary DNA profile, which allows for the inclusion or exclusion of a suspect. If no reference profiles are available or there is a large pool of suspects, DNA profiling has limited capacity to resolve a crime. In these cases, police investigators often depend on eyewitness statements, which are notoriously unreliable. By contrast, FDP is the process of inferring phenotypic traits from DNA and can be used as a biological witness. FDP utilises SNPs along with other markers such as insertions-deletions and microhaplotypes. SNPs are the most common markers for FDP because of their low mutation rates and account for more than 85 percent of variance in the human genome.

Genome-wide association studies have identified a variety of SNPs associated with phenotypic traits, such as eye colour, hair colour, skin colour, baldness and freckles. These SNPs must be typed appropriately to generate FDP profiles. A range of SNP genotyping technologies exist, including: real-time polymerase chain reaction (PCR) based assays with probe hybridisation such as TaqMan (Thermo Fisher Scientific—TFS); microfluidic technology such as Fluidigm Biomark or Open array (TFS); single base primer extension assays such as SNaPshot™ (TFS); and post-PCR assays such as high resolution melt (HRM) analysis. Most recently, massively parallel sequencing (MPS) assays incorporating sequencing by synthesis are represented in the forensic field by Ion Torrent (TFS) and Illumina technologies. These SNP-typing technologies differ in cost, throughput, detection methods and run times, which can make it difficult to choose between them for FDP purposes. Some forensically important criteria include simplicity of operation, reliability, reproducibility, flexibility and modularity. An ideal method should be cost effective, able to process degraded samples and have the ability to sequence a large battery of FDP SNPs. This thesis compares forensic SNP genotyping techniques for three categories of throughput: low, medium and high.

HRM analysis is a low-throughput genotyping method and was applied to the IrisPlex eye colour FDP panel of six SNPs. It is a simple and fast post-PCR real-time method. HRM produced reproducible profiles at 0.5 ng DNA input amounts. Its cost-effectiveness can be further increased by using half-volume reactions. IrisPlex includes a symmetrical SNP (rs16891982) and a SNP with high guanine-cytosine content regions (rs1800407) critical to eye colour inference. HRM underperformed in genotyping these SNPs, which might present a challenge in terms of their application for these types of panels. HRM also possesses limited multiplexing capability.

SNaPshot™ (TFS) is the most common forensic SNP-typing tool and was assessed as a medium-throughput genotyping method. This evaluation was also performed using the IrisPlex eye colour panel. The workflow involved a PCR step (amplification of templates) and minisequencing step (single base extension) that introduced a contamination risk due to multiple tube-to-tube transfers. SNaPshot generated reproducible profiles at 0.1 ng DNA and other studies confirmed their reproducibility at 0.062 ng. The assay is able to multiplex up to 40 SNPs and can be applied to both forensic identification—using identity informative SNPs—and FDP. This thesis includes a published review of SNaPshot forensic SNP genotyping assays.

The Illumina MiSeq MPS platform was evaluated as a high-throughput tool. It was used to simultaneously genotype 136 SNPs from five SNaPshot assays: the SNP*for*ID 52-plex, SNP*for*ID 34-plex, Eurasiaplex, Pacifiplex and IrisPlex. MPS libraries were generated from 0.05 ng input amounts for each multiplex. A total of 24 samples were pooled in a single run using unique oligonucleotide barcodes as sample identifiers. MPS was demonstrated to be applicable to degraded samples, UV-exposed samples and humic acid inhibited samples. Sequencing on the MiSeq produced genotypes that were 98 percent concordant with genotypes derived from SNaPshot and Ion Torrent sequencing. It generated 100 percent reproducible profiles. This unique approach demonstrated the capacity to multiplex SNP panels from existing SNaPshot assays (identity and phenotyping) and apply them to multiple samples with no requirement for investing in new panel designs. Further, this thesis describes an automated workflow in a forensic laboratory for routine application of MPS. Two major library normalisation procedures—magnetic bead-based and real-time PCR-based—were compared with real-time PCR to demonstrate the best performance.

In summary, this thesis compares and contrasts three FDP SNP genotyping methods available for forensic applications with different throughput requirements. It is anticipated that the

findings may serve as a starting point and guide for forensic laboratories in implementing FDP SNP-typing for routine cases.

Acknowledgements

I would like to acknowledge my supervisors, Ass. Prof. Tamsin Kelly, Prof. James Robertson, Prof. Dennis McNevin and Dr Runa Daniel, for all their efforts and support. They mentored and nourished me through every step of this journey. Without their moral, mental and intellectual support, I would not have completed this thesis.

The work included in this thesis was carried out at the following laboratories: National Centre for Forensic Studies, Faculty of Science, Technology and Mathematics, University of Canberra; the Office of the Chief Forensic Scientist (OCFS), Victoria Police Services Department, Melbourne; and Specialist Operations, Australian Federal Police, Canberra. I would like to acknowledge the support, especially, in terms of wet lab training provided by staff at all three laboratories.

I am thankful to the late Dr Bryan Found, former Chief Forensic Scientist at the OCFS, for including me in the OCFS family for two years, which paved the way for the work on MPS included in this thesis. I would like to especially thank Dr Paul Roffey, Dr Eric Wenger, Clifton Frost and Slazana Ristevska for accommodating me at the Australian Federal Police and offering support in carrying out MPS work. I would like to acknowledge Dr Stephen Doyle for allowing me to perform some work on his MiSeq sequencer at La Trobe University and for his technical advice. I also acknowledge Gareth Elvidge from Illumina for all his technical knowledge and training support for the MiSeq work included in this research.

I offer my thanks to students from the National Centre for Forensic Studies for their support. I appreciate the assistance and contribution from Dr Samantha Venables, especially regarding HRM and MPS work. I also thank Dr Chris Philips and his team at the Forensic Genetics Unit, University of Santiago de Compostela, for their collaboration and support. I acknowledge the co-authors of publications included in this thesis. I am grateful to the volunteers who provided their DNA samples for studies presented in this thesis.

I am thankful to my dear wife, Purvi Patel, for her immense love and encouragement, without which I would not have completed this journey. I am obliged to my parents for all their love, sacrifices and hard work that enabled me to get where I am today. They instilled in me the passion and ambition to strive with hard work. I am also thankful to my sister, brother-in-law and parents-in-law for their prayers, moral and emotional support.

Last but not least, I am grateful to my spiritual master, His Divine Holiness Hariprasad Swamiji, for his blessings, vision, persistence, belief in me and guidance, which helped me to remain focused, motivated, positive and optimistic throughout this journey.

Table of contents

List of publications

Mehta B, Daniel R, Phillips C, McNevin D (2017) Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. International Journal of Legal Medicine 131(1): 21-37.

Venables SJ, **Mehta B**, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35 (21-22):3036-3043.

Mehta B, Daniel R, McNevin D (2013) High resolution melting (HRM) of forensically informative SNPs. Forensic Science International: Genetics Supplement Series 4 (1):e376-e377.

Mehta B, Daniel R, McNevin D (2017) HRM and SNaPshot as alternative forensic SNP genotyping methods. Forensic Science, Medicine, and Pathology 13 (3):293-301.

Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37 (21):2832-2840

Mehta B, Venables S, Roffey P (2018) Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. International Journal of Legal Medicine 132 (1):125-132.

List of abbreviations

AIM	Ancestry informative marker
AISNP	Ancestry informative single nucleotide polymorphism
APEX	Arrayed primer extension
ASO	Allele specific oligonucleotide
ATP	Adenosine triphosphate
BAM	Binary alignment
BGA	Biogeographical ancestry
CCD	Charge coupled device
CE	Capillary electrophoresis
CMOS	Complementary metal oxide semiconductor
ddNTP	dideoxynucleotidetriphosphate
dNTP	deoxynucleotidetriphospate
DVI	Disaster victim identification
EVC	Externally visible characteristics
FDP	Forensic DNA phenotyping
FP	Fluorescence polarisation
FRET	Fluorescence resonance energy transfer
GA	Genome analyser
GC	Guanine-cytosine
GWAS	Genome-wide association studies
HID	Human identification
HRM	High resolution melt
IISNP	Individual identification single nucleotide polymorphism
INDEL	Insertion/deletion
ISFET	Ion-sensitive field effect transistor
ISP	Ion sphere particle
LC	Light Cycler
LIM	Lineage information marker
LISNP	Lineage informative single nucleotide polymorphism
MALDI-TOF	Matrix assisted laser desorption/ionisation time of flight
MCS	MiSeq control software
MGB	Minor groove binder
MPS	Massively parallel sequencing
MS	Mass spectrometry

mtDNA	Mitochondrial DNA
NGS	Next generation sequencing
PCR	Polymerase chain reaction
PGM	Personal Genome Machine
PIM	Phenotypic informative markers
PISNP	Phenotype informative single nucleotide polymorphism
SBE	Single base extension
SBL	Sequencing by ligation
SBS	Sequencing by synthesis
SFF	Standard flowgram format
SMRT	Single molecule real-time
SMS	Single molecule sequencing
SNP	Single nucleotide polymorphisms
SOLiD	Sequencing by oligonucleotide ligation and detection
STR	Short tandem repeat
TFS	Thermo Fisher Scientific
TS	Torrent suite
UAS	Universal analysis software
VNTR	Variable number tandem repeat
ZMW	Zero mode waveguide

List of figures

List of tables

Chapter 1: Introduction

1.1 Background

This project was a part of an Australian Research Council Linkage grant, *From Genotype to Phenotype: Molecular Photofitting for Criminal Investigations* (LP110100121). Many criminal investigations have neither knowledge of suspects who committed crimes nor how they may appear. In these cases, deoxyribonucleic acid (DNA) profiles obtained using forensic analysis of short tandem repeats (STRs) are uninformative unless a reference profile becomes available. Often, investigators look for eyewitness testimonies to find suspects to obtain reference profiles, although these eyewitness statements are notoriously inaccurate and unreliable [1]. DNA evidence may be considered to act as a 'silent witness' to a crime and has the potential to provide molecular photofits, independent of human eyewitnesses. The goal of molecular photofitting is to infer phenotypes from DNA genotypes. This process is known as forensic DNA phenotyping (FDP). Unlike forensic identification, which commonly uses STRs, FDP utilises other markers, including single nucleotide polymorphisms (SNPs) to make phenotypic inferences. Genome-wide association studies (GWAS) have discovered many FDP SNPs associated with phenotypic traits, such as biogeographical ancestry (BGA) [2], eye, hair and skin colour [3], fingerprint patterns [4] and facial composites [5]. These suggest hundreds of FDP SNPs would be required to construct a complete 'molecular photofit'. STRs are most commonly typed using the capillary electrophoresis (CE)-based fragment analysis method in forensics [6]. SNPs can be genotyped using a variety of technologies and chemistries [7]. Different SNP-typing methods could be applied depending on the FDP application throughput requirements. For example, a laboratory needing to run a small FDP assay may only require a low-throughput SNP-typing tool, while another laboratory typing hundreds of FDP SNPs would need a high-throughput tool. This thesis examines the applicability of three SNP-typing technologies for FDP: high resolution melt (HRM) analysis (low-throughput method); single base extension-based SNaPshot™ (Thermo Fisher Scientific: TFS) assays (medium-throughput technique); and sequencing by synthesis (SBS)-based Illumina massively parallel sequencing (MPS) (high-throughput tool). Chapters 2, 3 and 4 are peer-reviewed publications.

1.2 Forensic analysis of biological evidence

The precipin test was developed in 1901, which identified whether a blood stain is from human or animal origin. This test was used in a 1901 German child murder investigation and formed part of the evidence cascade that led to the execution of Ludwig Tessnow in 1904 [8]. ABO blood grouping test [9] was applied in the most famous Robert Payne's case, in which the

suspect's blood group matched with the non-victim sample found at the crime scene. In light of above evidence, along with other evidence including fibre analysis, the suspect, Robert Payne, was convicted [10]. Another historical case was in 1983 at Sheffield, UK in which the suspect Andrew Hutchinson was convicted of a triple murder based on blood group test evidence that identified his rare blood group—found in only 1 in 50,000 people—along with palm print evidence [11]. In 1953, Watson and Crick determined the structure of DNA [12]. DNA evidence is a powerful forensic investigative tool as their analysis generates valuable genetic code information that can be linked to a suspect or eliminate a suspect from suspicion of a crime [13].

1.2.1 DNA fingerprinting

In the 1980s, Sir Alec Jeffreys and co-workers brought a revolution in forensics with their DNA fingerprinting technique [14]. In a 1986 Leicester double-murder case, a 17-year-old kitchen potter confessed to two murders and confirmed circumstantial evidence against him. However, the DNA fingerprinting technique proved that he could not have been the murderer. Using this DNA technology as a means of screening all men in the area surrounding the crime scene led to Collin Pitchforkthe being identified as the real perpetrator [15]. This landmark case proved both innocence and guilt based on DNA technology.

The DNA fingerprinting technique used variable number tandem repeats (VNTRs)—genetic markers, also called minisatellites—which have nucleotide unit repeats ranging from eight to several thousands and vary from person to person [16, 17]. The method employed restriction endonucleases that cleaved the DNA sequence at specific sites (e.g., Hae III cuts the sequence repeat GGCC [18]), followed by gel electrophoresis to detect the different-sized fragments. Individual DNA fragments were extracted and processed using southern hybridisation [19]. Targeted single locus probes were utilised to obtain the VNTR locus banding pattern and these band patterns were called 'DNA profiles' [14, 17]. The profiles with VNTR fragments occupying corresponding positions were recognised as a match and others as a mismatch. VNTR alleles differing by one or two repeats were indistinguishable and hence the bands of similar size were grouped into bins. This reduced the total number of alleles from hundreds to twenty or thirty. The frequencies of allele falling into various bins were used to calculate match probability [13].

VNTRs provided better discriminatory power in comparison to restrictions fragment lengths polymorphism (RFLPs) markers. However, the technique was not very sensitive as it required

a large amount of input DNA (in the range of 50–100ng) and could not be used in polymerase chain reactions (PCR) due to the large size of VNTR markers [13]. In addition, the technique was time consuming and lacked applicability to samples with small amounts of DNA [13].

1.2.2 STR profiling

In the early 1990s, STRs—also referred to as microsatellites—were applied to forensic DNA evidence analysis. STRs consist of an array of tandem repeats, with each repeat generally 2–7 base pairs long [20, 21]. STRs were much shorter than VNTRs and could be analysed using PCR, which was invented in 1986 [22, 23]. Application of PCR to DNA analysis enabled the generation of profiles with low input amounts of DNA and made DNA profiling possible for a wide variety of biological material recovered from crime scenes. STRs were the first set of markers to be amplified using PCR technology in forensics [24].

In 1991, Alec Jeffery applied STR analysis on challenging casework samples to establish the identity of a murder victim in the United Kingdom [25]. In 1992, the identity of skeletal remains exhumed in Brazil was confirmed of Josef Mengele using STRs [26]. The early STR profiling methods contained STRs with dinucleotide repeats, which suffered from high levels of stutters and led to their replacement by tetra-nucleotide repeats, which decreased stutter levels and increased diversity [27]. PCR multiplexing of STR markers became easier with the development of fluorescent labels and one of first STR multiplexes (a quadraplex) was developed in 1994 by Forensic Science Services (FSS) [27]. This was followed by a six STR locus multiplex assay with the amelogenin gender identification marker [28]. The first commercial STR multiplex kit for silver stain analysis was released by Promega in 1994—commonly referred as ‘CTT’ triplex [29]. To date, a variety of commercial STRs multiplex kits exist, such as GlobalFiler™ (TFS) [30], Identifiler™ (TFS) [31] and Powerplex™ 21 (Promega) [32].

1.3 Forensically relevant DNA markers

STRs are the most commonly used marker for forensic identification. These identity markers have characteristics of being highly polymorphic, codominant, high heterozygosity, low FST—a measure of genetic distance—between global populations, low mutation rates and high discrimination power [33]. STRs have been firmly established in forensics since the early 1990s. More than 10^5 STRs are found in the human genome and forensics uses STRs from non-coding regions of human DNA for forensic identity profiling [33]. STRs, due totheir long repeat

regions, sometimes cannot be applied to degraded samples. Mitochondrial DNA (mtDNA) sequencing is an alternative method utilised when nucleus DNA is degraded or not available. The hypervariable regions of human mtDNA are sequenced and compared with reference for identification [34, 35]. The mtDNA sequencing process is laborious, time consuming, expensive and non-discriminative compared to STR analysis [35]. In addition, STRs—due to their mutation rates up to 0.05 nucleotide per generation—have limitations in applying to paternity cases in which there is no availability of the mother's DNA [36]. There are other forensically relevant marker sets such as SNPs, insertions-deletion (INDELs) and microhaplotypes (MHs), which can be considered in the above situations.

1.3.1 Single nucleotide polymorphisms

SNPs are a single base change in a DNA sequence, in which the least frequent allele must have an abundance of one percent or higher in a population, by convention [37]. SNP variation is observed when a single nucleotide (A, T, G or C) differs between members of a species or chromosome (paired) in the genome of an individual. The vast majority of SNPs in the genome are bi-allelic; however, tri- and tetra- allelic SNPs also exist [37, 38]. Of the approximately 14–15 million SNPs recorded in the human database, 94,000 are tri-allelic [39].

SNPs have a low mutation rate, approximately 2.5×10^{-8} per nucleotide site per generation [40]. Other studies estimate SNP mutation rates between 1×10^{-9} and 5×10^{-9} per nucleotide per year at neutral positions in mammals [41]. The average mutation rate is $\sim 1.1 \times 10^{-8}$ per nucleotide site per generation [42]. Therefore, the probability of two independent base changes occurring at a single position is very low.

More than 85 percent of human variance is derived from SNPs [43, 44], which are mainly bi-allelic. SNPs are robust markers in data interpretation, laboratory handling, inheritance stability and population genetic analysis [45]. SNPs are also applied to various forensic applications, including paternity testing and kinship testing. A battery of identification SNPs (usually 40–50) are used to get the discrimination power equivalent to 13 STRs [46]. Due to their smaller size, SNPs offer advantages for degraded samples over STRs and having low mutation rates makes them markers of choice for lineage and ancestry prediction applications. Due to occurrence in both coding and non-coding regions, SNPs are used to make the phenotypic association of human traits. While STRs are still the most accepted markers for identification, SNPs offer application versatility in both forensic identification and FDP [1].

1.3.1.1 SNP categories used in forensics

SNPs for forensic applications can be broadly classified into four categories [47, 48]:

- Individual identification SNPs (IISNPs): These SNPs are used to complement STRs in forensic identification. IISNPs have high heterozygosity and low FST between global populations.
- Lineage informative SNPs (LISNPs): These uniparental inherited markers are used to make lineage inferences of the sample. LISNPs are used in the identification of missing persons, paternity and maternity testing, sexual assault cases and kinship analyses. For example, mtDNA SNPs and Y SNPs are used to solve kinship cases.
- Ancestry informative SNPs (AISNPs): AISNPs can be used for inferring biogeographical ancestry—this term is generally reserved for autosomal ancestry informative markers. Contrary to IISNPs, AISNPs have low heterozygosity and high FST between global populations.
- Phenotype informative SNPs (PISNPs): PISNPs helps in inferring individual having particular externally visible characteristics (EVCs), such as eye, hair and skin colour.

1.3.1.2 SNPs for forensic identification

IISNPs are utilised in forensics for identification in the same way as identity STRs. These SNPs collectively provide extremely low probabilities of two individuals having the same multisite genotype as STR loci. It delivers the genetic information to distinguish two people and excludes suspects that cannot be a source of an evidentiary sample [47]. IISNPs to be used in forensic applications should have high heterozygosity and low FST between global subpopulations. Kidd (2011) indicates that ideal IISNPS for forensic use should satisfy the criteria of being easily typed unique locus, highly informative for the required purpose and possess well-documented relevant characteristics [48].

SNP identification panels such as 21 SNP panel [49] and a SNPforID 52-plex SNP panel [50] were developed earlier. Later, Kidd et al (2006) identified a set of 19 SNPs by interrogating a public database of 90,000 potential SNPs from 40 population groups [51]. Pakstis et al. (2007), developed a 40 SNP identification panel with FST<0.06 and heterozygosity >0.4. More recently, a 45 SNP universal identification panel was published [52, 53]. The discrimination power of bi-allelic SNPs is not as high as STR loci. Some studies shows 10–15 tri-allelic SNPs

have the same discrimination power as 40–50 bi-allelic SNPs for forensic identification [54, 55].

IISNPs variation does not have repetitive sequences like STRs, as IISNP profiling avoids stutter artefacts and can be used in cases with trace amounts of DNA [46]. Short amplicon lengths are generated when analysing IISNPs, which are useful in the analysis of degraded samples and in disaster victim identification (DVI) cases [1]. Tri-allelic SNPs are a useful option in reducing the large number of bi-allelic IISNPs required to achieve high discrimination power and in the resolution of mixed-source samples.

STR criminal databases are extensive and have been generated over decades. This is not the case with IISNP databases therefore, IISNPs are not likely to replace STRs for routine identity testing, although they are appropriate for use in small, closed populations, as encountered in DVI. Bi-allelic IISNPs have limited application to mixed samples or mixture analysis in comparison with STRs, which have higher heterozygosity and multiple alleles. STR markers remains the 'gold standard' for forensic identification purposes [1]. IISNPs may not replace STRs for identification, but they can complement STR profiling in the case of degraded samples.

1.3.2 Insertion/Deletions

Insertion/deletions (INDELs) are length polymorphisms created by insertions or deletions of one or more nucleotides in the genome. Di-allelic INDELs were known in 2002, which demonstrated the potential of these markers in genetic studies [56]. Mills et al. (2006) identified more than 400,000 unique INDEL polymorphisms and further estimated that the human genome would harbor more than 1.5 million INDELS [57]. A class of INDELs with allele length variations between 2bp and 10kb spanned ~41 percent of total INDELs across the genome, with nearly all of them under 100bp [57]. These small INDELs can be applied to analyses using PCR and CE. INDELs are the second most common class of mutation in the human genome [58]. These makers have low mutation rates of approximately ~2 x 10^{-8} [40]. INDELs finds their applicability in forensics due to the following characteristics: i) wide distribution across the genome; ii) low mutation rates; iii) allelic frequency differences among different populations groups is significant, which could be potential ancestry informative markers; iv) large scale multiplexing capability by analysing small INDELs in short amplicons and can be applied to degraded samples; v) genotyping easily existing PCR and CE technologies in forensics; and vi) suitability of small INDELs for high-throughput technologies [59]

38-plex human identification (HID) assay is an example of INDEL identification assay that can obtain profiles from 0.3 ng to 5 ng input DNA amount, with discrimination power greater than 99.999 percent [59]. A commercial DIPplex (Qiagen), 30-plex INDEL (with an amelogenin marker) assay also exists, which has shown to obtain profiles from 62 pg of input DNA amount [60]. INDELs are also known to be applicable in ancestry inferences. 46 AIM-INDELs assay can distinguish African, European, East Asian and North American populations [61].

1.3.3 Microhaplotypes

Haplotypes are sets of DNA variations or polymorphisms that are co-inherited. Haplotypes are a combination of alleles at multiple location on a single chromosome [62, 63]. SNPs in close proximity (<10kb) tend to have low recombination rates, less than 10^{-4}. Due to the history of origin of the variants at different sites, rare recombination events, and the impulses of random genetic drift and selection, suggests the existence of multiple haplotypes in the genome [64]. A microhaplotype (MH) marker is an SNP-based multiallelic locus. MH loci have two or more SNPs in close proximity, usually within less than 200 nucleotides and with three or more allelic combinations. These MHs have characteristics including: i) they can be covered in short amplicons and are applicable to high-throughput technologies; ii) no stutters as for STRs; iii) low mutation rates; and iv) all alleles at a locus are the same size [64, 65].

MHs can be forensically informative in mixture detection, deconvolution and identification of close biological relationships [66, 67]. Recently developed, 74-plex MHs assay differentiated African-American, European-American and South-West Hispanic populations [68]. One hundred and thirty MHs were recently published with their estimated allelic frequencies in 83 different populations. Many of those loci were shown to be highly informative for identification and mixture detection and deconvolution [66].

1.4 Need for alternative genotyping technologies for forensic DNA analysis

Electrophoresis was developed in 1955 [69] and successfully utilised in the development of the Sanger DNA sequencing method as the detection system [70, 71]. The development of CE in the 1990s assisted in the automation of the DNA sequencing technique [72]. CE-based fragment analysis method for STR profiling is now routinely used for forensic identification.

PCR and CE-based fragment analysis methods can be also utilised to genotype SNPs, INDELs and MHs. One limitation of CE is in analysing multiple samples together. CE-based technology is widely used in forensic identification that requires only a few STRs (15–23) to genotype [29,

33]. In FDP, the DNA association to phenotypic traits—eye, skin and hair colour—is depicted predominantly using genetic marker SNPs. INDELs and MHs also have the potential to be used as FDP markers. The expansion of the current repertoire of forensic DNA analysis tools is required to genotype forensically relevant markers, including SNPs, INDELs and MHs, along with STRs together for FDP analysis.

1.5 Forensic DNA phenotyping

In many forensic cases, there are no suspects or a large pool of suspects exists, and investigators are often left depending on eyewitnesses statements, which are notoriously unreliable. In such instances, the inference of EVCs and biogeographical ancestry (BGA) from DNA—which is called FDP—can provide intelligence or leads for the investigators in the form of 'biological eyewitness.'

Forensic DNA identification involves the analysis of markers in the non-coding region of the genome, whereas FDP analyses utilise markers in both the coding and non-coding regions of the genome [73]. The conventional DNA profiling aims to exclude a particular suspect from some population, whereas FDP can help determine the population in which a suspect can be included [74].

The FDP have raised ethical issues and ethicists emphasise the limits of regulation, education, active monitoring and appropriate guidelines for interpretation [73]. In Netherlands, FDP is explicitly regulated by legislation, with the permission of determining the race, gender and EVCs from birth, which solely contributes to criminal investigation. Germany—except the state of Bavaria [75]—and some states of the United States—such as Indiana, Rhode Island and Wyoming—disallow FDP by legislation [76]. In a majority of states in the United States, it is not written in legislation. In UK, FDP is implicitly implemented in legislation [76]. With the advancement of technology and forensic/scientific quests, specific legislation may arise relating to FDP, considering societal and parliamentary debates on their acceptability. Regulatory issues, such as privacy and data protection, stigmatisation and non-discrimination, the right not-to-know and avoiding slippery slopes must be considered but their certain relevance to FDP should not be overestimated [76]. FDP can be used in crime investigations, for inculpation or exculpation of suspects or groups for further investigation, suggesting criminal legislation for intelligence-led policing [73, 77].

1.6 FDP markers

The process of FDP is dependent on genetic markers used for their purpose. Currently, the information produced by the FDP approach can be broadly classified as inferring lineage and biogeographical ancestry and EVCs directly from DNA samples. For the above purposes, specific genetic markers are used. Genetic markers employed for inferring lineage ancestry are referred to as lineage informative markers (LIMs), for BGA are ancestry informative Markers (AIMs), and those inferring EVCs are known as phenotypic informative markers (PIMs)

1.6.1 Lineage informative markers

Lineage-based analyses utilise mitochondrial DNA (mtDNA) and the non-recombining portion of the Y-chromosome (NRY). mtDNA (inherited maternally) and NRY (inherited paternally) have been useful in human evolution and genealogical studies [47]. LIMs are haploid, uniparental, not subject to recombination and can be used to construct maternal and paternal lineages [78]. LIMs include SNPs, INDELs, MHs, and STRs.

1.6.2 Ancestry informative markers

Lineage ancestry markers analyses focus only on maternal or paternal lineages. Autosomal genetic ancestry markers can differentiate among biogeographical ancestral groups and are more commonly referred to as AIMs—previously referred to as population-specific alleles. AIMs demonstrate substantial differences in allele frequency across population groups [78].

A forensic scientist may need to perform AIMs analysis to infer ancestry, which aids in active investigations. BGA is used to express the heritable component of the ancestral group (population) and their inference from DNA using AIMs [79]. Low mutation rate AIMs are considered for BGA inference to make it applicable to a wide number of populations with accuracy. AIMs with high FST between global populations are preferred as these provide more genetic differentiation between subpopulations [79, 80]. Along with STRs [81], AIMs also includes SNPs, INDELs and MHs.

1.6.3 Phenotypic Informative markers

DNA markers with the capacity to describe phenotype traits can enable genetic prediction of appearance and can help investigators to identify offenders of a crime. These genetic markers are referred as PIMs, which can also be applied by forensic anthropologists for facial reconstruction of unknown human remains [47].

Some of the common phenotypic traits of individual appearances are eye colour, hair colour, skin pigmentation and androgenic alopecia or male pattern baldness. The single term used is EVCs. EVCs prediction using PIMs is rapidly growing in the forensics community [82]. The immense amount of genetic information on EVCs and human phenotypes is available due to GWAS [83]. SNPs, INDELs and copy number variants are the source of phenotypic variation [83, 84]. SNPs are the most common markers used for FDP.

1.7 SNP genotyping technologies

The discovery of FDP SNPs is an ongoing process for inferring phenotypic traits. However, an appropriate platform and chemistry for reliably genotyping these FDP SNPs to use in forensic laboratories for intelligence purposes is one of the most important forensic needs. Recently, various SNPs typing technologies have come into existence, based on different allelic discrimination and detection platforms. The products resulting from allelic discrimination reactions can be detected with many methods and the same detection method can be used to analyse products of various assay formats. Most of the existing SNP genotyping technologies can be divided into six types, based on molecular mechanisms: allele specific hybridisation, primer extension, allele specific oligonucleotide ligation, invasive cleavage, MPS and HRM (see Table 1.1). The products from these molecular mechanisms can be analysed using various detection methods such as fluorescence and mass measurements. The assay formats for the reactions can be categorised as solution-based—one occurring in solutions, also known as homogenous reactions—and solid support based, such as glass slide and chips—array hybridisation [7].

Basis for technique	**Representative assay**	**Detection method**	**Assay formats**	**Reference**
Allele specific hybridisation	Reverse Blot	Colorimetry	Membrane-based	[85]
	Light Cycler (Roche)	Fluorescence	Solution-based	[86]
	TaqMan (TFS)	Fluorescence	Solution-based	[87, 88]
	Molecular Beacons	Fluorescence	Solution-based	[89]
	Amplifluor assay	Fluorescence	Solution-based	[90]
	Gene Chip (Affymetrix)	Fluorescence	Solution-based	[91]
Primer Extension	SNaPshot (TFS)	Fluorescence	Solution-based	[50]
	PinPoint (Applied Biosystems)	Mass spectrometry	Solution-based	[92]

	Array primer extension (APEX)	Fluorescence	Chip-based	[93]
Allele specific oligonucleotide ligation	Infinium bead chip array (Illumina)	Fluorescence	Chip-based	[94]
	SNPlex (TFS)	Fluorescence	Chip-based	[95]
Invasive cleavage	Invader assay (Third Wave Technology)	Fluorescence	Solution-based	[96]
Next Generation Sequencing	Ion Torrent	pH based	Chip-based	[97]
	Illumina SBS	Fluorescence	Solid surface - based	[98]
Melting Temperature	High Resolution Meting Temperature (HRM) analysis	Fluorescence	Solution-based	[99]

Table 1.1: Classification of SNP genotyping methods on broad molecular mechanisms

1.7.1 Allele specific hybridisation

The selection between two DNA targets varying at a single nucleotide position using hybridisation is known as allele specific hybridisation or allele specific oligonucleotide (ASO) hybridisation (see Figure 1.1) [100]. In ASO-based reactions, two probes with a polymorphic base in a central position of the probe sequence are designed. These probes are allele specific and bind to the target DNA only if stable at optimised assay conditions. The mismatch probe-target hybrids are unstable. The reverse blot format ASO probes were used first to detect PCR analysed polymorphism in forensics [7, 85]. Most of the ASO hybridisation SNP-typing methods use fluorescence detection methods and can be divided into two groups, based on assay formats, as solution-based or chip-based.

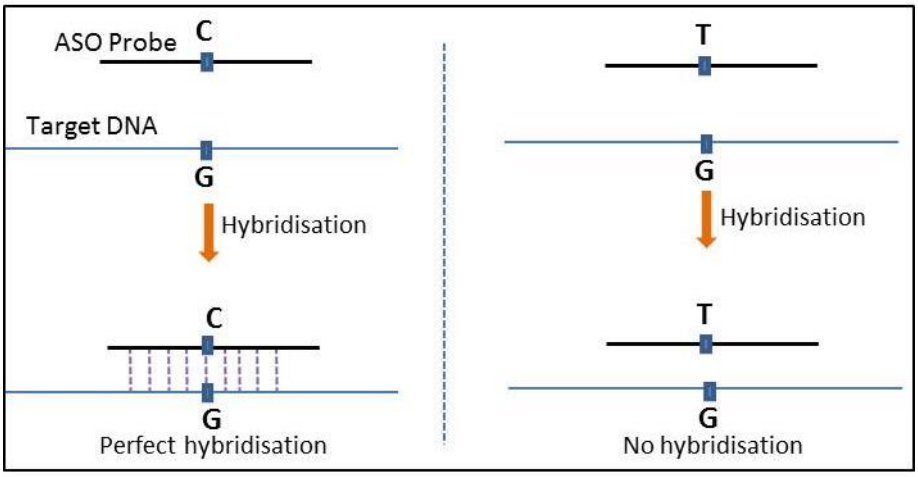


Figure 1.1: Allele specific hybridisation mechanism of SNP genotyping. The left hand side represents the presence of allele specific oligonucleotide (ASO) probes which facilitate perfect hybridisation. The right hand side represents the absence of ASO probes leading to no hybridisation.

1.7.1.1 Solution-based ASO hybridisation using fluorescence detection

The fluorescence resonance energy transfer (FRET) is the main principle in these assays. FRET occurs when two fluorescent dyes are close to each other and the emission spectrum of one fluorophore overlaps the excitation spectrum of the other [101]. These solution-based genotyping techniques combine ASO hybridisation allelic discrimination with real-time PCR reactions. Hence, two PCR primers are required in addition to the ASO probes. The intensity of fluorescence is measured in real-time PCR either during PCR or on completion of PCR. Many typing methods have been developed based on the above principle with some modifications. The significant advantage of solution-based methods is that no post-PCR step is required, enabling PCR and detection being performed in the same reaction [7].

1.7.1.1.1 TaqMan assay

TaqMan assay is also called 5’ nuclease allelic discrimination assay. It requires forward and reverse PCR primers and two differently labelled TaqMan minor groove binder (MGB) probes [88]. The assay is based on the 5’ nuclease activity of the Taq polymerase that displaces and cleaves the oligonucleotide MGB probes hybridised to the target DNA, which generates a fluorescent signal detected by real-time PCR [87]. Two MGB probes differ at the polymorphic site—one probe is complimentary to the variant allele and the other to the wild type allele. These probes are labelled with different fluorescent dyes—for example, FAM and VIC dyes if ABI Prism 7900HT detection system is used [88]—at 5’ end and a quencher at 3’end [87]. The quencher interacts with fluorophore by FRET only if the probes are intact, quenching their

fluorescence. In the PCR annealing step, MGB probes hybridise with target DNA. In the following extension step, Taq polymerase cleaves fluorescent dye, resulting in an increase of the reporter dye fluorescence. The wrongly hybridised probes or mismatch probes are dislodged without fragmentation. The intensity measurements of two different dyes determine the genotype of the unknown sample [88].

The other detection method for the 5' nuclease assays is fluorescence polarisation (FP) [7]. In this method, the emissions are obtained from the excited fluorophore by plane—polarised light remains polarised. In any allele discrimination reaction based genotyping method, FP can be used if the product of the reaction is larger or smaller than the starting fluorescent molecule [7]. Fluidigm Biomark real-time PCR systems is an alternative for 5' nuclease assays SNP-typing technique to achieve high sensitivity and high-throughput [102].

1.7.1.1.2 Molecular beacons

Molecular beacons are oligonucleotides probes with two complementary sequences flanking the complementary sequence to the target DNA. Molecular beacons have fluorophore in the 5' end and a quencher at the 3'end. There is no fluorescence observed when the fluorophore is quenched by the quencher if the probe and the target DNA are not hybridised.. In this non-hybridised condition the probe is designed to adopt a hairpin loop confirmation. The fluorescence appears when the molecular beacon is hybridised perfectly to the complementary target DNA [89]. Two molecular beacons are employed in SNP-typing, one for wild type allele and another for variant allele. Each is labelled with different fluorescent dyes enabling allele discrimination in a single PCR reaction [89, 103]. Various targets can be detected in a single reaction using different molecular beacons with different colour fluorescent dyes. The number depends on the capability of the detection platform available [7, 89].

1.7.1.1.3 Ampliflour assays

Ampliflour assays use the Ampliflour universal primer system [90]. These assays are also based on FRET from an excited fluorophore to a sophisticated acceptor moiety resulting in quenching (see Figure 1.2). The fluorophore and the acceptor 4-(dimethylamino) azo benzene sulfonic acid are linked to an oligonucleotide primer, which accomplishes quenching. Each of these uniprimers consists of a different 3' primer sequence and 5' hairpin region that is labelled with a unique and specific energy transfer dye. The primer sequences helps to bind unlabelled target-specific primers to the target DNA. These target-specific primers are designed in such a way

that their 5’ tail sequence is identical to the 3’ region of uniprimers, which allows them to hybridise to the PCR products. In the PCR reaction, the incorporated uniprimer acts as a template for DNA polymerisation, which results in replication and displacement of the hairpin sequence catalysed by DNA polymerase. The efficient fluorescence quenching is obtained in the hairpin confirmation when fluorophore is at 5’ base of the primer and quencher is linked to complimentary nucleotide of the 5’base. The fluorescent signal intensity signifies the amount of amplified DNA [104].

Ampliflour SNP genotyping assays utilises two Ampliflour SNPs primers (uniprimers) and three unlabelled primers—two allele specific primers and one common primer [105]. It is a single tube system. The pair of Ampliflour primers is attached with fluorescent dyes (FAM, SR or JOE). Two SNP-specific primers and a common reverse primer [106, 107] are designed to amplify over the SNP, each with 5’ tail corresponding to one of the Ampliflour SNPs primers. As the SNP-specific products are generated, the primer sequence of Ampliflour SNPs primers binds to the complementary tail sequence of the newly generated PCR product and amplifies further with the help of common reverse primer. Depending on the base present, fluorescent signal is generated [104].

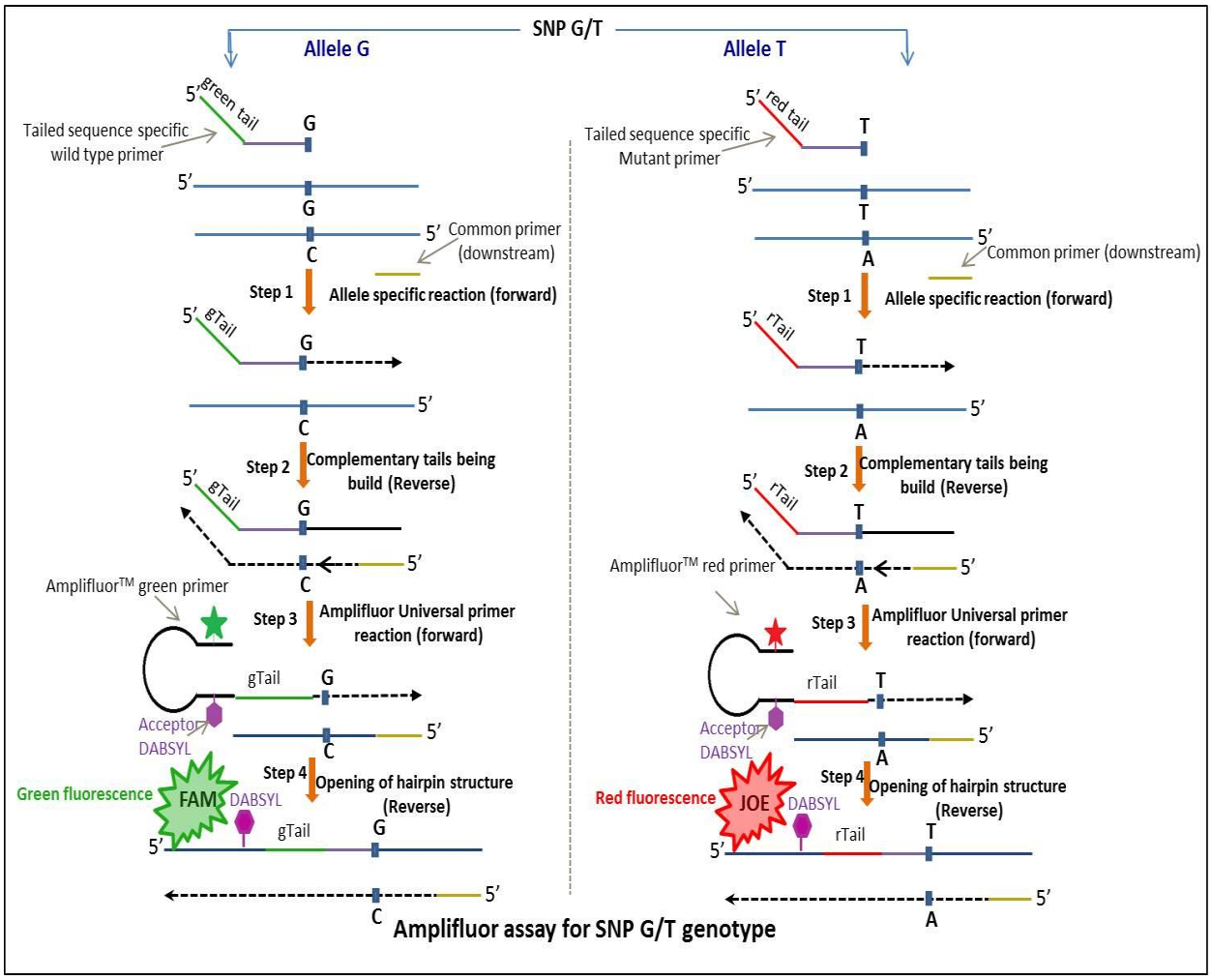


Figure 1.2: Steps of Amplifluor assay for SNP genotyping. The example illustrates the typing of SNP G/T with left hand side of the figure showing the steps for allele call G and right hand side for allele T.

1.7.1.1.4 Roche Light Cycler assay

The Roche Light Cycler (LC) assays employs two fluorescent labelled specific oligonucleotide probes. Probe 1 is attached with fluorescein label at 3' end and probe 2 carries another label (for example, LC red) at 5'end. These probes are designed to attach adjacent to each other on the target DNA sequence. As a result of probe hybridisation, the two fluorescent dyes come in close proximity to each other. The fluorescein dye's green emission excites the LC red dye to emit because of their close proximity positions. This FRET is dependent on the distance between the two dye molecules and high efficiency of FRET is only seen if the spacing between two fluorescent dyes is 1–5 nucleotides. The fluorescence is measured after the annealing step as LC red emits light after the hybridisation of both the oligonucleotide probe. LC red light intensity followed by filtration is measured by an LC instrument. The signal intensity is proportional to the amount of target DNA amplification product [86].

The target polymorphic SNP base is located in central position in one of the probes and the other must be adjacent to allow for FRET. A mismatch can significantly reduce the melting temperature of the oligonucleotide probe and this temperature reduction is primarily dependent on the length of the oligonucleotide and the position of mismatch. This phenomenon of reduced temperature can be measured using melting curve analysis. The combination of using different fluorescent labels with probes at different melting temperature can enable the genotyping of more than one SNP at a time [7].

1.7.2 Primer extension

Primer extension approach is based on the DNA polymerase activity and is dependent on the ability of DNA polymerase to incorporate specific deoxyribonucleotides complementary to the template DNA sequence. Further, many modifications of these reactions are known, but broadly classified into two single nucleotides—primer extension and allele specific primer extension. In single nucleotide primer extension—also known as minisequencing methods—the addition of dideoxynucleotidetriphosphate (ddNTP) complementary to the base cross-examined by DNA polymerase determines the polymorphic site (base). In allele specific extension, the amplification by DNA polymerase is done only if the primer-template hybrid is a perfect match [7].

1.7.2.1 Single nucleotide extension methods

In single nucleotide extension reactions, the primer anneals to the target DNA at the immediate adjacent to SNP position, which is extended by DNA polymerase with a single nucleotide complementary to the polymorphic site [108, 109].

The primer extension reaction products can be analysed by different methods depending on detection technique chosen such as using labelled or unlabelled nucleotide, only ddNTP or a combination of ddNTP and deoxynucleotidetriphospate (dNTP). The most common available technologies using single nucleotide extension approaches are electrophoresis with fluorescence detection (SNaPshot) and matrix assisted laser desorption/ionisation time of flight (MALDI-TOF) mass spectrometry (MS) and microarrays with fluorescence detection [7].

1.7.2.1.1 SNaPshot assay (TFS)

SNaPshot is the technology available in commercial kits supplied by Applied Biosystems and works on the principle of single nucleotide extension reactions followed by electrophoresis and fluorescence detection methods (see Figure 1.3). The method uses fluorescent ddNTPs. An

unlabelled primer is positioned in such a way that their 3'end is at the immediate base upstream to the SNP base and is extended by DNA polymerase with a single fluorescent labelled ddNTP. Each ddNTP is labelled with a different fluorescent dye. The reactions are multiplexed by spatial separation of the single nucleotide extension reaction products using tails at 5' end of the SNaPshot primers with varying lengths of non-human sequence. Further, the products are separated in an automated CE DNA sequencer [50, 110].

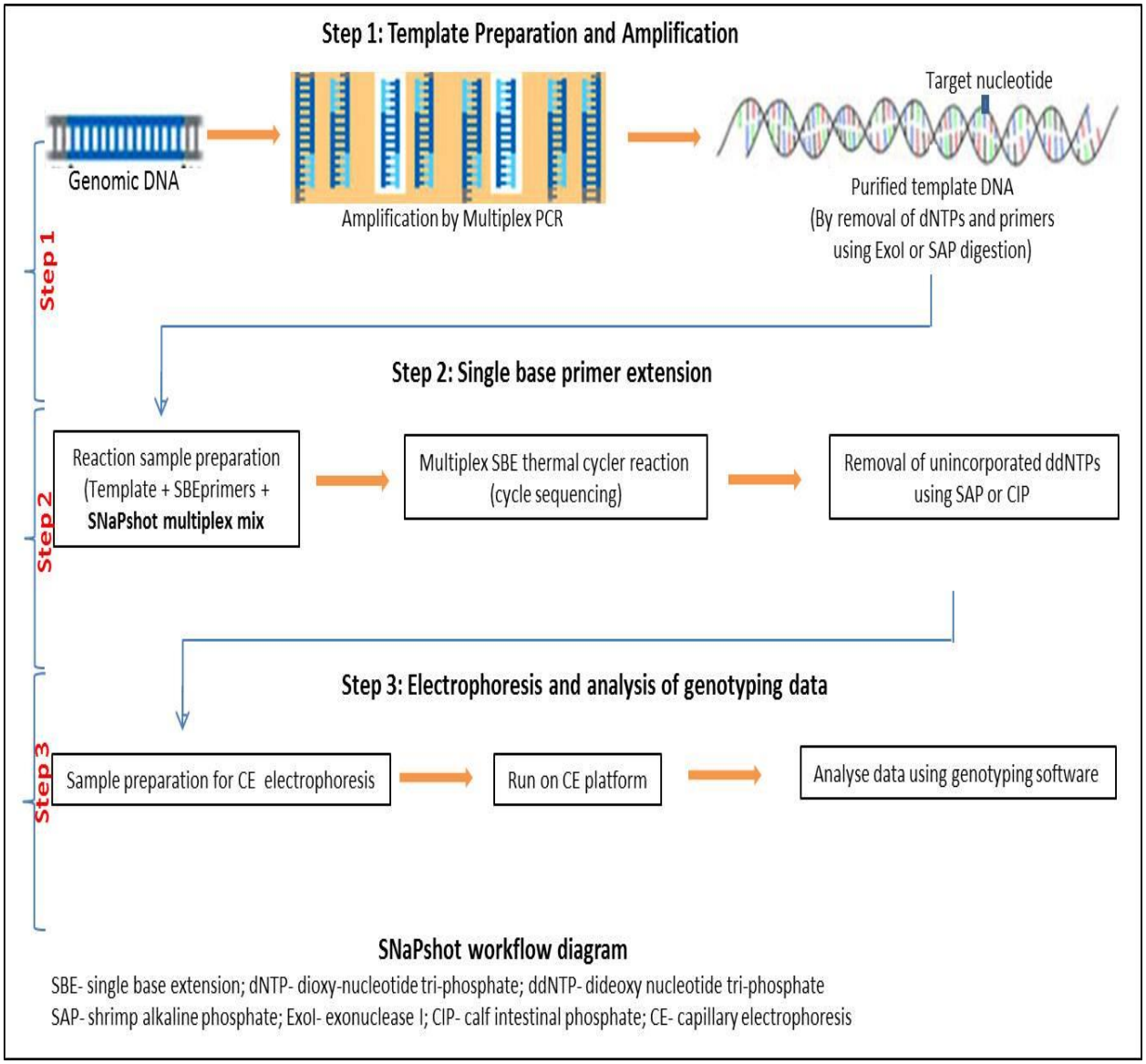


Figure 1.3: The figure illustrates the three main steps of the SNaPshot SNP typing: template preparation and amplification, single base primer extension and electrophoresis and analysis of data.

1.7.2.1.2 MALDI-TOF-MS

This method measures the molecular weight of the single nucleotide primer extension products. It is the direct method of detection in comparison to assay methods identifying products using

fluorescent emitted signals. The base mass added to the primer extended is known by the increasing mass of ddNTPs nucleotides added [92]. MALDI-TOF-MS resolution is high enough to detect the smallest mass different of nine Daltons between ddA and ddT and the specific ddNTP incorporated in primer extension can be detected. The primer extension products are run on the matrix placed on the chip or plate or slide. A desorption process is undertaken, in which both the DNA product and matrix are emitted by laser beam pulse. A flight tube collects the expelled DNA product, resulting from the vapours due to the laser beam energy transfer to the matrix. Time of flight is measured, which is the time between the application of laser beam pulse and collision of DNA product to the detector. This time of flight is correlated into mass measurements as lighter products travel faster than heavier products. The software linked to the instrument reveals the mass measurements [7].

The SNPs genotyping methods based on MALDI-TOF-MS are: PROBE assay (MassEXTEND, Sequenom), which combines the use of ddNTP and dNTPs in primer extension reactions to increase the mass differences between SNPs alleles; [111] and the PinPoint assay (Applied Biosystems), which uses only ddNTPs [92].

1.7.3 Arrayed primer extension (APEX)

APEX reactions are microarray-based with a single nucleotide primer extension performed on chip or solution [93]. In one method, the minisequencing reaction primers are attached to chips and extended by DNA polymerase using labelled ddNTPs and the microarray is checked for fluorescence. In the second method, single base extension is carried out with specific 5' sequence tag primers, distinct for each SNP [112]. The multiplex primer extension products done in solution are hybridised to the reverse complementary tag sequences arrays onto the chip [112].

1.7.3.1 Allele specific extension

This method is dependent on DNA polymerase extension efficiency between matched and mismatched 3' ends of primers. It means DNA polymerase extends the primer with a perfect hybridisation of their 3'end to the complementary DNA target. Two primers are used, one specific to each allele of a SNP and detecting which primer formed the product-determined SNP genotype. The product can be detected using fluorescent labelled nucleotides on a microarray [93]. There is another variation to this approach, known as allele specific PCR. This uses a common reverse primer in addition to allele specific primers and the matching primer

permits the amplification of a specific allele in the target DNA. The detection of PCR product reveals the genotype of the target DNA [106]. Germer et al. (1999) studied the use of tag primers based on melting curve analysis for identification of allele specific PCR products. FRET detection method can also be used [113].

1.7.4 Allele specific oligonucleotide ligation

ASO is based on the DNA ligase activity. The oligonucleotide ligation assay was developed on the ability of ligase to covalently join two oligonucleotides when they hybridise next to each other on a DNA template (see Figure 1.4) [114]. This method uses three probes, one common and two allele specific. The allele specific probes bind to each allele and the common probes anneals immediately downstream to the SNP target. The enzyme DNA ligase will only bind the perfectly matched allelic probe with the common probe. Ligase chain reaction, generating exponential ligation products, requires both strands of gDNA to consist of targets. The products from the first ligation become targets and the chain reaction continues. Many assay formats were developed for detection of this method, including use of biotinylated common probes with reporter group on allele specific probes and the use of fluorescent labelled dyes [7].

SNPlex™ technology utilised ASO ligation principles, in which the fragmented gDNA was analysed using three unlabelled ligation probes per SNP target in a multiplex assay. Following ligation, PCR amplification was performed using two universal primers with one carrying a biotin molecule. These biotinylated products were made bound to streptavidin-coated plates, in which fluorescent probes bind to the PCR products and detected using CE genetic analysers [7]. The array based detection technology like Bead Array™ (Illumina Inc) utilised by GoldenGate™ (Illumina Inc) assays uses a combination of ligation and allele specific extension principles [7].

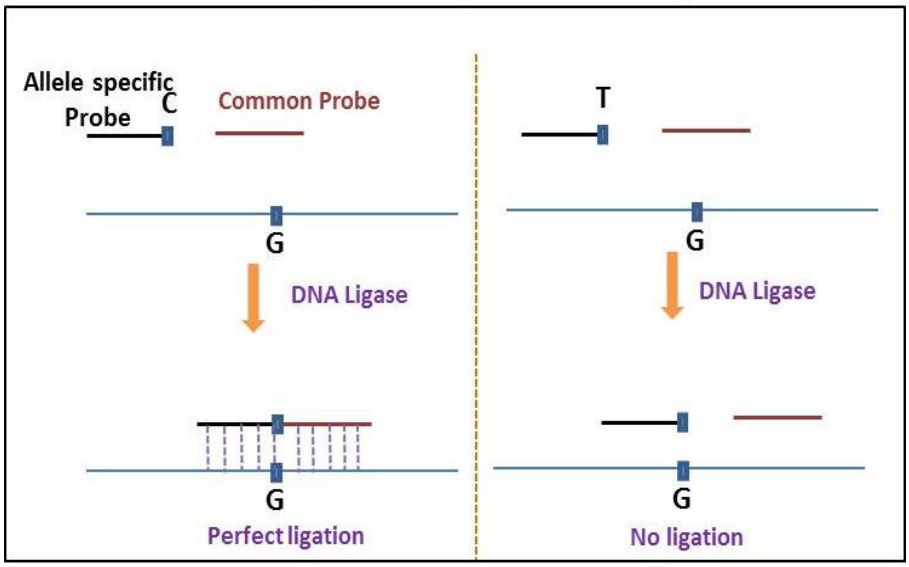


Figure 1.4: Allele Specific Oligonucleotide (ASO) ligation mechanism for SNP genotyping. The left hand side of the figure represents the presence of allele specific probe leading to perfect ligation. The right hand side of the figure shows no ligation in the absence of allele specific probe.

1.7.5 Invasive cleavage

The invader assay is based on the specificity of recognition, and cleavage, by Flap endonucleases, of the three-dimensional structure formed when two overlapping oligonucleotides hybridise perfectly to a target DNA (see Figure 1.5) [96]

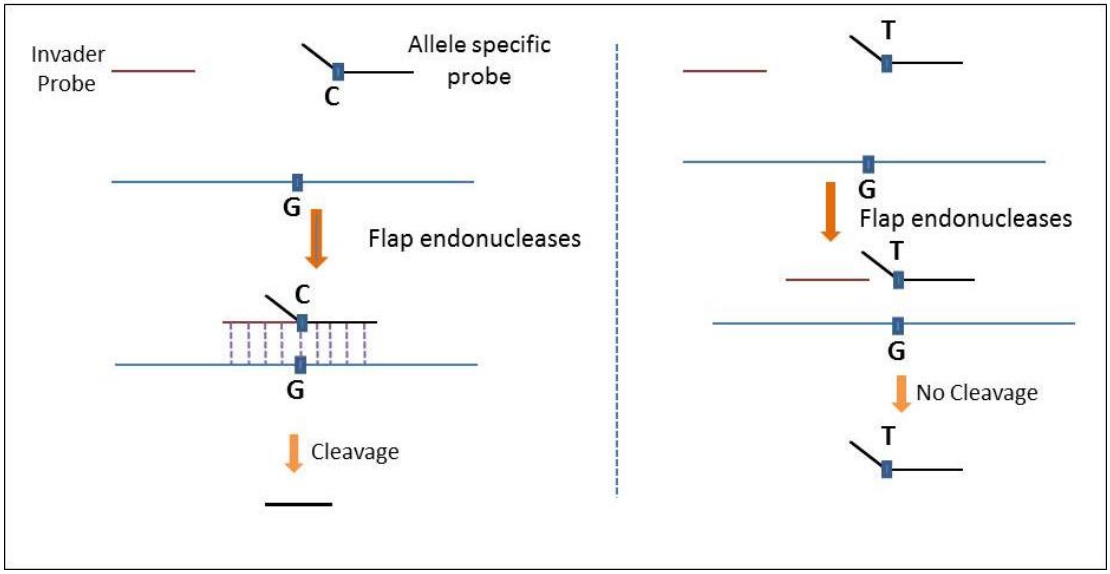


Figure 1.5: Invasive Cleavage mechanism for SNP genotyping. The left hand side of the figure represents the presence of allele specific probe leading to perfect cleavage. The right hand side of the figure shows no cleavage in the absence of allele specific probe.

1.7.6 Melting temperature

DNA melting temperature (melt curve) analysis is a post-PCR analysis method based on biophysical measurement of the amplified DNA. Figure 1.6 illustrates the steps involved in HRM analysis workflow. It has been used for various applications such as to detect primer/dimers and detection of genetic variation in DNA sequences. The most common method used today is referred to as HRM analysis. In HRM analysis, the target DNA is amplified by PCR in the presence of double stranded (ds) DNA binding fluorescent dyes, followed by gradual melting of PCR product through a range of temperatures. The emitted fluorescence is detected and characteristics melting curves are generated [99].

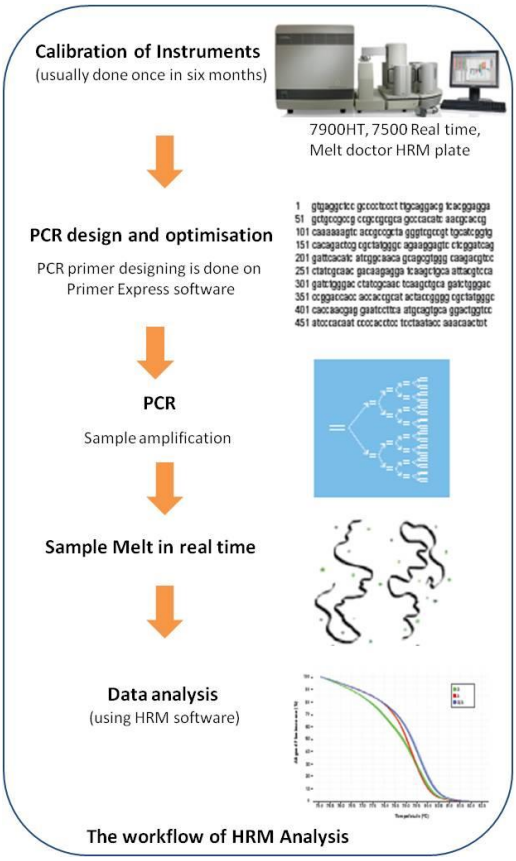


Figure 1.6: Steps involved in HRM analysis (adapted from [115]). The PCR design and optimisation is followed by PCR amplification. This is followed by post-PCR Melt analysis in real-time.

1.7.7 DNA sequencing technologies

DNA sequencing technologies were developed historically from first generation to third generation over a last few decades.

1.7.7.1 First generation sequencing

Alan Coulson and Frederick Sanger's 'plus and minus' system and Allan Maxam and Walter Gilbert's chemical cleavage technique in 1975 began the DNA sequencing era [71, 116]. The first DNA genome sequenced was of bacteriophage φX174 (PhiX), which used the plus and minus system [116]. Maxam and Gilbert's system used chemical cleavage to fragment DNA at specific bases, which was widely adopted [71]. However, the DNA sequencing revolution began in 1977, when Sanger developed the chain termination dideoxy technique for DNA sequencing [70]. This method used the chemical analogous of deoxynucleotidetriphospates (dNTPs). ddNTPs do not have 3' hydroxyl group, which is required for DNA extension and therefore terminates chain reaction by not forming bond with the 5' phosphate of the next dNTP. The first developed methods used radiolabelled ddNTPs electrophoresed in four parallel lanes of polyacrylamide gels, utilising autoradiography to identify the corresponding radioactive band to determine which nucleotide was being incorporated [70]. There were several improvements over the years in the Sanger sequencing methods, with radiolabelled ddNTPs replaced by fluorescently labelled ddNTPs and the use of CE as a detection system [117, 118]. This led to the development of automated sequencers with ABI PRISM range [119] used in the completion of the Human Genome Project and the first draft of human genome produced in 2001 [120].

1.7.7.2 Second generation sequencing

The need for higher throughput sequencing led to the development of sequencing technologies that could sequence multiple DNA regions in parallel. This is more commonly referred to as next generation sequencing (NGS) or MPS. Second generation sequencing is classified into main two types: SBS and sequencing by ligation (SBL) [121]. The overview of the workflow of SBS and SBL sequencing methods is provided in Figure 1.7.

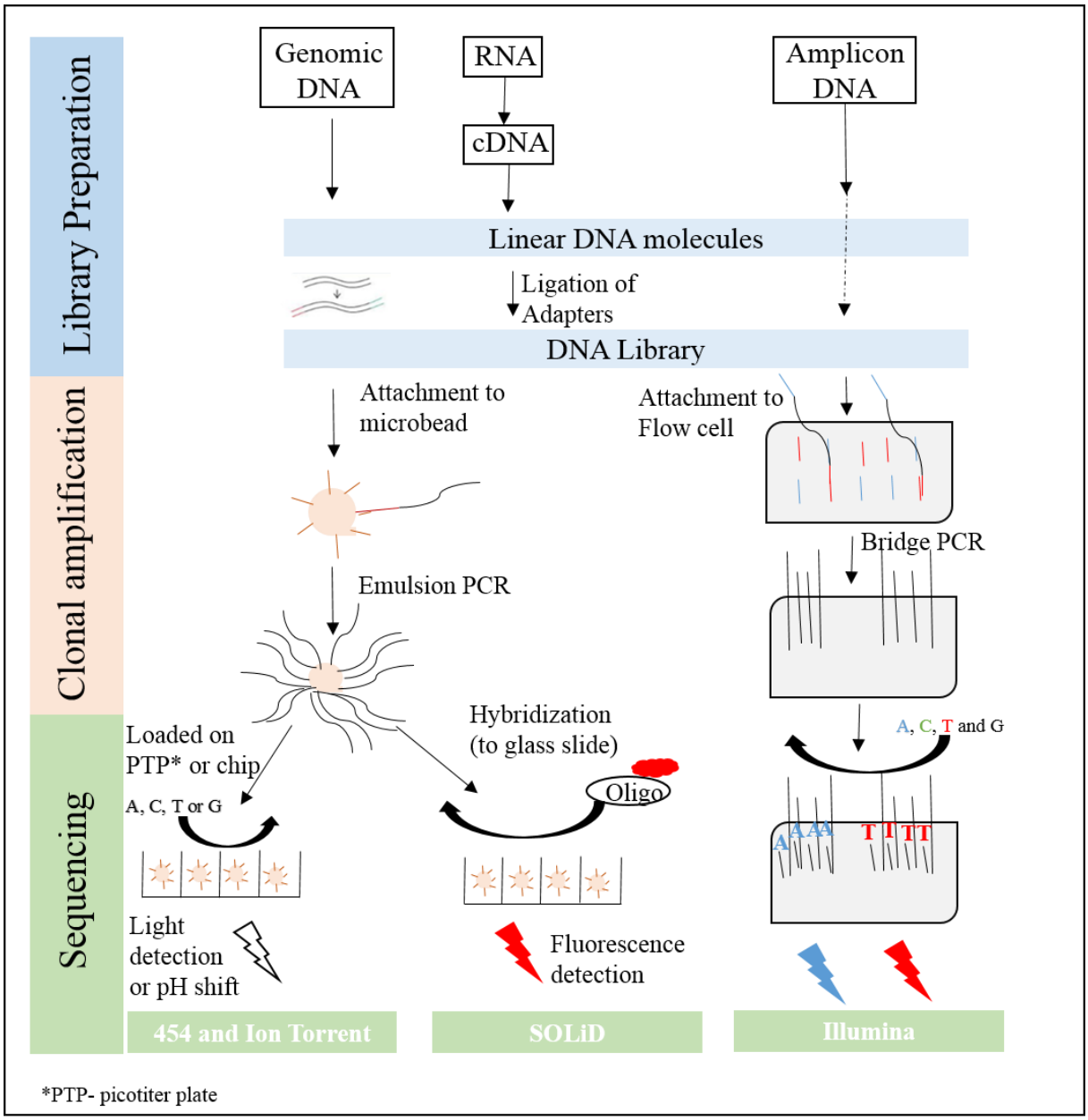


Figure 1.7: Workflow of the most common SBS and SBL chemistries (adapted from [122]). 454, Ion Torrent and Illumina sequencing chemistries are SBS whereas SOLiD represents SBL chemistry.

1.7.7.2.1 Sequencing by synthesis

There are three main types of SBS technologies: pyrosequencing, sequencing by reversible termination and sequencing by detection of hydrogen ions [121].

1.7.7.2.1.1 Pyrosequencing

The technique that established NGS used the luminescent method of measuring pyrophosphate sequencing. This method was commonly known as pyrosequencing. The sequence was inferred by pyrophosphate production as each nucleotide washed through the system of template DNA

affixed to the solid phase [123]. Despite differences between Sanger and pyrosequencing, both methods are SBS techniques, as they rely on DNA polymerase to produce observable output measurement. In this method, single-strand DNA with an annealed primers and four enzymes—DNA polymerase, luciferase, apyrase and adenosine triphosphate (ATP) sulfurylase—are present in the reaction mixture, which is followed by the addition of four nucleotides. If the added nucleotide is complimentary to the target DNA, the DNA polymerase incorporates nucleotide, resulting in the release of pyrophosphate, which is converted to ATP by ATP sulfurase. Luciferase uses this ATP to generate detectable light signal and intensity of this signal is proportional to the number of specific incorporated nucleotides. The excess nucleotides are degraded by apyrase (see Figure 1.8). The light signal is not produced if the added nucleotide is not incorporated by DNA polymerase [124, 125]. This technology requires the preparation of single stranded templates from PCR products library prior to analysis, which is one of the limiting factors along with their low multiplexing capabilities and inaccurate photopolymer sequencing [7, 121].

Pyrosequencing was licensed to 454 Life Sciences, a biotechnology company led by Jonathan Rothberg. The first commercial second generation sequencer was named 454, which allowed mass parallelisation of sequencing reactions and increased the amount of DNA that could be sequenced in a run. Libraries of DNA molecules first attached on the beads using adapters, in which ideally a single DNA molecule was coated on one bead and clonally amplified in their own emulsion droplet. These DNA coated beads were washed over a picotitic plate that fit one bead per well; pyrosequencing occurred as bead-linked enzymes and dNTPs were washed over the plate, and the release of pyrophosphate measured by charge coupled device (CCD) sensor under the well. The 454 sequencing was capable of producing long reads up to 500bp, for millions of wells together [126]. The first commercial high-throughput NGS sequencer widely used by customer was GS 20 in 2005, which was superseded by the GS FLX. The benchtop GS-Junior was also released. The 454 sequencer was bought by Roche in 2007, which closed the 454 sequencers business in the global market in mid-2016 [127].

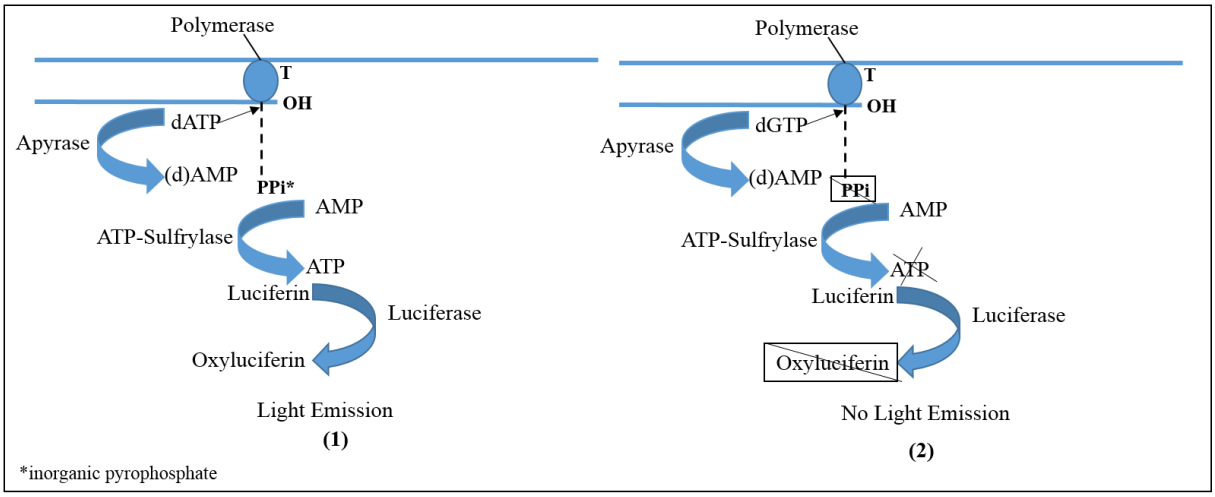


Figure 1.8: Enzymatic reactions involved in Pyrosequencing (1) during complementary nucleotide incorporation and (2) when nucleotide is not incorporated (adapted from [121]).

1.7.7.2.1.2 Sequencing by reversible termination

The earlier developed sequencing by reversible termination technology was Solera sequencing, which was acquired by Illumina in 2006 [128]. In this method, adapter-attached DNA libraries are passed over a flow cell with a lawn of bound complementary oligonucleotides. A clonal amplification PCR occurs at each DNA molecule, commonly known as bridge amplification (or bridge PCR) as the replicating DNA strands need to arch over to prime the next round of polymerisation off neighbouring surface-bound oligonucleotides [128]. This is followed by sequencing by reversible termination using fluorescently labelled reversible terminating dNTPs, which halts the binding of further nucleotide due to fluorophore occupying the 3' hydroxyl position. Fluorophore is cleaved for polymerase to continue their activity and sequencing happens in synchronous manner (see Figure 1.9) [129]. At each cycle, these modified dNTPs and DNA polymerase are washed on the flow cell and the incorporated base is identified by CCD, measuring the fluorophore excitation by specific laser. Initial genome analyser (GA) produced very short reads (up to 35bp) but offered paired-end advantage. Later, GAIIx was developed, replaced by four-channel sequencing system HiSeq that, which provided greater read lengths and depths as these instruments are used in whole genome and exome sequencing application [130]. In 2011, Illumina released the benchtop sequencer MiSeq, wherein each base was detected by individual image and provided less throughput compared to Hisses [98]. This technology performs paired-end sequencing, allowing users to sequence DNA targets from both ends [121].

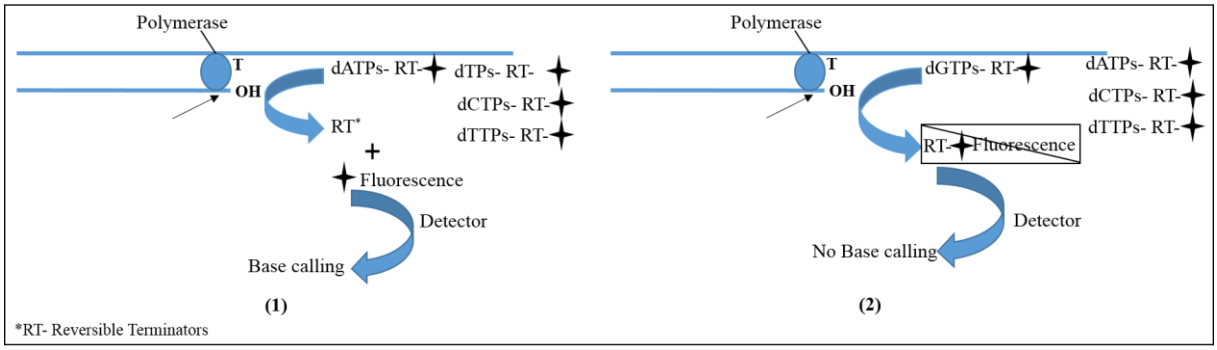


Figure 1.9: Sequencing by reversible termination (a) during complementary nucleotide incorporation and (b) when nucleotide is not incorporated (adapted from [121]).

1.7.7.2.1.3 Sequencing by hydrogen ion detection

Ion Torrent sequencing was developed by Jonathan Rothberg and then acquired by Life Technologies (now TFS) in 2010. The first Ion Torrent sequencer was the Ion Personal Genome Machine (PGM) and was the first 'post-light sequencing' technology as it neither used fluorescence nor luminescence. The method performed bead-based clonal amplification using emulsion PCR in a manner analogous to 454 sequencing, but did not measure pyrophosphate production, as it measured the difference in pH from the release of H+ ions during polymerisation using complementary metal oxide semiconductor (CMOS) technology from microprocessor chip manufacturing [97]. Thus, referred to as semiconductor sequencing or pH-mediated sequencing [121].

The DNA target post-library preparation and clonal amplification is bound on the proprietary ion sphere particles (ISPs) present in the microwells of the semiconductor chips. A single type of dNTP flows on the chip at a time and the release of H+ ions signifies their incorporation. This results in a change of pH, which is detected by a sensing layer underneath microwells, which converts the chemical signal to digital and is measured as a voltage (see Figure 1.10). In comparison to other sequencing technologies, which use indirect laser scanners or CCD cameras, the detection is direct and independent of any imaging devices [121].

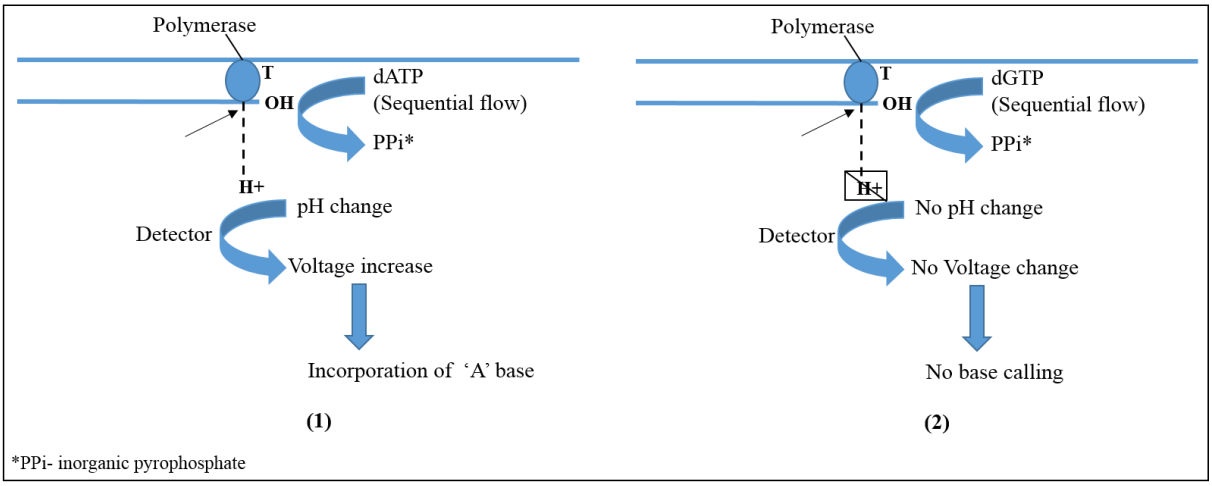


Figure 1.10: pH change involved in Sequencing by detection of hydrogen ions (1) during complementary nucleotide incorporation and (2) when nucleotide is not incorporated (adapted from [121]).

1.7.7.2.2 Sequencing by ligation

Another technology in second generation sequencing is sequencing by oligonucleotide ligation and detection (SOLiD) systems from Applied Biosystems (now TFS). SOLiD is an SBL-based system and does not use DNA polymerase (i.e., not SBS), based on 'polony' sequencing developed in 2005 [131].

This method involves the hybridisation and ligation of a variety of one or two base encoded probes to the target DNA. In this method, the oligonucleotide probe is generally eight to nine base pair long made up of one or two bases, followed by three degenerative bases and three universal bases, which are attached to a fluorescent label. A primer along with fluorescent labelled oligonucleotide probes, which anneals with target DNA having a complimentary sequence, are mixed. The probes ligates with primers using DNA ligase and are detected by fluorescent imaging. The non-ligated probes are washed away. The fluorescent label cleaves from the oligo-probes after each detection as they have cleavable linkage and become ready for the next round of ligation. This cascade continues until the target DNA sequence is complete (see Figure 1.11). At the end of each round, the sequence of known bases is only in positions other than the degenerative bases. The sequence of missing or skipped positions occur in the successive round with a shorter primer. Thus, the sequence of target DNA is completed using anchors of different lengths [121].

The sequencing chemistry only allowed read lengths from 35 bp to 85 bp, which was a major limitation [121]. The assembly of shorter sequences remained a challenge; however, SOLiD

offered a competitive cost per basis in comparison to Illumina [132]. SOLiD was discontinued by Life Technologies in 2016 [133].

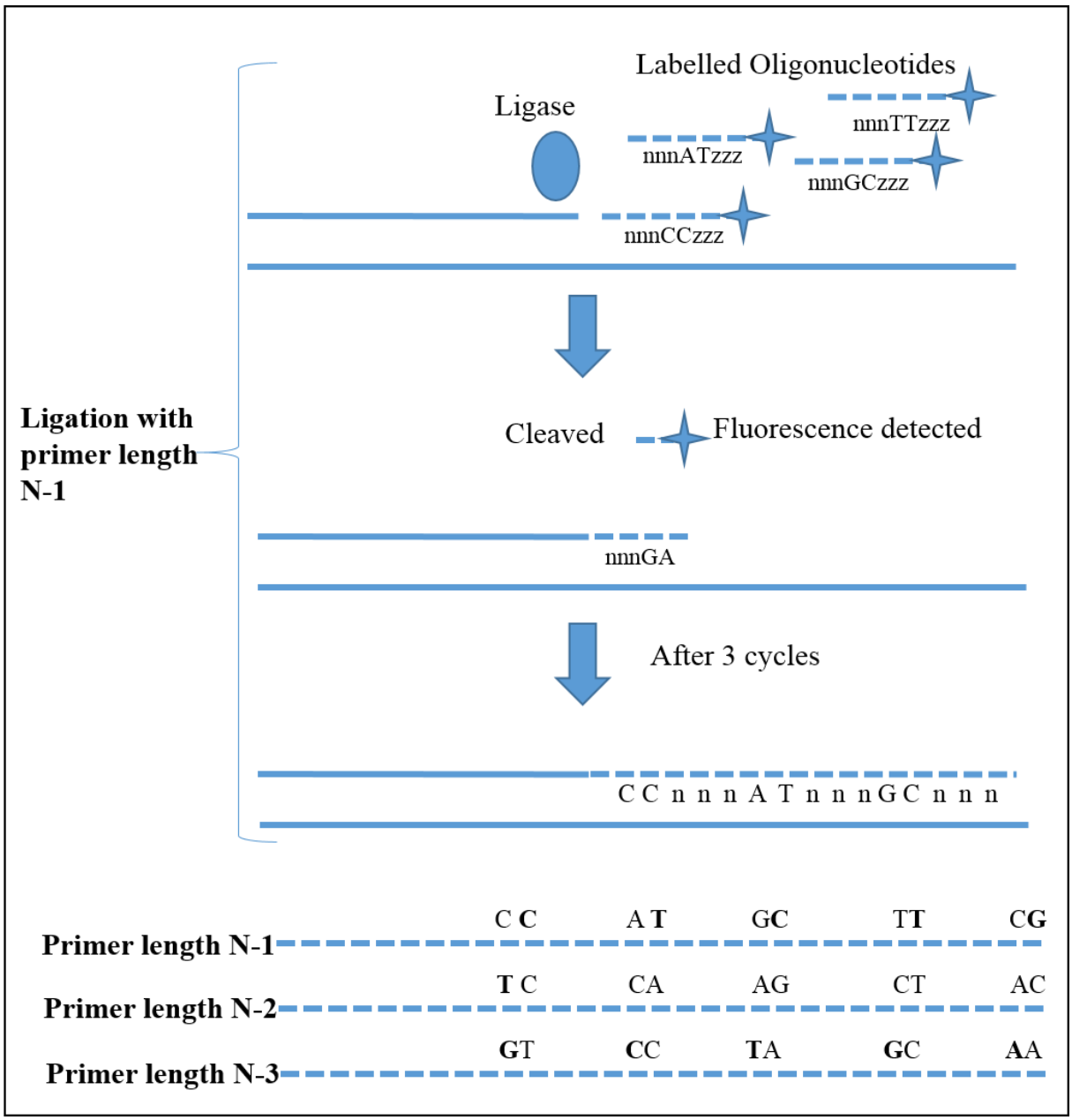


Figure 1.11: Diagrammatic representation of enzymatic reactions involved in SBL (adapted from [121]). The fluorescently labelled oligonucleotide probe ligates with a short primer which together hybridises with the target sequence being complementary followed by fluorescent imaging detection. The cleavable linkage cleaves after each detection and preparing the system for another round of ligation. The cascade continues until the target sequencing gets completed.

1.7.7.3 Third generation sequencing

The second generation sequencing technologies have two main general challenges: i) they cannot perform long read sequencing; and ii) PCR bias at the base detection/incorporation level

due to clonal amplification [127]. The third generation sequencing technologies were developed mainly to overcome above challenges. In these methods of sequencing, the DNA template is sequenced in real-time and are referred to as single molecule real-time (SMRT) sequencing. The use of biochemicals are minimised, which led to the miniaturisation of the entire process to nanoscale level. There are three main technologies in this category: Pacific Biosciences SMRT, Helicos single molecule sequencing (SMS) and Oxford Nanopore DNA sequencing [121].

1.7.7.3.1 Helicos single molecule sequencing

The first SMS was developed by Stephen Quake [134], which was then commercialised by Helicos Biosciences. The library preparation does not require ligation or amplification. The DNA template is sheared, tailed with poly-A and blocked at 3'-OH end using terminal transferases and a dNTP. These poly-A fragments are hybridised on to the flow cell surface with oligo-dT for initiating SBS. The Helicos Scope sequencers utilise fluorescent labelled nucleotides for sequencing DNA template that are attached to the flow cell through poly-T tails [121]. The DNA template gets attached to the planar surface and then proprietary fluorescent reversible terminating dNTPs, as virtual terminators[135] are washed over one base at a time and imaged.

This first non-amplified DNA-based technology was relatively slow and expensive compared to second generation sequencing [136]. The technique also allowed sequencing and quantitation of RNA directly without requiring cDNA conversion. The technology only offered short read sequencing lengths (24 to 70 bases) and generated data output up to 20 GB [121]. Helicos filed for bankruptcy in 2012 [137].

1.7.7.3.2 Pacific BioSciences SMRT

Pacific Biosciences is another third generation sequencing platform based on SMRT technology. DNA polymerisation happens in zero mode waveguides (ZMWs), which are nanostructures on the metallic film covering a chip. ZMWs have illuminated regions and DNA polymerases are deposited in these regions, so sequencing happens with the wash of DNA libraries and fluorescent labelled dNTPs. The extension of DNA is monitored in real-time and fluorescence is detectable only from nucleotides in which dye is cleaved [138]. This allows SMS in a very short time [139]. PacBio platforms are capable of producing long reads, up to ~40KB in length and the technology can be utilised in de novo assemblies [136, 140]. However,

the technology has an accuracy of 85 percent, which is much lower than the second generation sequencing technologies [121].

1.7.7.3.3 Oxford Nanopore DNA sequencing

Oxford Nanopore Technologies was the first company to release nanopore sequencing platforms, GridION and MinION [141, 142], the latter being only 10cm in size and 100g in weight. The system is primarily comprised of a nanopore embedded in the artificial membrane and a motor protein molecule assists in the translocation of DNA molecules from one side of the membrane to the other through the nanopore. An electric current runs through the nanopore and as DNA molecules passes through the nanopore, the current is modified, with each base having their signature effect. These voltage changes are converted to bases and help reconstruct the DNA sequence. The technology currently has very high error rates, especially in homopolymers stretches of DNA [142, 143]. The technology is capable of directly sequencing RNA molecules, eliminating the need of performing reverse transcriptase PCR [144].The MinION nanopore sequencer has 512–2000 nanopores and each nanopore has a sequencing speed of 120–1000 bases per minute. The sequencer is like a USB and can be used only once. This technology can allow sequencing to be performed directly in field, reducing cost, time and effort tremendously, although it requires improvement in their current sequencing error rates [121].

1.7.7.4 MPS for forensic applications

Sanger sequencing would not be suitable for forensics identification and FDP as it is low-throughput. Sanger sequencing is used in forensics for mtDNA sequencing, but the resolution may be low [145]. On the contrary, third generation sequencing technologies are not mature enough to gain applicability in forensics [146]. Second generation sequencing, or MPS, has been around for more than a decade and displays potential for forensic applications. MPS offers a battery of capabilities for forensics including: i) simultaneous analysis of multiple markers in multiple samples, which reduces depletion of evidence material [147]; ii) improved mixture analysis [148]; iii) FDP capability [149]; iv) whole mitochondrial DNA sequencing [145]; v) RNA sequencing of body fluid specific markers [145]; and vi) sequencing of epigenetic markers for age estimation [150].

MPS technologies can perform a range of applications, including whole genome (de novo) sequencing, exome sequencing, transcriptomic sequencing and targeted amplicon sequencing.

Whole genome sequencing capabilities are useful for mitochondrial applications in forensics. Targeted sequencing is the main capability of MPS useful for forensics as it allows analysis of specific forensically informative markers—SNPs, STRs and INDELs. Targeted MPS in forensics will enable the genotyping of hundreds of markers for multiple samples. The technology offers potential to be used in the complete forensic DNA analysis spectrum of identification and phenotyping (see Figure 1.12). MPS benchtop platforms would be most suitable to forensics for the required throughput for targeted sequencing applications. There are two main platforms for forensic applications: Ion™ PGM and GeneStudio™ S5 (TFS); and MiSeq (Illumina)/MiSeq Forensic Genomics (FGx™) (Verogen).

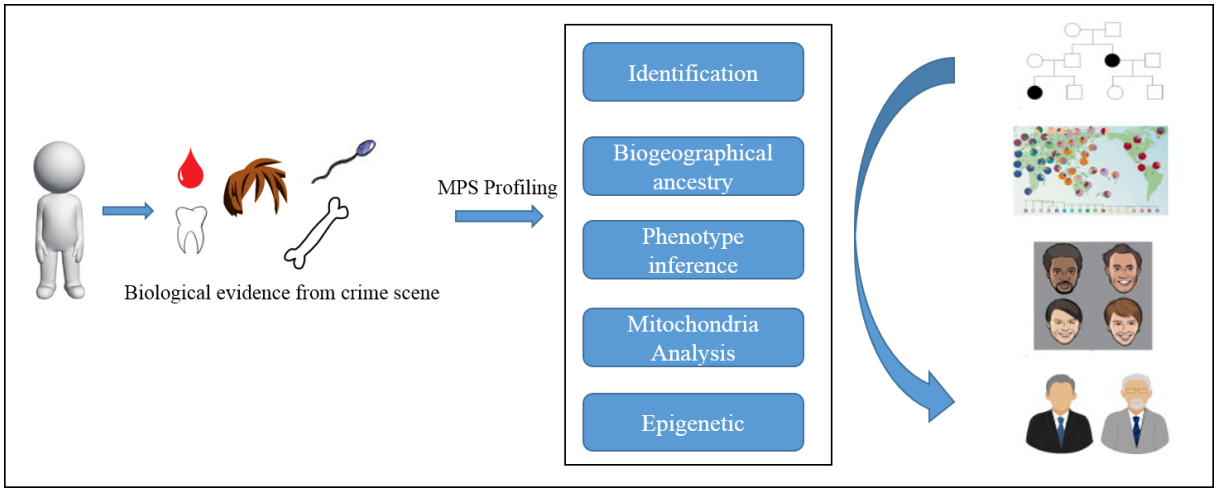


Figure 1.12: Potential of MPS to provide forensic identification and forensic phenotyping information spectrum (adapted from [145]).

1.7.7.4.1 Ion Torrent sequencers

Ion Torrent benchtop sequencers, Ion PGM or Ion GeneStudio S5 (TFS), offer simple, scalable, faster technology for forensic applications. Ion sequencers convert the pH change to voltage signals, which is detected by semiconductor chips. Each voltage change corresponds to the nucleotide base (A, T, G or C) being sequenced [97]. The entire process of sequencing involves three main steps: library preparation, template preparation and Ion chip sequencing.

1.7.7.4.1.1 Ion library preparation

The process involves the amplification of DNA templates and attachment of DNA barcodes. DNA barcode labelling enables multiplexing of multiple samples [151]. Once the genomic DNA is isolated and quantitated, the DNA template is subjected to library preparation. The general input amount of genomic DNA required to initiate library preparation is 1ng. Ion AmpliSeq library preparation chemistry also allows the use of lower quality and input of DNA,

with slightly higher PCR amplification cycles [152]. The first step involves the multiplex PCR amplification of targets using the specific primers. Then the amplicons are subjected to partial digestion using FuPa reagents to create blunt ends for adapters and DNA barcodes ligation [152]. The next step is the ligation of Ion P1 adapters and Ion Xpress Barcodes, with the partial digested amplicons using DNA ligase. This is followed by library purification, which involves removal of excess dNTPs—non-ligated adapters and barcodes. Library purification is generally performed used magnetic beads, such as Agencourt™ AMPure XP reagents (Beckman Coulter) [152]. Following purification, library normalisation is performed using equaliser beads, qPCR or Qubit (TFS)/Bioanalyzer™ (Agilent Technologies). The aim of library normalisation is to add the normalised (equimolar) quantities of each libraries for the next step of sequencing. The Ion Library Equilizer™ kit (TFS) provides a method of library normalisation to ~100pM using equaliser beads [152]. qPCR provides more accurate quantitation using the specific TaqMAN probes targeting the Ion P1 adapter sequence. qPCR method takes little longer compare to bead equaliser method [152, 153]. Figure 1.13 demonstrates the steps of Ion Torrent AmpliSeq library preparation (adapted from [154]).

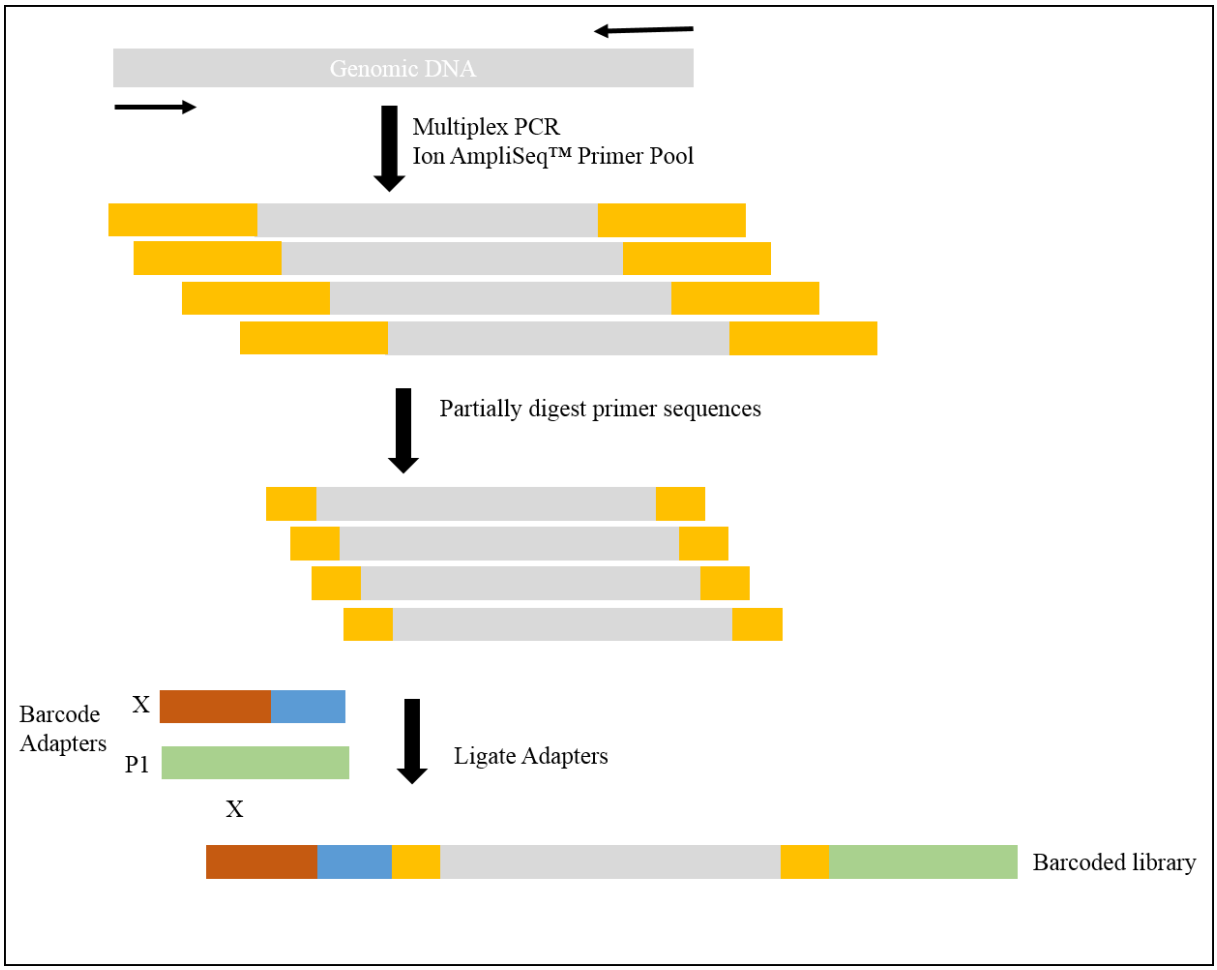


Figure 1.13: Overview of the Ion AmpliSeq library preparation step (adapted from [154]). The first step is multiplex PCR which generates amplicons followed by partial digestion of primer sequences. These are further ligated to DNA barcodes and adapters which makes a barcoded library.

1.7.7.4.1.2 Ion Template preparation and chip loading

The main purpose of template preparation is to create clonal copies of each library, which is also referred to as clonal amplification. Each ISP consists of a lawn of complementary oligonucleotides to Ion P1 adapter sequence. The amplification is monoclonal only if one library DNA fragment binds to an ISP. This is done using oil-in-water emulsion, known as emulsion PCR. Template preparation using manual approach requires Ion One Touch™ 2 (OT2) System (TFS) [147].

The normalised libraries, along with all standard PCR components—buffers, dNTPS, DNA polymerase—are added to ISPs, which provide primers and oil. The Ion OT2 system creates an emulsion environment optimised to achieve 1:1 ratio of bead to library fragment. The library fragment binds to the ISP via the P adapter. Emulsion PCR begins simultaneously for millions of ISPs. As a final step, NaOH is added to separate the strands. The strands not attached to ISPs will be washed away, leaving clonally amplified ISPs ready for sequencing. These ISPs are loaded onto semiconductor chips for sequencing. The manual loading is carefully performed to ensure all the microwells of chips are filled with ISPs. The size of microwells are designed in the chip in such a way that only one ISP fits each well [147]. The Ion Chef System (TFS) is another option that automates the procedure of template preparation and loading of chips [155]. After the chip is loaded, it is ready to be sequenced on Ion PGM or Ion Gene Studio S5.

1.7.7.4.1.3 Ion chip sequencing and data analysis

Once the chip is filled with clonally amplified ISPs, it is loaded onto the sequencer for sequencing. Each microwell on the semiconductor chip contains clonally amplified copies of a single stranded DNA template and DNA polymerase. The chip is flooded sequentially with unmodified dNTPs (A, C, G or T). If an introduced dNTP is complementary to the nucleotide on DNA template, H+ ions are released as it gets incorporated into the growing strand. These H+ ions bring a change in pH, which is detected by chips. The unattached dNTPs are washed away before the next round of flowing other dNTP into the chip [156, 157].

Ion-sensitive field effect transistor (ISFET) sensor layer is engineered beneath the layers of microwells. All layers are contained within a CMOS of the semiconductor chip, similar to a

chip used in computers. These ISFET sensors detect the pH change electric signals, which are translated into a DNA sequence on a computer [157, 158].

Data analysis involves four basic steps—signal processing, base calling, alignment and variant calling (see Figure 1.14). Signal processing—from electric to digital—occurs on the Ion Torrent suite (TS) software. Base calling is then performed by the TS software, which converts digital signals into bases. These unaligned bases are stored in FASTQ format files. The next step is alignment of the sequence to the reference genome and is stored in binary alignment (BAM) file formats. The Torrent Variant caller plugin is used to make variant calls from the BAM files and is stored in the variant call file format [147]. Ion Torrent have also developed special plugins suitable for forensics, such as the HID SNP genotype plugin, which provides SNP genotypes of forensically interested SNPs for the samples included in the run [159].

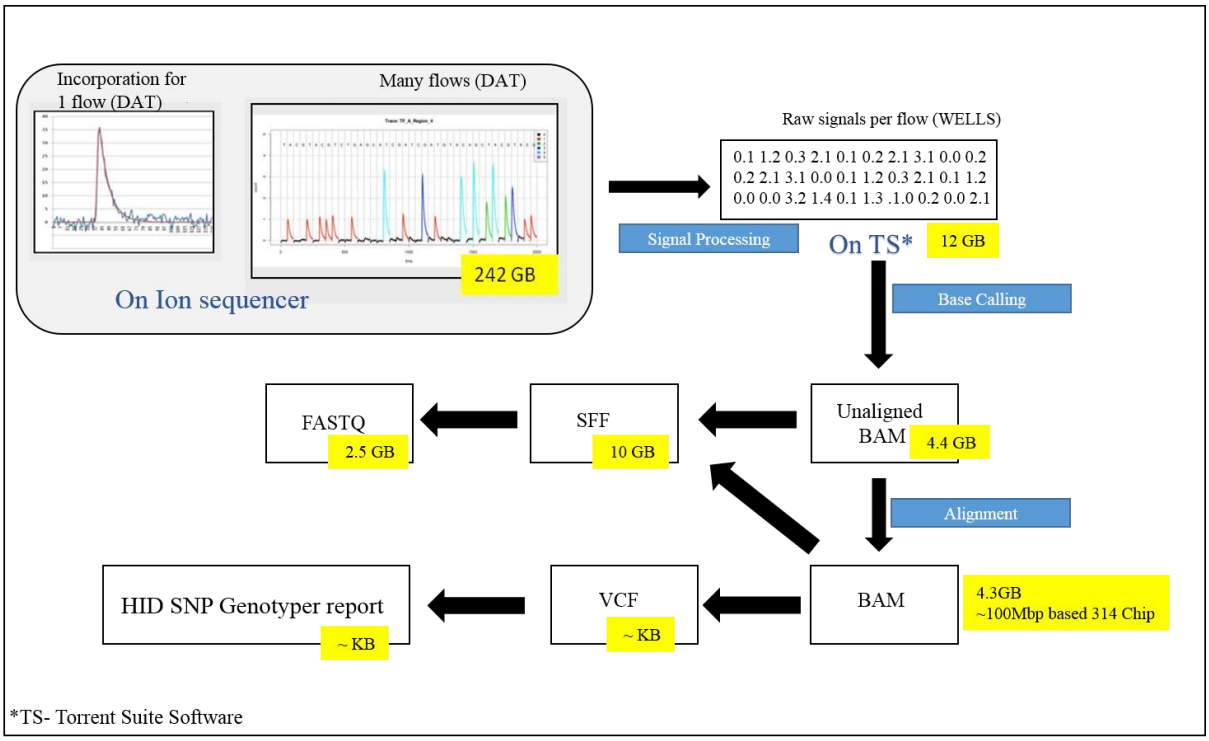


Figure 1.14: Overview of the Ion Torrent analysis steps. Conversion of signal processing input files (DAT) to binary raw signals represents the signal processing step. These binary signals are in turn converted to unaligned binary alignment (UBAM) files after a baseCalling step by the Torrent Suite Software. UBAM files are then converted to FASTQ files via a binary file in standard flowgram format (SFF) as well as to BAM files which produce a variant calling file (VCF) after a variant calling step. The HID SNP Genotyper plugin provides genotypes for forensic application from VCF files.

Ion PGM systems offer scalable capabilities using three chips: Ion 314™ chip v2 (Output: ~60–100 MB); Ion 316™ chip v2 (Output: ~600 MB–1 GB); and Ion 318™ chip v2 (Output: ~1.2–2 GB). The recently released Ion GeneStudio S5 series of MPS platforms offer scalable options

from 2 million reads to 130 million reads using their new chips series: Ion 510™ chip (Output: ~300–500 MB); Ion 520™ chip (Output: ~600 MB–1 GB); Ion 530™ chip (Output: 3 GB–5 GB); Ion 540™ chip (Output: ~10–15 GB); and Ion 550™ chip (Output: ~20–25 GB). Ion Torrent have many commercial MPS panels released for forensic application, such as: Precision ID GlobalFiler (TFS), a NGS STR panel capable of genotyping 21 autosomal STRs for forensic identification; Precision ID Ancestry panel (TFS), with the ability to genotype 165 autosomal SNPs for inferring BGA of a sample; and Precision ID mtDNA Whole Genome Panel (TFS) [160], which allows forensic users to interrogate the mtDNA for paternity and kinship applications. The Ion AmpliSeq Designer (TFS) tool also offers potential to build custom panels and access DNA phenotyping community panels such as HIrisPlex [161].

1.7.7.4.2 Illumina MiSeq sequencer

Illumina MiSeq or Verogen MiSeq FGx benchtop sequencers involve the incorporation of fluorescently labelled dNTPs into the DNA template strand during sequencing cycles of DNA synthesis catalysed by DNA polymerases. The nucleotides are identified by fluorophore excitation at the time of incorporation during each cycle [162]. There are three steps in the sequencing process: library preparation, sequencing and data analysis.

1.7.7.4.2.1 Library preparation

TruSeq Amplicon library preparation protocol is used for preparing libraries via a ligation approach [163]. The input material is generally PCR amplicons or products. The input amount can vary between 1ng and 100ng [165]. PCR product quantitation is recommended using Qubit (TFS) and size check using Bioanalyser (Agilent Technologies). PCR products are first subjected to an end repair step, which converts the 5’ overhangs from incomplete polymerisation during PCR to blunt ends. The phosphorylation of the 5’ ends also happens in this step [165]. The magnetic bead clean-up removes excess and unattached reagents. The next step involves the adenylation of 3’ ends, in which a single ‘A’ nucleotide is added to the 3’ blunt end of PCR amplicons mainly to avoid their ligation to each other during adapter ligation step [165]. The adapters have corresponding ‘T’ nucleotide on the 3’ end, which becomes complimentary to the added A nucleotide. Adenylation is followed by RNA adapter ligation [165]. Each index has overhangs of P5 (5’ AAT GAT ACG GCG ACC ACC GA 3’) and/or P7 (5’ CAA GCA GAA GAC GGC ATA CGA GAT 3’) adapters. Both P5 and P7 adapters are required for paired-end sequencing. The entire index-adapter oligonucleotide is ligated with DNA amplicon/template [165]. The ligated product is referred to as a library (see Figure 1.15).

This is followed by PCR enrichment of libraries, in which only those fragments that are ligated with adapters are selectively enriched, using PCR primers complementary to the P5 and P7 adapter sequences. Library purification is then undertaken, using magnetic beads to remove excess reagents and non-adapter ligated DNA fragments. This is followed by library evaluation, using Bioanalyser (Agilent Technologies) to ensure the correct sized libraries are present. Library normalisation is undertaken, using Qubit (TFS), Bioanalyser (Agilent Technologies), qPCR or equaliser beads. The normalised libraries pool (~10 nM) is made single stranded, using NaOH, and a small amount is loaded onto the flow cell for sequencing on MiSeq [163, 164].

Verogen ForenSeq DNA library preparation uses a different library preparation workflow than the one detailed above. It is a PCR-based library preparation workflow rather than a TruSeq ligation-based approach. Genomic DNA is first subjected to PCR, using ForenSeq primer sets to generate PCR amplicons. Then, a second PCR performed to attach i5 and i7 indices to the DNA fragments (see Figure 1.16). They use universal primer sequences complimentary to adapter sequences for attaching indices, barcodes and adapters to the DNA template [165]. The libraries are then purified, using magnetic beads to remove excess reagents and unattached oligonucleotides. Purified and normalised libraries are recommended, using normalisation beads to achieve consistent cluster density for each library [165]. The normalised libraries are pooled and denatured to single stranded, using NaOH and loaded onto the flow cell for sequencing on MiSeq FGx (Verogen) [166].

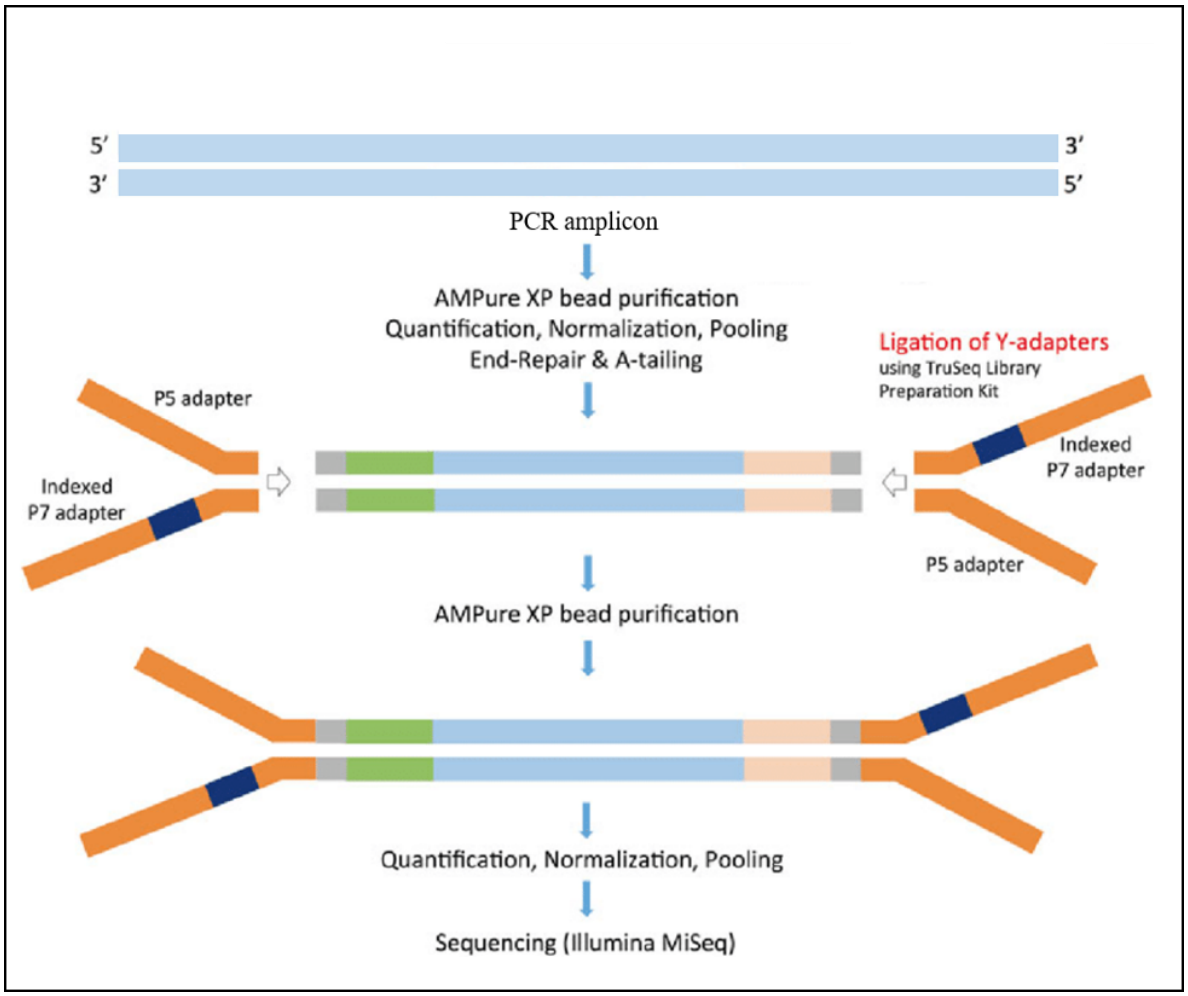


Figure 1.15 Overview of the library preparation step using Illumina TruSeq ligation-based chemistry (adapted from [167]). The steps involves magnetic bead clean up, End-repair & A-tailing of PCR amplicons followed by ligation of P5/P7 Y adapters which makes a barcodes library.

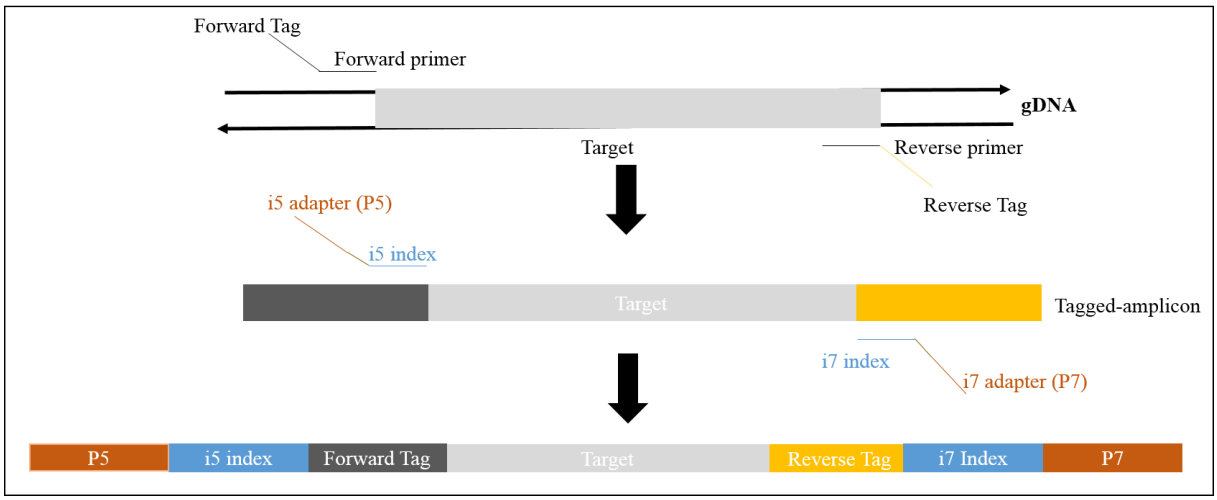


Figure 1.16: Verogen ForenSeq DNA Signature library preparation workflow (adapted from [168]). This process involves the PCR amplification to generate PCR amplicons followed by attachment of barcodes using universal PCR step to generate libraries.

1.7.7.4.2.2 Sequencing and data analysis

The normalised libraries are loaded onto the flow cell for sequencing. MiSeq flow cell is a glass-based substrate, in which cluster generation and sequencing occurs. These are single use and have a lawn of hanging oligos complementary to P5 (i5) or P7 (i7) adapters, which are found in DNA libraries [164, 169]. DNA libraries bind to flow cells using the above oligos and clonal amplification process—also known as cluster generation—via a bridge PCR process (see Figure 1.11). Once the strand binds to complementary oligos on the flow cell, DNA polymerase extends the strand. The forward strand is washed away and a cluster generation of newly synthesised (reverse strand) strand begins. The newly extended strand bends and attaches to another adapter complimentary oligo sequence and the polymerase extends again, denaturing and washing the first strand once the extension is complete. The bridge PCR clonal amplification of hundreds to thousands of DNA libraries happens simultaneously, potentially generating millions of clonal clusters [169].

At the end of clonal amplification, the reversed strands are washed away, leaving only forward strands. Sequencing primers anneals to the forward strand and polymerase extension commences, using fluorescently labelled dNTPs. A single base is added per cycle as reversible terminators on each nucleotide, which prevents the addition of other bases during the same cycle. The fluorescence is emitted and detected using a CCD camera. Reversible terminators are cleaved before the next cycle of sequencing and the process repeats. Once the forward strand sequencing is complete, it will sequentially sequence from Index 1, Index 2 and finally from the reverse strand until the paired-end sequencing is complete [169, 170] (see Figure 1.17).

Figure 1.17: MiSeq (Illumina) paired-end sequencing sequential order: 1) sequencing of forward strand; 2) sequencing of Index 1; 3) sequencing of Index 2; and 4) sequencing of reverse strand (adapted from [171]).

Data analysis on MiSeq occurs in the following sequence. First, Miseq control software (MCS) on the sequencer process the images and perform base calling (.bcl file), along with quality scoring of bases. Sequencings Analysis viewer is used to monitor data. This is followed by the raw base calling files (fastq files) being processed by MiSeq Reporter (MSR) software for alignment (BAM files), assembly and variant calling (vcf files; see Figure 1.18) [172].

BaseSpace Hub (Illumina) is a cloud option of MSR for performing alignment, assembly and variant calling [173].

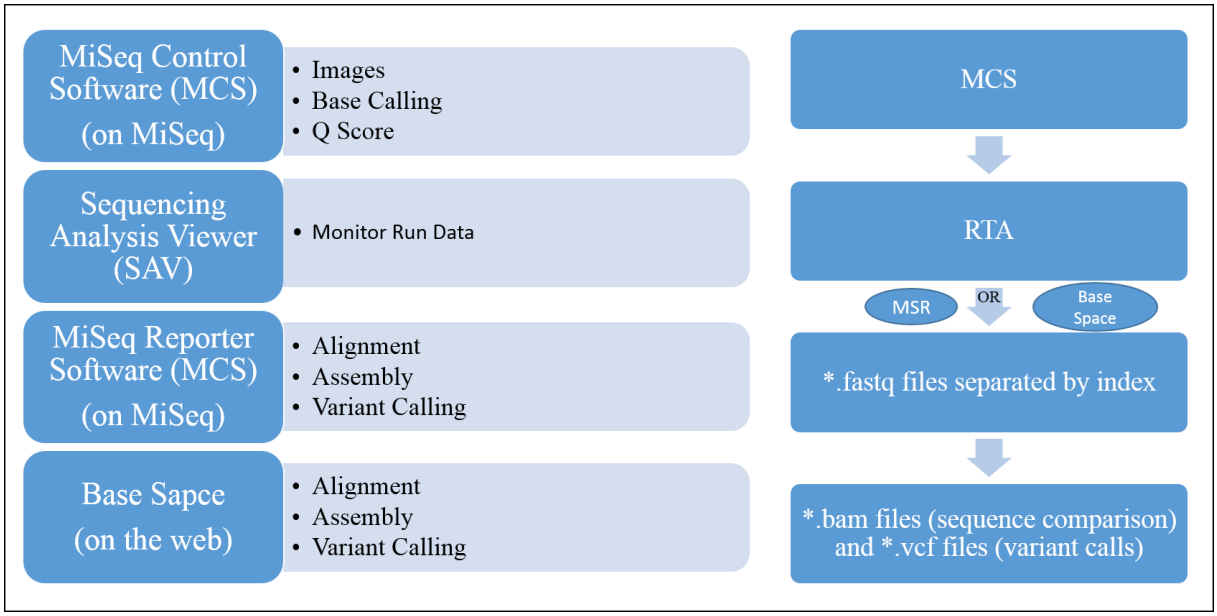


Figure 1.18: MiSeq (Illumina) data analysis pipeline. The primary analysis involves calling of bases followed by secondary analysis of alignment and variant calling.

Miseq FGx employs a different data analysis pipeline for ForenSeq DNA Signature Kit (Verogen) data. The sequencing is partial paired-end and only the last 31 bp is sequenced in reverse direction compared to 351 cycles in forward direction [174]. The analysis software package is ForenSeq Universal Analysis Software (UAS) (see Figure 1.19). The MCS component of UAS controls the image processing process, while the real-time analysis software component performs image analysis, base calling and quality scoring. ForenSeq UAS performs alignment, allele calling, genotyping and reporting [175], and can also generate pdf reports.

ForenSeq Universal Analysis Software	Run Set-Up
MiSeq Control Software	Sequencing Chemistry
MiSeq Control Software	Cycle-by-cycle chemistry
Real time Analysis Software	Image Analysis, Basecalling and Quality Scoring
ForenSeq Universal Analysis Software	Alignment, Allele Calling, Genotyping and Reporting

Figure 1.19: ForenSeq UAS analysis workflow (adapted from [176]).

Illumina have commercial mtDNA analysis kits [177], which allow forensic users to interrogate mtDNA for paternity and kinship applications [160]. For example, Verogen ForenSeq DNA Signature Prep Kit (Illumina) was the first all-in-one NGS solution for forensics, comprising of 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 IISNPs, 56 AISNPs and 22 PISNPs [168].

1.7.7.4.3 Other forensic MPS solutions

PowerSeq™ Auto/Mito/Y-system (Promega) has the ability to simultaneously analyse autosomal and Y-STRs, along with mitochondrial data from one sample [178]. PowerSeq uses the MiSeq (Illumina) platform. More recently, GeneReader (Qiagen) technology integrates fluorescent-based SBS chemistry with detection of the respective fluorescent signal templates that have been clonally amplified using GeneRead QIAcube [179]. Pyrosequencing is also known to be applied for methylation forensic application of age estimation [180].

1.8 Research aims

Forensic DNA identification cannot be utilised for cases in which there are no database matches, no suspects and missing eyewitness statements. FDP has the potential to provide leads for investigators, acting as a 'molecular eyewitness' for such cases when forensic DNA identification has failed to draw conclusions. SNPs are the most common markers for forensic DNA phenotyping, requiring forensic laboratories to type these markers to generate FDP profiles. There are a large number of SNP-typing technologies and each has their own requirements—such as the number of SNPs that can be typed, throughput suitability, platform availability—which makes it difficult for forensic laboratories to choose the most suitable technology. In addition, forensic laboratories also require platforms that are:

- simple to use
- cost effective
- flexible and modular
- sensitive and reproducible
- able to work with degraded DNA or low input amounts
- able to obtain maximum information from scarce evidentiary DNA material
- able to sequence a large battery of FDP markers.

Each platform is subjected to forensics validation criteria for it to be considered for routine samples testing. Due to the limited availability of resources, forensic laboratories cannot

conduct evaluations of all SNP-typing technologies and platforms using forensic validation criteria for FDP. The aim of this doctoral project is to assess and evaluate three SNP-typing techniques on key forensic validation criteria useful for FDP, to be used as a preliminary guide for forensic scientists. The three methods were representative of low-, medium- and high-throughput genotyping technologies. The research presented in this thesis provides a starting point for forensic personnel to investigate FDP processes.

Each SNP genotyping technique described above differs on throughput levels, in which each method can genotype a variable number of SNPs and samples in a run. The throughput requirements of each FDP application may vary. For example, one laboratory may be required to genotype only eye colour FDP assay, and that requires typing less than ten SNPs. Another laboratory may be required to genotype hair and eye colour, ancestry and baldness FDP assays, which require typing of larger number of SNPs (~100). Hence, the SNP-typing techniques investigated in this thesis are representative of three categories: low-, medium- and high-throughput.

HRM is a low-throughput SNP genotyping technique capable of genotyping a small number of SNPs (<10). SNaPshot™ (TFS) is the most common SNP-typing method currently employed in forensics and a variety of SNaPshot™ FDP assays exist. It is a medium-throughput SNP genotyping technique. This thesis also aims to list the common forensically relevant SNaPshot™ SNP-typing assays. Finally, MPS is a high-throughput SNP genotyping technique, which has the potential to simultaneously genotype hundreds of SNPs for multiple samples.

The research aims of this thesis are:

1 Survey the most common forensic SNP assays that can be typed, using existing CE-based detection systems in forensic laboratories.
2 Assess and evaluate HRM analysis as a low-throughput genotyping tool for FDP SNP-typing.
3 Assess and evaluate SBE-based SNaPshot™ as a medium-throughput tool for FDP SNP-typing.
4 Assess and evaluate Illumina MiSeq as a high-throughput tool for FDP SNP-typing.
5 Compare and contrast these three indicative SNP genotyping technologies, describing their advantages and disadvantages for particular applications.

1.9 Chapter descriptions

1.9.1 Chapter 2: Forensically relevant SNaPshot™ assays for forensic SNP genotyping

CE-based fragment analysis, SNaPshot (TFS) is the most common method used for forensic SNP-typing. A large number of forensic assays have been typed using this technique. This chapter lists the forensically relevant SNaPhsot™ SNP assays categorised in to four: identity informative SNP assays; lineage informative SNP assays; ancestry informative SNP assays; and phenotype informative SNP assays. Chapter 2 addresses the gap in the field, in which a quick guide is required for forensic scientists to select the most appropriate set of markers for their respective applications. Chapter 2 is presented as a published paper:

> Mehta B, Daniel R, Phillips C, McNevin D (2017) Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. International Journal of Legal Medicine 131(1): 21–37.

1.9.2 Chapter 3: Low- and medium-throughput genotyping tools: High resolution melting and single base extension (SNaPshot™)

HRM analysis is a post-melting PCR technique that utilises real-time PCR technology. The method has been utilised in clinical diagnostics for SNP genotyping for a small number of SNP panels [181, 182]. Chapter 3 describes HRM forensic assessment on criteria including: sensitivity, reproducibility, multiplexing capability and the effects of different DNA isolation methods. The assay was assessed using the six SNP IrisPlex assay for eye colour prediction.

The SNaPshot™ single base extension assay uses the CE-based detection system available in forensic laboratories and, is capable of multiplexing up to 40 SNPs [183]. Chapter 3 also describes the forensic evaluation of SNaPshot™ SNP genotyping on criteria, including sensitivity, reproducibility, multiplexing capability, mixture detection and cost efficiency. It is compared with the HRM method using the same IrisPlex eye colour prediction assay. This chapter is presented as three **published** papers:

> Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35 (21–22):3036–3043.

> Mehta B, Daniel R, McNevin D (2013) High resolution melting (HRM) of forensically informative SNPs. Forensic Science International: Genetics Supplement Series 4 (1):e376-e377.

Mehta B, Daniel R, McNevin D (2017) HRM and SNaPshot as alternative forensic SNP genotyping methods. Forensic Science, Medicine, and Pathology 13 (3):293–301.

1.9.3 Chapter 4: High-throughput genotyping tools: Illumina MiSeq Massively Parallel Sequencing

MPS, also referred to as NGS, can genotype multiple markers simultaneously for many samples [149]. Markers for hundreds of BGA and EVCs are known and genotyping them together may be beneficial in providing investigators with FDP leads quicker and more cost efficiently. Two MPS technologies have dominated the forensic sphere: Ion Torrent (TFS) and fluorescent-based SBS (Illumina). Chapter 4 focuses on the forensic evaluation of Illumina SBS chemistry using the MiSeq (Illumina) sequencer. The evaluation was performed on forensic criteria, including sensitivity, reproducibility, multiplexing capability, mixture detection, and the ability to type difficult samples—UV degraded and humic acid inhibited. The assessment further included forensic casework samples and genotype concordance with SNaPshot™ genotypes. One of the important features described in the chapter is the capability of MPS to genotype amplicons from existing SNaPshot™ PCR assays and resulting benefits to the forensic community. MPS is also capable of typing SNPs and STRs together. For example, the ForenSeq DNA Signature Prep kit (Verogen) is such an assay that can be typed on the MiSeq FGx (Verogen) [148]. Chapter 4 also describes the workflow forensic laboratories can utilise for MPS with automation of sample and library preparation steps, using the ForenSeq DNA Signature Kit (Verogen). It includes a comparison between different library quantitation methods required in the MPS sample processing workflow. Chapter 4 is presented as two published papers:

Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37 (21):2832–2840

Mehta B, Venables S, Roffey P (2018) Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. International Journal of Legal Medicine 132 (1):125–132.

1.9.4 Chapter 5

This chapter draws conclusions comparing and contrasting three FDP typing technologies. It also outlines potential future directions.

1.10 References

1. Kayser, M. and P. de Knijff, *Improving human forensics through advances in genetics, genomics and molecular biology.* Nature Reviews Genetics, 2011. **12**(3): p. 179-192.

2. Phillips, C., et al., *Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs.* Forensic Science International: Genetics, 2007. **1**(3–4): p. 273-280.

3. Sulem, P., et al., *Genetic determinants of hair, eye and skin pigmentation in Europeans.* Nature Genetics, 2007. **39**(12): p. 1443-1452.

4. Ho, Y.Y.W., et al., *Genetic variant influence on whorls in fingerprint patterns.* The Journal of investigative dermatology, 2016. **136**(4): p. 859-862.

5. Claes, P., H. Hill, and M.D. Shriver, *Toward DNA-based facial composites: Preliminary results and validation.* Forensic Science International: Genetics, 2014. **13**: p. 208-216.

6. Butler, J.M., M.D. Coble, and P.M. Vallone, *STRs vs. SNPs: thoughts on the future of forensic DNA testing.* Forensic Science, Medicine, and Pathology, 2007. **3**(3): p. 200-205.

7. Sobrino, B., M. Brion, and A. Carracedo, *SNPs in forensic genetics: a review on SNP typing methodologies.* Forensic Science International, 2005. **154**: p. 181-194.

8. Tilstone, W.J., K.A. Savaga, and L.A. Clark, *Forensic science : an encyclopedia of history, methods, and techniques*. 2006, Santa Barbara, Calif: ABC-CLIO. 307.

9. Landsteiner, K., *Zur Kenntnis der antifermentativen, lytischen und agglutinierenden Wirkungen des Blutserums und der.* Lymphe. Zentbl. Bakt. Orig., 1900. **27**: p. 357-362.

10. Cameron, J.M., *R. V Payne.* Medico-Legal Journal, 1972. **40**(1): p. 18-26.

11. TheTrueCrimeEnthusiast. *When life means life – The Dore Massacre*. 2016.

12. Watson, J.D. and F.H.C. Crick, *Molecular structure of Nucleic Acids.* Nature, 1953. **171**: p. 737-738.

13. NIJ, *The Future of Forensic DNA Testing : Predictions of the research and development working group*. 2000, National Institute of Justice.

14. Jeffreys, A.L., V. Wilson, and S.L. Thein, *Individual specific 'fingerprints' of human DNA.* Nature, 1985b. **316**: p. 75-79.

15. Aronson, J.D., *DNA fingerprinting on trial: the dramatic early history of a new forensic technique.* Endeavour, 2005. **29**(3): p. 126-131.

16. Wyman, A.R. and R.A. White, *Highly polymorphic locus in human DNA.* Proceedings of National Academic Science, 1980. **77**: p. 6754-6758.

17. Jeffreys, A.L., V. Wilson, and S.L. Thein, *Hypervariable 'minisatellite' regions in human DNA.* Nature, 1985a. **314**(67-72): p. 67.

18. Budowle, B., et al., *Hae III--a suitable restriction endonuclease for restriction fragment length polymorphism analysis of biological evidence samples.* Journal of Forensic Science, 1990. **35**(3): p. 530-536.

19. Southern, E., *Detection of specific sequences among DNA fragments separated by gel electrophoresis.* Journal of Molecular Biology 1975. **98**: p. 503-517.

20. Andersen J, et al., *Report on the third EDNAP collaborative STR exercise.* Forensic Science International, 1996. **78**: p. 83-93.

21. Smith, R.N., *Accurate size comparison of short tandem repeat alleles amplified by PCR.* BioTechniques, 1995. **18**(1): p. 122-128.

22. Mullis K, et al., *Specific enzymatic amplification of DNA in vitro: The polymerase chain reaction.* Cold Spring Harbor Symposia on Quantitative Biology, 1986. **51**: p. 263-273.

23. Saiki, R., et al., *Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia.* Science, 1985. **230**(4732): p. 1350-1354.

24. Butler, J., *Genetics and genomics of core short tandem repeat loci used in human identity testing.* Journal of Forensic Science, 2006. **51**(2): p. 253-265.

25. Hagelberg, E., I.C. Gray, and A.J. Jeffreys, *Identification of the skeletal remains of a murder victim by DNA analysis.* Nature, 1991. **352**: p. 427.

26. Jeffreys, A., et al., *Identification of the Skeletal Remains of Mengele, Josef by DNA Analysis*. Vol. 56. 1992. 65-76.

27. Kimpton, C., et al., *Evaluation of an automated DNA profiling system employing multiplex amplification of four tetrameric STR loci*. Vol. 106. 1994. 302-311.

28. Sparkes, R., et al., *The validation of a 7-locus multiplex STR test for use in forensic casework. (II), Artefacts, casework studies and success rates*. Vol. 109. 1996. 195-204.

29. Butler, J.M., *Chapter 5 - Short Tandem Repeat (STR) Loci and Kits*, in *Advanced Topics in Forensic DNA Typing: Methodology*. 2012, Academic Press: San Diego. p. 99-139.

30. AppliedBiosystems, *GlobalFiler® PCR Amplification Kit User Guide*. Life Technologies Corporation, Carlsbad, CA, 2012.

31. Collins, P., et al., *Developmental Validation of a Single-Tube Amplification of the 13 CODIS STR Loci, D2S1338, D19S433, and Amelogenin: The AmpFℓSTR® Identifiler® PCR Amplification Kit.* 2004.

32. Promega, *Technical manual: PowerPlex1 21 System*. Promega Corporation, Madison, WI, 2012.

33. Butler, J., *Forensic DNA Typing: Biology, Technology and genetics of STR markers*. 2005, London: Elsevier Academic Press, London.

34. Wilson, M.R., et al., *Validation of mitochondrial DNA sequencing for forensic casework analysis.* International Journal of Legal Medicine, 1995. **108**(2): p. 68-74.

35. Budowle, B., et al., *Forensics and Mitochondrial DNA: Applications, Debates, and Foundations.* Annual Review of Genomics and Human Genetics, 2003. **4**(1): p. 119-141.

36. Schneider, P.M., *Beyond STRs: The Role of Diallelic Markers in Forensic Genetics.* Transfusion Medicine and Hemotherapy, 2012. **39**(3): p. 176-180.

37. Brookes, A.J., *The essence of SNPs.* Gene, 1999. **234**: p. 177-186.

38. Vignal, A., et al., *A review on SNP and other types of molecular markers and their use in animal genetics.* Genetics Selection Evolution 2002. **34**(2002): p. 275-305.

39. Sherry ST, et al., *dbSNP: the NCBI database of genetic variation.* The Nucleic Acid Research, 2001. **29**(1): p. 308-311.

40. Nachman, M.W. and S.L. Crowell, *Estimate of the Mutation Rate per Nucleotide in Humans.* Genetics, 2000. **156**(1): p. 297-304.

41. Martinez-Arias R, et al., *Sequence variability of a human pseudogene.* Genome Research, 2001. **11**: p. 1071-1085.

42. Roach, J.C., et al., *Analysis of Genetic Inheritance in a Family Quartet by Whole-Genome Sequencing.* Science, 2010. **328**(5978): p. 636-639.

43. Cooper, D.N., et al., *An estimate of unique DNA sequence heterozygosity in the human genome.* Human Genetics, 1985. **69**(3): p. 201-205.

44. L Holden, A., *The SNP Consortium: Summary of a private consortium effort to develop an applied map of the human genome*. Vol. Suppl. 2002. 22-4, 26.

45. Li, M., M. Boehnke, and G.R. Abecasis, *Efficient Study Designs for Test of Genetic Association Using Sibship Data and Unrelated Cases and Controls.* American Journal of Human Genetics, 2006. **78**(5): p. 778-792.

46. Budowle, B. and A. vanDaal, *Extracting evidence from forenisc DNA analyses: Future molecular biology directions.* Biotechniques, 2009. **46**(5): p. 339-350.

47. Budowle, B. and A. vanDaal, *Forensically relevant SNP classes.* Biotechniques, 2008. **44**(5): p. 603-610.

48. Kidd, K.K., *Population Genetics of SNPs for forensic purposes (updated)*. 2011, National Institute of Justice.

49. Dixon, L.A., et al., *Validation of a 21-locus autosomal SNP multiplex for forensic identification purposes.* Forensic Science International, 2005. **154**(1): p. 62-77.

50. Sanchez, J.J., et al., *A multiplex assay with 52 single nucleotide polymorphisms for human identification.* Electrophoresis, 2006. **27**(9): p. 1713-1724.

51. Kidd, K.K., et al., *Developing a SNP panel for forensic identification of individuals.* Forensic Science International, 2006. **164**(1): p. 20-32.

52. Pakstis, A.J., et al., *Candidate SNPs for a universal individual identification panel.* Human Genetics, 2007. **121**(3): p. 305-317.

53. Pakstis, A.J., et al., *SNPs for a universal individual identification panel.* Human Genetics, 2010. **127**: p. 315-324.

54. Westen, A.A., et al., *Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples.* Forensic Science International: Genetics, 2009. **3**(4): p. 233-241.

55. Zha, L., et al., *Analysis of tri-allelic SNPs for forensic purpose in Chinese Han population.* Forensic Science International: Genetics Supplement Series. **3**(1): p. e107-e108.

56. Weber, J.L., et al., *Human Diallelic Insertion/Deletion Polymorphisms.* American Journal of Human Genetics, 2002. **71**(4): p. 854-862.

57. Mills, R.E., et al., *An initial map of insertion and deletion (INDEL) variation in the human genome.* Genome Research, 2006(16): p. 1182-1190.

58. Mullaney, J.M., et al., *Small insertions and deletions (INDELs) in human genomes.* Human Molecular Genetics, 2010. **19**(R2): p. R131-R136.

59. Pereira, R., et al., *A new multiplex for human identification using insertion/deletion polymorphisms.* ELECTROPHORESIS, 2009. **30**(21): p. 3682-3690.

60. Wei, Y.-L., et al., *A validation study of a multiplex INDEL assay for forensic use in four Chinese populations.* Forensic Science International: Genetics. **9**: p. e22-e25.

61. Pereira, R., et al., *Straightforward Inference of Ancestry and Admixture Proportions through Ancestry-Informative Insertion Deletion Multiplexing.* PLOS ONE, 2012. **7**(1): p. e29684.

62. Gabriel, S.B., et al., *The structure of haplotype blocks in human genome.* Science, 2002. **296**: p. 2225-2229.

63. Ge, J., et al., *Haplotype block: a new type of forensic DNA markers*. Vol. 124. 2010. 353-61.

64. Kidd, K.K., et al., *Current sequencing technology makes microhaplotypes a powerful new type of genetic marker for forensics.* Forensic Science International: Genetics. **12**: p. 215-224.

65. Kidd, K.K., et al., *Microhaplotype loci are a powerful new type of forensic marker.* Forensic Science International: Genetics Supplement Series. **4**(1): p. e123-e124.

66. Kidd, K.K., et al., *Evaluating 130 microhaplotypes across a global set of 83 populations.* Forensic Science International: Genetics, 2017. **29**: p. 29-37.

67. Kidd, K.K. and W.C. Speed, *Criteria for selecting microhaplotypes: mixture detection and deconvolution.* Investigative Genetics, 2015. **6**(1): p. 1.

68. Oldoni, F., et al., *Microhaplotypes for ancestry prediction.* Forensic Science International: Genetics Supplement Series, 2017. **6**: p. e513-e515.

69. Smithies, O., *Zone electrophoresis in starch gels: group variations in the serum proteins of normal human adults.* Biochemical Journal, 1955. **61**(4): p. 629-641.

70. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(12): p. 5463-5467.

71. Maxam, A.M. and W. Gilbert, *A new method for sequencing DNA.* Proceedings of the National Academy of Sciences of the United States of America, 1977. **74**(2): p. 560-564.

72. Karger, B.L. and A. Guttman, *DNA Sequencing by Capillary Electrophoresis.* Electrophoresis, 2009. **30**(Suppl 1): p. S196-S202.

73. Kayser, M. and P.d. Knijff, *Improving human forensics through advances in genetics, genomics and molecular biology.* Nature Reviews, 2011. **12**: p. 179-192.

74. M'charek, A., *Contrasts and Comparison: three practices of forensic investigation.* Comparative Sociology, 2008. **7**: p. 387-412.

75. Vogel, G., *German law allows use of DNA to predict suspects' looks*. Science, 2018. **360**(6391): p. 841-842.

76. Koops, B.-J. and M. Schellekens, *Forensic DNA phenotypong: regulatory issues.* The Columbia Science and Technology Law Review, 2008. **9**: p. 158-202.

77. Ossorio, P.N., *About face: forensic genetic testing for race and visible traits.* Journal of Law Medicine and Ethics, 2006. **34**: p. 277-292.

78. Shriver, M.D. and R.A. Kittles, *Genetic ancestry and the search for personalized genetic histories.* Nature Reviews, 2004. **5**: p. 611-618.

79. Frudakis, T.N., *Molecular photofitting: predicting ancestry and phenotype from DNA*. 2008, Amsterdam, Netherlands: Academic Press Publishers (Elsevier).

80. Rosenberg, N.A., et al., *Informativeness of genetic markers for inference of ancestry.* The American Journal of Human Genetics, 2003. **73**: p. 1402-1422.

81. Phillips, C., et al., *Development of a novel forensic STR multiplex for ancestry analysis and extended identity testing.* ELECTROPHORESIS, 2013. **34**(8): p. 1151-1162.

82. Kayser, M. and P.M. Schneider, *DNA-based prediction of externally visible characteristics in forensics: Motivations, scientific challlanges and ethical consideration.* Forensic Science International: Genetics, 2009. **3**: p. 154-161.

83. Giardina, E., A. Spinella, and G. Novelli, *Past, present and future of forensic DNA typing.* Nanomedicine, 2011. **6**(2): p. 257-270.

84. Beckmann, J.S., K. Estivill, and S.E. Antonarakis, *Copy number varaints and genetic traits: closer to the resolution of phenotypic to genotypic varaibility.* Nature Reviews: Genetics, 2007. **6**: p. 639-646.

85. Saiki, R.K., et al., *Genetic analysis of amplified DNA with immobilized sequence-specific oligonucleotide probes.* proceedings of National Academic Science USA, 1989. **86**(16): p. 6230-6234.

86. Lareu, M., et al., *The use of the LightCycler for the detection of Ychromosome SNPs.* Forensic Science International, 2001. **118**(2-3): p. 163-168.

87. Livak, K.J., *Allelicdiscrimination using fluorogenicprobes and the 5' nuclease assay.* Genetic Analysis: Biomolecular Engineering, 1999. **14**(5-6): p. 143-149.

88. Shen, G.-Q., K.G. Abdullah, and Q.K. Wang, *The TaqMan method for SNP genotyping*, in *Single nucleotide polymorphims:Methods and protocols*, A.A. Komar, Editor. 2009, Humana Press. p. 293-306.

89. Tyagi, S., D.P. Bratu, and F.R. Kramer, *Multicolor molecular beacons for allele discrimination.* Nature Biotechnology, 1998. **16**(1): p. 49-53.

90. Nazarenko, I.A., S.K. Bhatnagar, and R.J. Hohman, *A closed tube format for amplification and detection of DNA based on energy transfer.* Nucleic Acid Research, 1997. **25**(12): p. 2516-2521.

91. Mei, R., et al., *Genome-wide Detection of Allelic Imbalance Using Human SNPs and High-density DNA Arrays.* Genome Research, 2000. **10**: p. 1126-1137.

92. Haff, L.A. and I.P. Smirnov, *Multiplex genotyping of PCR products with MassTag-labeled primers.* Nucleic Acid Research, 1997. **25**(18): p. 3749-3750.

93. Pastinen, T., et al., *A System for Specific, High-throughput Genotyping by Allele-specific Primer Extension on Microarrays.* Genome Research, 2000. **10**(7): p. 1031-1042.

94. Ding, C. and S. Jin, *High throughput methods for SNP genotyping*, in *Single nuceotide polymorphisms: methods and protocols*, A.A. Komar, Editor. 2009, Humana Press. p. 245-276.

95. Delahunty, C., et al., *Testing the feasibility of DNA typing for human identification by PCR and an oligonucleotide ligation assay.* American Journal of Human Genetics, 1996. **58**(6): p. 1239-146.

96. Lyamichev, V., et al., *Polymorphism identification and quantitative detection of genomic DNA by invasive cleavage of oligonucleotide probes.* Nature Biotechnology, 1999. **17**(3): p. 292-296.

97. Rothberg, J.M., et al., *An integrated semiconductor device enabling non-optical genome sequencing.* Nature, 2011. **475**: p. 348.

98. Quail, M.A., et al., *A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers.* BMC genomics, 2012. **13**(1): p. 341.

99. Reed, G.H., J.O. Kent, and C.T. Wittwer, *High resolution DNA melting analysis for simple and efficient molecular diagnostics.* Pahrmacogenomics, 2007. **8**(6): p. 597-608.

100. Wallace, R.B., et al., *Hybridisation of synthetic oligodeoxyribonucleotides to phi 174 DNA: the effect of single base pair mismatch.* Nucleic Acid Research, 1979. **6**: p. 3543-3557.

101. Clegg, R., *Clegg, R.M. Fluorescence resonance energy transfer and nucleic acids. Methods Enzymol. 211, 353-388.* Vol. 211. 1992. 353-88.

102. Fluidigm, *Fuidigm system for rapid, reliable and cost effective SNP genotyping*. 2012, Fluidigm.

103. Kostrikis, L.G., et al., *Spectral Genotyping of Human Alleles.* Science, 1998. **279**(5354): p. 1228-1229.

104. International, C., *Amplifluor SNPs genotyping system for assay development*, in *Chemicon International*. 2010. p. 1-45.

105. Warden, D.R. and H. Refsum, *Detection of Single-Nucleotide Polymorphisms by PCR with Universal Energy Transfer–Labeled Primers: Application to Folate- and Cobalamin-Related Genes.* Clinical Chemistry, 2005. **51**(9): p. 1713-1716.

106. Okayama, H., et al., *Rapid, nonradioactive detection of mutations in the human genome by allele-specific amplification.* Journal of Laboratory and Clinical Medicine, 1989. **114**(2): p. 105-113.

107. Newton, C.R., et al., *Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS).* Nucleic Acids Research, 1989. **17**(7): p. 2503-2516.

108. Syvänen, A.-C., et al., *A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E.* Genomics, 1990. **8**(4): p. 684-692.

109. Sokolov, B.P., *Primer extension technique for the detection of single nucleotide in genomic DNA.* Nucleic acids research, 1990. **18**(12): p. 3671.

110. Quintáns, B., et al., *Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing.* Forensic science international, 2004. **140**(2): p. 251-257.

111. Braun, A., D.P. Little, and H. Köster, *Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry.* Clinical chemistry, 1997. **43**(7): p. 1151-1158.

112. Hirschhorn, J.N., et al., *SBE-TAGS: An array-based method for efficient single-nucleotide polymorphism genotyping.* Proceedings of the National Academy of Sciences of the United States of America, 2000. **97**(22): p. 12164-12169.

113. Myakishev, M.V., et al., *High-Throughput SNP Genotyping by Allele-Specific PCR with Universal Energy-Transfer-Labeled Primers.* Genome Research, 2001. **11**(1): p. 163-169.

114. Landergren, U., et al., *A ligase-mediated gene detection technique.* Science, 1988. **241**: p. 1077-1080.

115. AppliedBiosystems, *A guide to High Resolution Melting (HRM) analysis*. 2012, Applied Biosystems. p. 1-20.

116. Sanger, F. and A.R. Coulson, *A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.* Journal of Molecular Biology, 1975. **94**(3): p. 441-448.

117. Luckey, J.A., et al., *High speed DNA sequencing by capillary electrophoresis.* Nucleic Acids Research, 1990. **18**(15): p. 4417-4421.

118. Smith, L.M., et al., *The synthesis of oligonucleotides containing an aliphatic amino group at the 5' terminus: synthesis of fluorescent DNA primers for use in DNA sequence analysis.* Nucleic Acids Research, 1985. **13**(7): p. 2399-2412.

119. Smith, L.M., et al., *Fluorescence detection in automated DNA sequence analysis.* Nature, 1986. **321**: p. 674.

120. Cousins, M.M., et al., *Comparison of a High-Resolution Melting Assay to Next-Generation Sequencing for Analysis of HIV Diversity.* Journal of Clinical Microbiology, 2012. **50**(9): p. 3054-3059.

121. Ambardar, S., et al., *High Throughput Sequencing: An Overview of Sequencing Chemistry.* Indian journal of microbiology, 2016. **56**(4): p. 394-404.

122. Knief, C., *Analysis of plant microb interactions in the eaa of next generation sequencing technologies*. Vol. 5. 2014. 216.

123. Hyman, E.D., *A new method of sequencing DNA.* Analytical Biochemistry, 1988. **174**(2): p. 423-436.

124. Ronaghi, M., M. Uhlén, and P. Nyrén, *A Sequencing Method Based on Real-Time Pyrophosphate.* Science, 1998. **281**(5375): p. 363-365.

125. Ronaghi, M., et al., *Real-Time DNA Sequencing Using Detection of Pyrophosphate Release.* Analytical Biochemistry, 1996. **242**(1): p. 84-89.

126. Margulies, M., et al., *Genome sequencing in microfabricated high-density picolitre reactors.* Nature, 2005. **437**: p. 376.

127. AllSeq. *454 (Roche)*. 2017.

128. Voelkerding, K.V., S.A. Dames, and J.D. Durtschi, *Next-Generation Sequencing: From Basic Research to Diagnostics.* Clinical Chemistry, 2009. **55**(4): p. 641-658.

129. Turcatti, G., et al., *A new class of cleavable fluorescent nucleotides: synthesis and optimization as reversible terminators for DNA sequencing by synthesis.* Nucleic Acids Research, 2008. **36**(4): p. e25-e25.

130. Balasubramanian, S., *Sequencing nucleic acids: from chemistry to medicine.* Chemical Communications, 2011. **47**(26): p. 7281-7286.

131. Shendure, J., et al., *Accurate Multiplex Polony Sequencing of an Evolved Bacterial Genome.* Science, 2005. **309**(5741): p. 1728-1732.

132. GLENN, T.C., *Field guide to next- generation DNA sequencers.* Molecular Ecology Resources, 2011. **11**(5): p. 759-769.

133. Perkel, J.M. and P.A. Fung, *2016 Next-Gen Sequencing Buyer's Guide.* 2016.

134. Harris, T.D., et al., *Single-Molecule DNA Sequencing of a Viral Genome.* Science, 2008. **320**(5872): p. 106-109.

135. Bowers, J., et al., *Virtual Terminator nucleotides for next generation DNA sequencing.* Nature methods, 2009. **6**(8): p. 593-595.

136. Schadt, E.E., S. Turner, and A. Kasarskis, *A window into third-generation sequencing.* Human Molecular Genetics, 2010. **19**(R2): p. R227-R240.

137. GenomeWeb. *Helicos BioSciences Files for Chapter 11 Bankruptcy Protection*. 2012; Available from: https://www.genomeweb.com/sequencing/helicos-biosciences-files-chapter-11-bankruptcy-protection.

138. Eid, J., et al., *Real-Time DNA Sequencing from Single Polymerase Molecules.* Science, 2009. **323**(5910): p. 133-138.

139. Flusberg, B.A., et al., *Direct detection of DNA methylation during single-molecule, real-time sequencing.* Nature methods, 2010. **7**(6): p. 461-465.

140. van Dijk, E.L., et al., *Ten years of next-generation sequencing technology.* Trends in Genetics. **30**(9): p. 418-426.

141. Clarke, J., et al., *Continuous base identification for single-molecule nanopore DNA sequencing.* Nature Nanotechnology, 2009. **4**: p. 265.

142. Eisenstein, M., *Oxford Nanopore announcement sets sequencing sector abuzz.* Nature Biotechnology, 2012. **30**: p. 295.

143. Branton, D., et al., *The potential and challenges of nanopore sequencing.* Nature Biotechnology, 2008. **26**: p. 1146.

144. Haque, F., et al., *Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA.* Nano Today, 2013. **8**(1): p. 56-74.

145. Yang, Y., B. Xie, and J. Yan, *Application of Next-generation Sequencing Technology in Forensic Science.* Genomics, Proteomics & Bioinformatics, 2014. **12**(5): p. 190-197.

146. Sobiah R, et al., *Implications of Targeted Next Generation Sequencing in Forensic Science.* Journal of Forensic Research, 2018. **9**(2): p. 416.

147. Daniel, R., et al., *A SNaPshot of next generation sequencing for forensic SNP analysis.* Forensic Science International: Genetics, 2015. **14**(0): p. 50-60.

148. Churchill, J.D., et al., *Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for Use in Genetic Profiling.* Forensic Science International: Genetics, 2016. **20**: p. 20-29.

149. Mehta, B., et al., *Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq.* Electrophoresis, 2016. **37**(21): p. 2832-2840.

150. Vidaki, A., et al., *DNA methylation-based forensic age prediction using artificial neural networks and next generation sequencing.* Forensic science international. Genetics, 2017. **28**: p. 225-236.

151. Binladen, J., et al., *The Use of Coded PCR Primers Enables High-Throughput Sequencing of Multiple Homolog Amplification Products by 454 Parallel Sequencing.* PLOS ONE, 2007. **2**(2): p. e197.

152. ThermoFisherScientific, *Ion AmpliSeq™ Library Kit 2.0 User Guide*. 2019.

153. Mehta, B., S. Venables, and P. Roffey, *Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping.* INTERNATIONAL JOURNAL OF LEGAL MEDICINE, 2018. **132**(1): p. 125-132.

154. Floragenex, *Technical Brief: Ion Torrent™ NGS and AmpliSeq™ Genotyping*, in *www.floragenex.com*. 2015.

155. Mogensen, H.S., C. Børsting, and N. Morling, *Template preparation of AmpliSeqTM libraries using the Ion ChefTM.* Forensic Science International: Genetics Supplement Series, 2015. **5**: p. e368-e369.

156. Rusk, N., *Torrents of sequence.* Nature Methods, 2010. **8**: p. 44.

157. Pennisi, E., *Semiconductors Inspire New Sequencing Technologies.* Science, 2010. **327**(5970): p. 1190.

158. Merriman, B., I.T. D Team, and J.M. Rothberg, *Progress in Ion Torrent semiconductor chip based sequencing.* ELECTROPHORESIS, 2012. **33**(23): p. 3397-3417.

159. Al-Asfi, M., et al., *Assessment of the Precision ID Ancestry panel.* International journal of legal medicine, 2018.

160. ThermoFisherScientific, *Integrate NGS into your forensic DNA lab workflow*. 2017.

161. ThermoFisherScientific. *Ion AmpliSeq Designer*. 2019.

162. Jäger, A.C., et al., *Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories.* Forensic Science International: Genetics, 2017. **28**: p. 52-70.

163. Illumina, *TruSeq Forensic Amplicon : Preparing Samples for the Illumina Sequencing platform*. 2014.

164. Illumina *MiSeq System User Guide*. 2018. **Document # 1000000061014 v00**.

165. Illumina, *ForenSeq™ DNA Signature Prep*. 2015.

166. Verogen, *MiSeq FGx™ Instrument : Reference Guide*. 2018.

167. Leray, M., Q. Haenel, and S. Bourlat, *Preparation of Amplicon Libraries for Metabarcoding of Marine Eukaryotes Using Illumina MiSeq: The Adapter Ligation Method*. 2016. p. 209-218.

168. Illumina, *ForenSeq™ DNA Signature Prep Kit* 2016.

169. Illumina. *Illumina Sequenicng Technology*. 2010 05.08.15].

170. Jeon, Y.-S., et al., *Improved pipeline for reducing erroneous identification by 16S rRNA sequences using the Illumina MiSeq platform.* Journal of Microbiology, 2015. **53**(1): p. 60-69.

171. Technologies, I.D., *Unique, dual-matched adapters mitigate index hopping between NGS samples*. 2017.

172. Illumina, *MiSeq® Reporter Software Guide*. 2017. p. 25-26.

173. Illumina. *BaseSpace Sequence Hub*. 2018.

174. Silvia, A.L., N. Shugarts, and J. Smith, *A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex.* International Journal of Legal Medicine, 2016: p. 1-14.

175. Jäger, A.C., et al., *Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories.* Forensic Science International: Genetics, 2017. **28**: p. 52-70.

176. Verogen, *ForenSeq Universal Analysis Software*

2018: https://verogen.com/wp-content/uploads/2018/08/ForenSeq-analysis-software-data-sheet-VD2018004.pdf.

177. Illumina. *Introduction to mtDNA Analysis*. 2018; Available from: https://sapac.illumina.com/areas-of-interest/forensic-genomics/forensic-analysis-methods/mtdna-analysis.html.

178. Faith, S.A. and M. Scheible, *Analyzing Data from Next Generation Sequencers Using the PowerSeq® Auto/Mito/Y System.* PromegaCoorporation, 2016.

179. Qiagen. *GeneReader Platform*. 2018; Available from: https://www.qiagen.com/cn/products/ngs/mdx-ngs-genereader/library-preparation/qiagen-genereader-platform/#productdetails.

180. Zbieć-Piekarska, R., et al., *Development of a forensically useful age prediction method based on DNA methylation analysis.* Forensic Science International: Genetics, 2015. **17**: p. 173-179.

181. Tindall, E.A., et al., *Assessing high-resolution melt curve analysis for accurate detection of gene variants in complex DNA fragments.* Human Mutation, 2009. **30**(6): p. 876-883.

182. Twist, G.P., et al., *High-resolution melt analysis to detect sequence variations in highly homologous gene regions: application to CYP2B6.* Pharmacogenomics, 2013. **14**(8): p. 10.2217/pgs.13.66.

183. AppliedBiosystems, *ABI PRISM® SNaPshot™ Multiplex Kit*. Thermo Fisher Scientific. p. 1-42.

Chapter 2: Forensically relevant SNaPshot™ assays for forensic SNP genotyping

2.1 Mehta B, Daniel R, Phillips C, McNevin D (2017) Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. International Journal of Legal Medicine 131(1): 21–37

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 2

Declaration by candidate

In the case of Chapter 2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Proposal and writing of the review; writing the first draft of the manuscript	65

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student UC Y/N
Dennis McNevin	Revising and editing of the manuscript, critical feedback	15	N
Runa Daniel	Revising and editing of the manuscript, critical feedback	15	N
Chris Phillips	Critical feedback	5	N

Candidate's Signature I.S.M McNEFF **Date 04/04/2018**

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who

accepts overall responsibility for the publication;

4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	**Name of the co-author**	**Signature**
04/04/2018	Dennis McNevin	Dennis McN
5/4/18	RUNA DANIEL	Runa Daniel

[* Please insert additional rows as needed.]

accepts overall responsibility for the publication;

4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	National Centre for Forensic Studies, Faculty of STeM, University of Canberra

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	Dennis McN
06/04/2018	Chris Phillips	

[* Please insert additional rows as needed.]

[Mehta, B., Daniel, R., Phillips, C. et al. Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Int J Legal Med 131, 21–37 (2017). https://doi.org/10.1007/s00414-016-1490-5]

Int J Legal Med (2017) 131:21–37
DOI 10.1007/s00414-016-1490-5

REVIEW

Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review

Bhavik Mehta[1] · **Runa Daniel**[2] · **Chris Phillips**[3] · **Dennis McNevin**[1]

Received: 11 August 2016 / Accepted: 31 October 2016 / Published online: 14 November 2016

Abstract Short tandem repeats are the gold standard for human identification but are not informative for forensic DNA phenotyping (FDP). Single-nucleotide polymorphisms (SNPs) as genetic markers can be applied to both identification and FDP. The concept of DNA intelligence emerged with the potential for SNPs to infer biogeographical ancestry (BGA) and externally visible characteristics (EVCs), which together enable the FDP process. For more than a decade, the SNaPshot® technique has been utilised to analyse identity and FDP-associated SNPs in forensic DNA analysis. SNaPshot is a single-base extension (SBE) assay with capillary electrophoresis as its detection system. This multiplexing technique offers the advantage of easy integration into operational forensic laboratories without the requirement for any additional equipment. Further, the SNP panels from SNaPshot® assays can be incorporated into customised panels for massively parallel sequencing (MPS). Many SNaPshot® assays are available for identity, BGA and EVC profiling with examples including the well-known SNP*for*ID 52-plex identity assay, the SNP*for*ID 34-plex BGA assay and the HIrisPlex EVC assay. This review lists the major forensically relevant SNaPshot® assays for human DNA SNP analysis and can be used as a guide for selecting the appropriate assay for specific identity and FDP applications.

Keywords Single-nucleotide polymorphism (SNP) · SNaPshot · Forensic genotyping · Capillary electrophoresis (CE) · Forensic DNA phenotyping (FDP) · DNA intelligence

✉ Bhavik Mehta
bhavik.mehta@hotmail.com

[1] National Centre for Forensic Studies, Faculty of Education, Science, Technology and Mathematics (ESTeM), University of Canberra, Bruce, ACT 2617, Australia

[2] Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Macleod, VIC 3079, Australia

[3] Forensic Genetics Unit, Institute of Forensic Sciences, University of Santiago de Compostela, 15782 Santiago de Compostela, Spain

Introduction

Short tandem repeats (STRs) are the markers of choice for forensic human identification due to their highly polymorphic nature and, therefore, their ability to differentiate between individuals [1]. In the past decade, genome-wide association studies (GWASs) have flooded databases with novel single-nucleotide polymorphisms (SNPs) [2]. In addition, numerous studies have contributed population data associated with SNPs to these databases, creating a valuable scientific resource [3]. The forensic community has been utilising these resources to apply SNPs to forensic DNA analysis, for both human identification and intelligence. SNPs can offer some key advantages over STRs including lower mutation rates (ideal for ancestry affiliation), higher abundance in the human genome, short PCR amplicon length suitable for high multiplexing capability and the analysis of degraded DNA, amenity to high-throughput genotyping and application to many forensic applications outside of human identification [4–6]. SNPs can be classified according to their forensic application such as identity-informative SNPs (IISNPs) for human identification, lineage-informative SNPs (LISNPs) for inferring genealogies (especially useful in kinship analysis and paternity testing), ancestry-informative SNPs (AISNPs) for inferring biogeographical ancestry (BGA) and phenotypic-informative SNPs (PISNPs) for inferring externally visible characteristics (EVCs) (such as eye, hair and skin

colours). IISNPs are less informative than the equivalent number of STRs because many of them are bi-allelic, and hence, 50–60 IISNPs are required to provide approximately the same discriminating power as 13 STRs [7]. In addition, bi-allelic markers have limited mixture resolution capability. Thus, while STRs are considered the gold standard for identification purposes, IISNPs can be used to supplement STR profiling and other LISNPs, AISNPs and PISNPs can provide valuable forensic intelligence by inferring lineage, BGA and EVCs, respectively, in cases when no STR inclusions are obtained and/or STR profiles are partial and non-informative [8].

The first PCR-based genotyping system interrogated SNPs at the HLA-DQA1 locus [9, 10]. The AmpliType® PM and DQA1 PCR amplification-reverse blot DNA typing system (Applied Biosystems, formerly PerkinElmer) was very popular in forensic laboratories nearly two decades ago [11, 12]. The system consisted of six loci and was developed as an alternative to the use of restriction fragment length polymorphisms (RFLPs) employing variable number tandem repeats (VNTRs). It was useful in casework applications, particularly when the evidentiary samples yielded low amounts of DNA or degraded DNA which could not be utilised for RFLP profiling [11, 13, 14]. However, it had a lower discrimination power than the RFLP method (~1:2000) due to the limited number of alleles available, was limited in its application to mixtures and was discontinued [12, 14]. VNTRs were later replaced by STRs [15] where DNA was amplified by PCR using commercial amplification kits and detected by capillary electrophoresis (CE) or 'genetic analysers'.

A variety of SNP genotyping techniques are available such as high resolution melting (HRM) analysis [16–18], TaqMan™ hybridisation probes (Applied Biosystems), invader technology [19], hybridisation microarrays [20], massively parallel sequencing (MPS) [21] and the SNaPshot® (Applied Biosystems) minisequencing method. Sobrino et al. [19] provided a comprehensive review on SNP genotyping methodologies [19]. Of these, SNaPshot®, based on minisequencing, has been most commonly applied to forensic DNA analysis due to its sensitivity and high multiplexing capability with the added advantage of not requiring additional equipment to that already utilised in forensic laboratories [22].

Background to the minisequencing method

Minisequencing is a genotyping method that falls under the broad category of primer extension techniques [19] which also includes other methods such as arrayed primer extension [23], primer oligo base extension [24] and pin-point assay [25]. In minisequencing, a detection primer is designed to anneal to the target DNA adjacent to the SNP of interest and is extended by a DNA polymerase using fluorescently labelled single nucleotides [19, 26]. The primer extension technique for detecting single nucleotides was developed in 1990 and was used mainly for diagnosis of genetic disorders and genotyping proteins [27, 28]. Earlier, singleplex minisequencing assays were performed using detection methods including gel and ELISA formats [26]. Later, multiplex assays that could simultaneously detect many sequence polymorphisms were developed [29]. The availability of enhanced detection methods including electrophoresis and fluorescence detection, matrix-assisted laser desorption/ionisation time-of-flight (MALDI-TOF) mass spectrometry (MS) and microarrays [19] enabled multiplex SNP genotyping.

The development of a multiplex solid-phase fluorescent minisequencing assay for the detection of 12 mitochondrial DNA sequence polymorphisms showed the potential of this method for forensic applications [29]. This assay was subsequently validated for forensic casework [30]. The minisequencing assay was then applied to genotype markers associated with phenotypes. Grimes et al. [31] developed a multiplex minisequencing assay detecting 12 mutations in the human melanocortin 1 receptor (*MC1R*) gene of which eight had an association with the red hair colour phenotype. The availability of the robust and accurate SNaPshot® (Applied Biosystems) kit [32], which involves electrophoresis and a fluorescence detection method, has led to the development of a series of forensically relevant multiplex assays.

Principle and workflow of SNaPshot® assays

The SNaPshot® assay is the most common commercial minisequencing method that works on the principle of single-base extension (SBE). Firstly, the DNA template is subjected to multiplex PCR to generate the target amplicons containing the SNPs of interest. Purification of the PCR product is then performed by adding exonuclease I and shrimp alkaline phosphatase (SAP) to degrade unbound primers and unincorporated dNTPs, which would interfere with the subsequent SBE reaction. The 3′ end of the oligonucleotide SNaPshot® (detection) primer binds immediately adjacent to the SNP of interest and is extended by *Taq* DNA polymerase which incorporates a fluorescently labelled dideoxyribo nucleotide triphosphate (ddNTP) complementary to the base on the opposite strand at the SNP position. Each detection primer can be augmented with (non-binding) oligonucleotide tails at the 5′ end (for example, poly C or CT) that assist in the spatial separation of the SBE products when detected by CE [22, 33]. The manufacturer's protocol recommends a maximum 10-plex; however, highly multiplexed custom assays often incorporating in excess of 20 SNP targets have been developed [32, 34–37]. The SBE products are further purified by adding SAP to eliminate unincorporated ddNTPs that potentially interfere with fluorescence detection. The purified products are prepared for CE and spatially separated on genetic analysers [32, 33]. Software such as GeneMapper™

ID-X (Applied Biosystems) is used to analyse the data and genotype the samples. Figure 1 illustrates the principle and workflow of the SNaPshot® method.

Primer design for multiplex assays

Primer design is critical for the successful development of multiplex SNaPshot® assays. The PCR and SBE primers can be designed using freely available primer designing tools such as Primer 3 [38] for primer sequences, AutoDimer [39] and IDT OligoAnalyzer [40] for secondary structure analysis (including the formation of primer dimers) and NCBI primer blast [41] for specificity. It is important that all primers in the same multiplex have similar melting temperatures (±3 °C) as they will all be subject to the same PCR cycling conditions [42]. In the same way, GC content of all the primers in the multiplex should be in the range 40–70% [35]. The tails added to the 5′ end of the SBE primers should not bind to any region of the genome in order to avoid non-specific binding in the SBE assay [33].

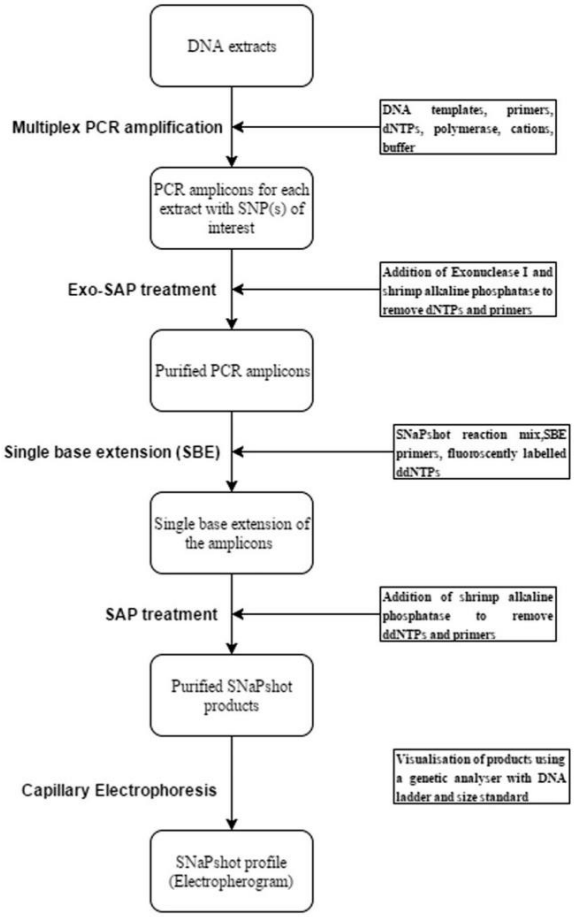


Fig. 1 SNaPshot® genotyping workflow including target amplification, enzymatic clean-up, single-base extension (*SBE*), second clean-up and electrophoresis

SNaPshot® multiplex optimisation

Each SNaPshot® multiplex assay requires optimisation in this order: multiplex PCR, multiplex SBE reaction and SBE product mobility. Generally, HPLC grade purified PCR and SBE primers are recommended for SNaPshot® assays to remove artefacts [43]. Prior to multiplex PCR optimisation, singleplex PCR is carried out to assess the performance of PCR primers and optimise assay conditions such as concentrations of primers, $MgCl_2$, dNTPs, *Taq* polymerase and thermal cycling conditions. The singleplex products are run on an agarose gel or microfluidic capillary electrophoresis (e.g. 2100 Bioanalyzer; Agilent Technologies) to verify the amplicon size distribution. The primers that yield low or no PCR product are redesigned at this stage. The multiplex PCR is then optimised, ensuring that all PCR products are more or less equally amplified [43, 44]. Exonuclease–shrimp alkaline phosphatase (ExoSAP–IT®; Affymetrix) purification of the optimised multiplex PCR products involves the removal of unbound primers and nucleotides that may hinder the subsequent analysis steps. After clean-up, the enzyme is inactivated by heating the product mixture to 80 °C for 15 min [35, 42, 45].

Singleplex SBE primer reactions without PCR template are performed to check for SBE primer self-extension. SBE primers should be redesigned if self-extension occurs. Multiplex PCR product is then added to singleplex SBE reactions to assess their specificity. The absence of one or more peaks requires redesign of the SBE primer. In order to proceed to multiplex SBE optimisation, the singleplex SBE products should show peaks higher than 500 relative florescence units (RFU) using the purified multiplex PCR amplicons [43]. During multiplex SBE optimisation, the concentration of SBE primers is adjusted depending on the signal intensities of the peaks. The sensitivity of the assay can be optimised by adjusting the volume of the SNaPshot® reaction mix (relative to DNA template amount), adjusting the number of PCR cycles and/or adjusting the number of SBE amplification cycles. Artefacts in the electropherogram may be due to factors such as poor quality of SBE primers, SBE primer interactions with other SBE primers, a non-specific PCR primer or very short PCR products. These artefacts can be neglected if they do not hinder accurate allele calling; otherwise, each SBE primer is removed serially from the multiplex to diagnose the cause of the artefact. A multiplex SBE reaction without PCR template can be carried out to identify any issues regarding the SBE primer interactions [33, 42].

It is also necessary to optimise the mobility of SBE primers to prevent SBE product overlap during CE. This optimisation

generally involves the redesign of the SBE primers with different length non-binding tails. It may also require redesign of both PCR and SBE primers and repetition of the entire process if the multiplex optimisation is unsuccessful [33].

CE and analysis

Purified SBE products are electrophoresed to obtain an electropherogram which is used to genotype samples. CE parameters including the dye set, type of polymer, capillary array length and run protocol are dependent on the type of platform used [33, 42]. For example, the 3500xl (Applied Biosystems) genetic analyser with POP-4 polymer and 36-cm array uses SNaPshot_pop4 default run protocol, GeneScan™ 120 LIZ™ Size Standard (Applied Biosystems) and Matrix Standard Set DS-02 (Applied Biosystems) for dye set E5. The electropherograms are analysed using fragment analysis software such as GeneMapper™ ID-X (Applied Biosystems) or GeneMarker™ (SoftGenetics®) [33]. The horizontal axis of the electropherogram represents the product length in the number of base pairs (bp), and the vertical axis represents the fluorescence signal intensity in RFUs. The product lengths derived from the horizontal axis will not correlate exactly with the SBE product lengths because the incorporated fluorophores will affect their electrophoretic mobilities. Each peak in the electropherogram represents the allele of a SNP with each of the four possible bases (A, C, G and T) labelled with a different fluorophore (ddATP-dR6G, ddCTP-dTAMRA™, ddGTP-dR110 and ddUTP-dROX™, respectively). Within the analysis software, bins and panels are used for genotype calling, and for each assay, different bins and panels are designed. Panels are determined based on the SNPs included in each assay while bins mark the bp range within which each peak falls. When a peak falls in a particular bin, the software makes a corresponding allele call. A single peak is obtained for homozygote genotypes, and two peaks (of different colours) are seen for heterozygous genotypes [22].

Forensically relevant SNaPshot® assays

Forensic SNaPshot® assays are here broadly classified into four categories based on their application such as IISNP assays, LISNP assays, AISNP assays and PISNP assays.

IISNP SNaPshot® assays

Forensic identity markers require the following characteristics: (i) minimal deviation from Hardy–Weinberg equilibrium (HWE), ensuring within-locus independence of alleles; (ii) minimal deviation from linkage equilibrium (LE), ensuring between locus independence; (iii) high heterozygosities to maximise the polymorphic nature of the loci, resulting in higher discrimination between genotypes; and (iv) these properties should apply both within and between multiple sub-populations [46]. One indication of this is that there is little genetic distance between these sub-populations, as measured by Wright's F_{ST}, the inbreeding coefficient within sub-populations relative to the total population [47]. IISNPs with these properties can be useful for providing identity information from degraded DNA or low DNA template amounts when STR profiling is not informative. Table 1 shows some of the potential IISNP assays.

Blood grouping assays

One of the earliest SNaPshot® assays differentiated between ABO blood groups. The multiplex assay was designed using six SNPs from the ABO gene. The assay correctly identified ABO genotypes when tested on casework samples involving bones, teeth and nails. ABO genotypes from semen-contaminated vaginal fluid casework samples were completely concordant with ABO phenotypes. This assay was species specific for human and higher-order primates. The assay was sensitive enough to generate ABO profiles from 0.1 ng of DNA [48].

Subsequently, a 17-plex SBE assay was developed to type all 10 of the known blood group systems. These SNPs represent the antigen and amino acid changes associated with each blood group. The method was tested on 29 samples with blood groups previously determined from hemagglutination, and it accurately predicted all the blood groups for all the samples. The assay was designed in three multiplexes (Multiplex I–Duffy and Dombrock blood group systems; Multiplex II–Landsteiner-Wiener, Colton, Scianna, Diego, Kidd, Lutheran and MN blood group systems; and Multiplex III–Kell and SS grouping systems) using genomic DNA extracted from 200 μl blood, and the turnaround time was 12 h for 32 samples [49].

Other (non-blood group) human identification SNP assays

One of the earliest human identification assays was a 24-plex SNaPshot® assay developed for the Korean population. The probability of identity for this assay was 2×10^{-10}, but the power of exclusion was 98.9% which is lower than that for Profiler Plus® (99.9%) [50]. The SNP*for*ID 52-plex assay, incorporating a set of two multiplexes (a 23-plex and a 29-plex) [35], provided a combined power of exclusion greater than 99.999% and a mean random match probability value in the order of 5×10^{-19}. The assay was forensically validated as a modified 49-plex assay [51] to be used in ISO 17025-accredited laboratories [52]. The 52-plex assay has also shown its versatility when applied in paternity testing [53]. It could be adapted by forensic casework laboratories as a supplementary method along with STR profiling and is useful for trace DNA amounts [54]. The 52-plex assay was tested in an Italian

Table 1 Identity-informative single-nucleotide polymorphism (IISNP) SNaPshot® assays

Assay	Target	Number of SNPs	Sensitivity (ng DNA)	Application	References
Blood grouping assays					
6-plex	Autosomal SNPs	6	0.1	ABO blood group genotyping	Doi et al. [48]
17-plex	Autosomal SNPs	17	–	Genotyping of 10 blood group systems	Palacajornsuk et al. [49]
Non-blood grouping assays					
SNP*for*ID 52-plex	Autosomal SNPs	52	0.5–70	Human identification	Sanchez et al. [35]
24-plex	Autosomal SNPs	24	1–2	Korean population identification	Lee et al. [50]
16-plex	X-SNPs	16	0.05–10	Identification of degraded samples	Oki et al. [60]
18-plex	Histone SNPs	18	0.078	Identification of highly degraded samples	Freire-Aradas et al. [61]
55-plex	Autosomal SNPs	55	0.125	Human identification	Wang et al. [22]

population, and the power of discrimination obtained was >99.99% with a 99.98% power of exclusion [55]. A sensitised version of SNP*for*ID 52-plex with increased SBE cycles (100 from 30) and use of the AmpFlSTR® SEfiler Plus™ Master Mix enhanced typing success from degraded and challenging DNA samples [56]. More recently, a 55-plex IISNP assay has been developed based on the Kidd 92 IISNP panel, enhanced from their earlier 44-plex assay [22, 57]. This assay was effective at 125 pg DNA with a power of exclusion >99.99% for the Hebei Han population [22].

IISNP SNaPshot® assays have also been developed for tri-allelic SNPs, which improve the resolution of degraded and mixed DNA samples. Westen et al. [58] developed a 16-plex SNP assay using Dutch sample sets. The assay was designed in three multiplexes of seven, four and five markers, and 15 SNPs were confirmed to be tri-allelic. When SGM Plus® and SNaPshot® profiles were compared, 14% of the alleles were lost with 5 min of UV degradation for the SGM plus® STR kit while 120 min of UV exposure was required for the same loss in the tri-allelic SNaPshot® assay. The first allelic loss for the SNaPshot® assay was observed after 60 min of UV irradiation. Resolution of two-person mixtures was possible for up to 1:8 ratios [58]. Another 20-plex SNaPshot® assay with tri-allelic SNPs generated reproducible identity profiles at 1 ng DNA input amount. The assay did not produce any profiles when tested on animal species including pig, chicken, rabbit, rat and loach fish [59]. A 16-X SNP SNaPshot™ assay (in combination with mini X-STRs) enhanced the identification of degraded DNA in Japanese samples. The sensitivity of the assay ranged from 50 pg to 10 ng DNA input amounts. The combined power of discrimination and the power of exclusion were greater than 99.99 and 99%, respectively [60].

The histone–DNA complexes of nucleosomes are known to be the sites preventing DNA degradation including apoptosis due to bacterial and environmental degradation [61]. An 18-plex SNaPshot® assay including the SNPs from these nucleosome regions had a sensitivity of 78 pg and was more effective than the SNP*for*ID 52-plex as well as the AmpFlSTR® Minifiler™ (Applied Biosystems) and AmpFlSTR® Identifiler™ (Applied Biosystems) STR assays for degraded DNA [61].

LISNP SNaPshot® assays

Lineage markers mostly include Y chromosome and mitochondrial (mt) DNA markers. Many Y-SNP and mtSNP SNaPshot® assays are available as shown in Tables 2 and 3. Before these, a solid-phase fluorescent minisequencing multiplex assay for 12 mtDNA polymorphism markers was developed and validated to identify British Caucasians and British Afro-Caribbeans [29, 30].

Y-SNP SNaPshot™ assays

In 2003, a 35-Y-SNP preliminary SNaPshot® minisequencing assay was developed with the intention of setting forensic parameters for SNaPshot® typing [62]. The sensitivity of the assay was in the range 100 pg to 10 ng DNA with an optimum of 1–2 ng. Reproducibility was demonstrated by concordant results for 194 male Danish samples typed in duplicate. The assay also illustrated the importance of primer design for SBE multiplex assays [62]. This work laid the platform for using SNaPshot® chemistry for a variety of other forensically relevant assays. The European SNP*for*ID consortium identified many potential Y-SNP and mtSNP for predicting lineages which assisted in the development of a series of SNaPshot® assays for haplogroup typing.

Vallone and Butler [63] examined 50 Y-SNPs able to differentiate US, African-American and Caucasian samples. Forty-two SNPs were typed using allele-specific hybridization

Table 2 Y chromosome lineage-informative single-nucleotide polymorphism (LISNP) SNaPshot® assays

Assay	Number of SNPs	Sensitivity (ng DNA)	Application	References
Global population assays				
Major Y chromosome haplogroup typing kit	29	0.25	Differentiates 12 world genealogies	Brión et al. [65]
28-plex	28	1–2	Differentiates major continental paternal lineages worldwide	van Oven et al. [66]
European population assays				
37-plex	37	0.05–1	Major clades of European sub-population	Onofri et al. [67]
Asian population assays				
16-plex assay	16	0.062–1	Differentiation of Asian haplogroup O	Park et al. [69]
Ancient DNA assay				
13 SNP aDNA assay	13	0.05–1	Ancient DNA samples	Bouakaze et al. [45]

(ASH) with flow cytometry detection, and 18 SNPs were typed using SBE with fluorescence detection. The SBE assay was designed in three multiplexes of six SNPs each. Ten SNPs were typed with both ASH and SBE methods. The results for both the genotyping methods were concordant, but the SBE method offered advantages of less time requirement and greater cost-effectiveness in re-analysing a sample in comparison to ASH. The study identified the need for additional Y-LISNPs. A comparative study between MALDI-TOF MS and SNaPshot® on eight Y-SNPs differentiating four European haplogroups also demonstrated the speed and accuracy of the SNaPshot® technique [64].

Global Y-SNP assays A major Y chromosome haplogroup typing kit consisting of 29 Y-LISNPs was subsequently developed which differentiated 1126 unrelated males from 12 worldwide populations into Y haplogroup lineages [65]. Only 12 SNPs were selected to divide the samples into 12 major clades, and the remaining SNPs subdivided some of these clades. This multiplex assay was subject to inter-laboratory validation using 10 human samples, and the assay overall defined 31 haplogroups. The Asian population samples were classified into six haplogroups out of 31, and 93% accuracy was reported for the detection of Southeast Asian population samples. Three haplogroups defined African samples with an additional haplogroup defining sub-Saharan African. Most of the African samples used in the study were of Somali origin. European samples were most abundant in the sample set and were classified into seven haplogroups. The assay had limitations in assigning admixed population

Table 3 Mitochondrial DNA (mtDNA) lineage-informative single-nucleotide polymorphism (LISNP) SNaPshot® assays

Assay	Number of SNPs	Sensitivity (ng DNA)	Application	References
Global population assays				
12-plex	12	0.007	Differentiates world genealogies	Nelson et al. [73]
71-plex	71	0.025	Differentiates R0 macro-haplogroup	Mosquera-Miguel et al. [74]
11-plex	11	1–2	Differentiation of genealogies in the Australian population	McNevin et al. [44]
36-plex	37	0.004	Differentiates 43 global haplotypes	van Oven et al. [75]
42-plex	42	0.01	Resolves Latin American admixture efficiently	Paneto et al. [76]
European population assays				
16-plex	16	0.0005–0.1	Differentiates West European haplogroups	Brandstätter et al. [77]
11-plex	11	0.0002–2	Resolution of European Caucasians	Vallone et al. [71]
17-plex	17	10	Differentiation of West Eurasian haplotypes	Quintáns et al. [78]
22-plex	22	–	Differentiates nine major European haplogroups	Köhnemann et al. [80]
Asian population assays				
20-plex	20		Differentiates the haplogroups in Andamanese populations	Endicott et al. [81]
32-coding mtSNP assay	32	5	Differentiates East Asian phylogeny	Álvarez-Iglesias et al. [82]

samples from Greenland and South America. This panel could differentiate the major population groups of the world but was more limited in differentiating closely related population groups [65].

More recently, a 28-Y-SNP SBE multiplex assay enabling the discrimination of major Y chromosome haplogroups worldwide has been developed. The assay was divided into two multiplexes to allow hierarchical typing. The recommended DNA amount was 1–2 ng, and further sensitivity tests were not conducted, but the PCR amplicon lengths were kept short in the range of 46–178 bp to make it suitable for degraded DNA samples. This assay can provide an assessment of the continental biogeographical male lineage only and requires additional SNPs for detailed phylogenetic classification [66].

European population Y-SNP assays A 37-Y-LISNP assay has been developed in six multiplexes for European lineage [67]. Two multiplexes with a total of 15 SNPs differentiate the major clades of the Y haplogroup tree (A-R) belonging to specific continents. The other four multiplexes differentiate European haplogroups. The sensitivity of the assays was in the range 50 pg–1 ng. The assay was shown to work on degraded DNA with quantities as low as 50 pg. The reproducibility of the assay was assessed by genotyping all the samples in duplicate which produced concordant results. The hierarchical multiplexes were designed in a way that at most, two amplification steps were required for determining the haplogroup for each sample. The first amplification step indicates the major continental clade, and the second amplification would depict the corresponding sub-clade of that sample. This assay was developed to differentiate closely related European population haplogroups but was not able to differentiate closely related population groups from other continents. This type of hierarchically designed multiplex assay could determine the specific genealogy of a sample in a forensic context [67].

Asian population Y-SNP assays A 16-Y-SNP SBE multiplex assay discriminating the males of haplogroup O mostly found in East and Southeast Asia has also been developed [68]. The assay uses smaller amplicon sizes in the range of 45–123 bp, making it applicable to degraded DNA. The assay could be used when a sample is found with a haplogroup O status from a global Y-SNP assay and further sub-lineage information is required [68]. Y-SNP miniplex assays are also available which can help dissect the high-occurrence haplogroups O and C in East Asian populations [69]. These Y-SNP miniplexes are a combination of four individual multiplexes with a total of 22 Y-SNPs. The first multiplex of six SNPs differentiates worldwide haplogroups. The other three multiplexes are designed to identify sub-haplogroups O, O3 and C. The sensitivity of the multiplexes was in the range of 62 pg–1 ng. When applied to an artificially degraded DNA, the assay produced concordant results with non-degraded controls, showing its reproducibility and reliability. The assay effectively typed 10 DNA samples from 55-year-old skeletal remains, and allele drop-in was not observed even when the amplification cycle number was increased from 33 to 35 or 37 cycles. This demonstrated the versatility of SNPs in comparison to STRs, which are prone to induce amplification errors when cycle numbers are increased. When used for typing 300 Korean samples, the assay correctly reported that the majority belonged to haplogroup O followed by haplogroup C. Haplogroup O3 signifies the migration patterns of modern East Asian populations, and this sort of information could be of forensic relevance [69].

Ancient DNA Y-SNP assays A 13-Y-SNP SNaPshot® assay designed specifically for typing of ancient DNA (aDNA) was developed on 11 bone samples from south Siberia [45]. As a single multiplex could not type any aDNA samples, it was redesigned in two PCR multiplexes of six and seven SNPs. This showed that a single multiplex assay with more markers might decrease amplification efficiency in the case of aDNA samples. Nine samples were successfully typed, and two very ancient samples failed to yield a result. The sensitivity of the assays was in the range of 50 pg–1 ng [45].

mtSNP SNaPshot® assays

Traditionally, when STR profiling fails, the typing of hypervariable parts of the mtDNA control regions (HV1 and HV2) often provides some identification information due to the high copy number of mtDNA. The inheritance of mtDNA is maternal, and due to the lack of recombination, it provides lower discrimination and identical HV1 and HV2 haplotypes are frequently encountered [70, 71]. The typing of mtSNPs in coding regions was used as an alternative to improve the discrimination power of mtDNA [72]. As a result, mtSNP typing can be used as a screening tool for eliminating multiple suspects or rapidly differentiating between many samples in high-volume cases.

Global mtSNP assays A 12-SNP multiplex assay defining the broad mtDNA haplogroups for different population samples requiring only 7 pg DNA to generate full profiles has been used to assign two World War II-era samples to their corresponding haplogroups [73]. A SNaPshot® assay defining different branches of macro-haplogroup R0 was designed using 71 mtSNPs in three multiplexes and was demonstrated to be accurate at 25 pg of DNA input amount. The assay was robust, and no contamination or amplification of NUMTs (nuclear mitochondrial DNA pseudo sequences) was found [74].

An Australian mtDNA SNP assay was used to assign 145 samples to 12 haplogroups using an 11-SNP multiplex assay [44]. The study demonstrated the difficulty in accounting for admixture with mtDNA markers. More recently, a 36-mtSNP

multiplex system has been employed to efficiently infer maternal ancestry at the continental level. The assay differentiated 43 different haplotypes with sensitivity down to 4 pg DNA. It was designed in three multiplexes of 12 SNPs each. The haplogroup assignment was consistent and concordant with full-sequence profiles [75].

Another 42-plex SNaPshot® assay was used to classify Latin American samples in an admixed population. The majority of the population (46.6%) was found to have African maternal lineage, 27.3% had European origin and 26.1% had an Asian origin. The complete profiles were obtained with only 10 pg DNA input amounts [76].

European population mtSNP assays A 16-SNP SNaPshot® mtSNP assay exists that is capable of discriminating between West European Caucasian haplogroups. The assay was developed in two equal multiplexes as a rapid screening method for elimination of multiple suspects. The power of discrimination and preliminary sensitivity were 88.6% and 25 pg, respectively [77]. Another 11-plex mtSNP assay was developed around the same time in an attempt to differentiate individuals with identical HV1/HV2 mitotypes in Caucasians [71]. The sensitivity of the assay was in the range of 0.2–2000 pg with an optimum of 1–2 pg for robust reproducibility. This assay displayed a limitation in detecting heteroplasmy with one highly ambiguous SNP [71]. A 17-plex SNaPshot® mtSNP assay was developed to allow differentiation of West Eurasian haplotypes in two PCR multiplexes. The first multiplex allocated samples to the most common European haplogroups, and the second multiplex differentiated the sub-haplogroups of the high-frequency European haplogroup H. Both the PCR multiplex products were combined in a single-SBE assay [78]. This assay separated haplogroup H into its sub-categories, thus complementing the 16-plex assay developed by Brandstätter et al. [77] and the 11-plex assay developed by Vallone et al. [71]. Grignani et al. [79] developed an assay using 25 mtSNPs to sub-type haplogroup H into sub-clades H1–H15. The assay was designed in two PCR multiplexes. Multiplex A was adapted from the 17-plex assay of Quintáns et al. [78] which separated H1–H7 sub-clades whereas multiplex B of eight SNPs differentiated H8–H15 sub-clades [79]. Another 22-mtSNP multiplex assay was developed to detect nine major European haplogroups and some of the sub-haplogroups [80].

Asian population mtSNP assays A 20-plex SNaPshot® assay was designed to target the M31 and M32 haplogroups found in Andaman Islanders using 20 ancient Andaman samples. The assay defined the fine structure of haplogroup M31, and two new sub-classes M31a1a and M31a1b were identified, supporting a division between greater Andamanese and Onge–Jarawa-speaking people. A sub-clade M32a1 was also identified to be specific to the Onge–Jarawa population [81]. A 32-coding mtSNP assay was developed to haplotype East Asian phylogeny with its Native American-derived branches [82]. A 15-plex mtSNP SNaPshot® assay was developed with a haplotype diversity of 0.9136, differentiating 28 haplotypes of the Chinese Yi population group. The assay showed the close relationship between the Chinese Yi and Bai populations [83].

Indigenous American population mtSNP assays Recently, a 26-plex SNaPshot® assay, AmericaPlex26, has been designed to genotype the human mitochondrial founder lineages of America [84]. The assay targeted sites within haplogroups A2, C, C1c, D, D4e, D2a and X, which commonly occur in American lineages. The assay was shown to work on degraded DNA and could be used as a screening tool to assess the sample preservation strategy and the presence of lineages other than above [84].

AISNP SNaPshot® assays

AISNPs are designed to distinguish between populations; therefore, their ideal characteristics are opposite to those of IISNPs. AISNPs have low heterozygosity and high F_{ST} between populations [7]. AISNP assays provide BGA information about the donor of a DNA sample (Table 4). As autosomal markers are co-inherited maternally and paternally, their advantage over lineage markers is that they are able to indicate recent admixture in individuals. In addition, mtDNA and Y chromosome markers will not reveal genetic inheritance from maternal grandfathers and paternal grandmothers which may bias ancestry estimates [42].

SNPforID 34-plex assay

The SNP*for*ID 34-plex SNaPshot® assay is a well-established ancestry-informative assay, differentiating between Europeans, Asians and Africans [42]. A naïve Bayesian classifier implemented via the *Snipper* web portal [85] can be used to estimate likelihood ratios of population membership. The assay has been demonstrated to be effective at 200 pg DNA amounts, but an optimum of 1–2 ng DNA was recommended [42].

The 11-M Madrid Bombings in 2004 represented a successful casework application of the SNP*for*ID 34-plex autosomal ancestry SNP assay. In this case, seven STR profiles from evidential samples were unmatched. Ancestry was assigned according to the 34-plex predictor model. Three samples were found to be of North African origin and one of European origin. The remaining three samples were not assigned, as their probabilities were lower than the predictor threshold. In one case, the 34-plex system revealed North African origin where Y and mtDNA loci did not. Later, familial searching of a Spanish DNA database indicated that this sample had an Algerian origin. This intelligence information led investigators to the suspect [86].

Table 4 Ancestry-informative single-nucleotide polymorphism (AISNP) SNaPshot® assays

Assay	Target	Number of SNPs	Sensitivity (ng DNA)	Application	References
SNP*for*ID 34-plex	Autosomal	34	0.2–2	Differentiation of continental populations	Phillips et al. [42]
Eurasiaplex	Autosomal	23	1	Differentiation of European and South Asian populations	Phillips et al. [36]
Pacifiplex	Autosomal	29	≥0.125	Differentiation of Oceania populations	Santos et al. [37]
EurEas_Gplex	Autosomal	14	0.5–2	Differentiation of European and East Asian populations and gender identification	Daca-Roszak et al. [88]
Global AIMs Nano	Autosomal	31	0.064	Differentiation of African, European, East Asian, Oceanian and Native American populations	de la Puente et al. [89]

Recently, the revised 34-plex assay was published, where the SNP rs727811 in the original panel was replaced by rs3827760 to improve resolution and performance of the assay [34]. The SNaPshot® assay design was re-optimised with new PCR and SBE primers. The amplification cycles for the PCR step were reduced from 35 to 30 and, for the SBE step, from 30 to 28. The assay includes two tri-allelic SNPs, which are useful for identifying contributors to a mixture. Samples with three or more admixed ancestries were difficult to resolve [34]. When applied to US population samples, the 34-plex assay showed that non-admixed samples and samples with two dominant co-ancestries were classified accurately. It was more difficult to identify ancestral populations in the highly admixed Hispanic samples [87].

Eurasiaplex assay

Eurasiaplex is a 23-plex SNaPshot® multiplex assay designed to complement the SNP*for*ID 34-plex assay [36]. Eurasiaplex, in combination with the 34-plex assay, differentiates Europeans and South Asians (especially west of Europe, India and Pakistan towards Afghanistan). However, misclassification errors were observed towards Eastern Europe (Turkey, South Caucasus) and were worst for Middle Eastern populations. This indicated that small-scale forensic multiplex assays are limited in discriminating continuous genetic variation among geographically close populations. The authors suggested that more than 100 SNPs would be needed for proper separation of Middle Eastern populations [36].

Pacifiplex

Pacifiplex was designed to complement the SNP*for*ID 34-plex assay in differentiating East Asians and Oceanian populations, in a 29-plex SNaPshot® multiplex assay [37]. The sensitivity of the assay was 125 pg, and the assay could be complementary to Y/mtDNA analyses if highly degraded and admixed samples are encountered. The assay potentially genotyped 50-year-old serum samples providing evidence of its versatility on challenging DNA samples. The combined Pacifiplex and SNP*for*ID 34-plex assay was able to differentiate Aboriginal Australians and Papua New Guineans [37].

EurEAs_Gplex

More recently, a 14-SNP sub-classification SNaPshot® assay known as EurEAs_Gplex has been published to be capable of discriminating European and East Asian ancestries along with gender identification [88]. The sensitivity of the assay was 500 pg, and the recommended optimum DNA input amount was 2 ng. This reduced SNaPshot assay was shown to differentiate continental populations when applied to artificially sonicated DNA samples [88].

Global AIMs Nano assay

The 31-plex nano SNaPshot assay is a compact version of the EUROFORGEN Global AIMs panel [89]. It consists of 28 bi-allelic and 3 tri-allelic SNPs and has been designed to differentiate between African, European, East Asian, Oceanian and Native American populations. The sensitivity of the assay enables the analysis of 64 pg that makes it suitable for degraded samples. The inclusion of tri-allelic SNPs may assist with mixture detection [89].

16-plex assay

A 16-plex SNaPshot® assay successfully inferred the BGA of six major ethnic population groups in Australia. Assignment accuracies of 93.5, 91.9, 100 and 94.1% were reported for classifying samples as Asian, Caucasian, sub-Saharan African and North African, respectively. The prediction accuracies for Middle Eastern (71.4%) and Continental Asian (82.8%) assignments were slightly lower. The sensitivity of the assay ranged from 140 pg to 2 ng DNA amounts in the evidentiary-type samples used for its development [90].

Native American admixed assays

A 128-SNP TaqMan AIM assay has been demonstrated to differentiate between American admixed populations [91]. A subset of 14 of these SNPs were combined in a SNaPshot® assay, developed as two multiplexes, to differentiate between African, European and Latin American populations, but it had limited ability to differentiate admixed Latin American populations from Southeast Brazil [92]. This smaller assay was designed as a cost-effective option for low-throughput labs [91, 92]. Another 28-plex SNaPshot assay was able to differentiate admixed Brazilian Native Amerindians (five regions in Brazil) and STRUCTURE multi-locus genotype clustering, which indicated that more than 90% of the admixed samples used in the study came from the European ancestry [93]. Similarly, a 24-plex SNaPshot assay developed by Corach et al. [94] was able to identify admixture in an Argentinean population with all samples used in the study displaying European and Native American admixture [94].

PISNP SNaPshot® assays

Prediction of EVCs can provide forensic intelligence about the physical characteristics of a DNA donor (such as eye, hair and skin colours) (Table 5). Currently, two well-established SNaPshot®-based phenotypic assays have been validated for the European population: IrisPlex [95] and HIrisPlex [96].

IrisPlex assay

IrisPlex is a blue and brown eye colour classification system comprised of six highly predictive eye colour SNPs [95, 97]. The sensitivity of the assay was 15–500 pg with reproducible profiles obtained at 31 pg DNA input amounts. The accuracy of blue and brown eye colour prediction was greater than 90% in a European population dataset when using a multiple logistic regression (MLR) prediction algorithm [95]. SBE primers for two SNPs (rs1800407 and rs12203592) were subsequently redesigned to increase the resolution at low template amounts and avoid sporadic effects encountered in the original IrisPlex assay [97]. Blind trials were performed on artificially created single-source and mixed (two contributors) casework-type samples from the blood, semen, saliva and touched surfaces. These revealed 100% genotyping consistency for single-source samples, but mixtures were difficult to detect due to the limited polymorphic nature of bi-allelic SNPs [97]. Non-blue and non-brown eye colours were classified as an intermediate category. The individual prediction of intermediate eye colour (such as green, grey and hazel) has lower prediction accuracy with the currently available SNPs and prediction tools [97].

The IrisPlex assay was also assessed in a Slovenian population, and this revealed prediction accuracies of 96.6, 91.3 and 79.6% for blue, brown and intermediate eye colours, respectively. The sensitivity (proportion of correct eye colour predictions) was highest for blue eye colour (93.6%) while brown and intermediate eye colours were less sensitive (58.1 and 0%, respectively). The zero sensitivity of the assay in predicting intermediate eye colour confirmed that more predictive markers are required. SNP rs1800407 that is claimed to be the next best predictive eye colour marker after rs121913832 had a weak effect on this population [98].

The IrisPlex system has been evaluated in a North American US population. The assay was performed in two PCR multiplexes of four and two SNPs (compared to the original one PCR multiplex) [95, 97], and then PCR products were pooled for a single-SBE multiplex reaction. Iris colour was determined using MLR as well as a Bayesian network model. This study had a greater number of intermediate eye colour phenotypes than in the original IrisPlex study, and hence, more inconclusive results were encountered. The Bayesian model offered better predictions than multinomial logistic regression (MLR) as well as offering the flexibility of calculating likelihood ratios which could be more convenient for reporting [99].

HIrisPlex assay

HIrisPlex is a 24-plex assay (23 SNPs and 1 INDEL) capable of predicting eye and hair colour collectively and includes the six IrisPlex SNPs. The sensitivity of the assay was in the range 31–500 pg with allelic dropout observed at 31 pg input amount. The profiles were reproducible at 63 pg template input [96]; hence, HIrisPlex demonstrated greater sensitivity

Table 5 Phenotype-informative single-nucleotide polymorphism (PISNP) SNaPshot® assays

Assay	Target	Number of SNPs	Sensitivity (ng DNA)	Application	References
IrisPlex	Autosomal	6	0.015–0.5	Eye colour prediction	Walsh et al. [97]
HIrisPlex	Autosomal	24	0.031–0.5	Hair and eye colour prediction	Walsh et al. [96]
8-Plex	Autosomal	8	≥0.1	Skin and eye colour prediction	Wurmbach [101]

than IrisPlex. Hair colour prediction was classified into four categories: blond, brown, red and black. The prediction accuracies were 69.5% for blond, 78.5% for brown, 80% for red and 87.5% for black hair colours in the European test dataset. The inaccurate predictions of age-related hair colour change, grey hair and intermediate eye colours were highlighted as major limitations of this tool. These limitations could be improved with the future discovery of DNA markers capable of resolving these highly variable traits [96].

The HIrisPlex assay has been tested on degraded, ancient DNA samples. Twenty-one tooth samples with ages ranging from 1 to 800 years and five contemporary bone samples were used. Of the 26 samples, 24 delivered full profiles with prediction accuracies consistent with those above for both eye and hair colours. The HIrisPlex profile from the DNA of a World War I Polish General revealed the same phenotype (blue eyes, blond hair) as mentioned in historical documents. The research provided evidence that accurate EVC prediction from degraded and ancient DNA depends on sample storage and environmental effects. The sensitivity of the assay remained 62 pg, as originally reported [96], except that one of the skeletal remains generated a full profile at 31 pg [100]. The study showed the applicability of HIrisPlex for skeletal and degraded remains.

8-plex assay

An 8-plex SNaPshot® assay was developed to predict eye and skin colour that had three SNPs in common with the IrisPlex assay. The sensitivity of the assay was shown to be 100 pg. Five of the eight SNPs were used in eye colour prediction and six in skin colour prediction. Skin colour is predicted in light, medium and dark categories using the predictor tool developed by the group using a training set of 803 independent samples. Eye colour is predicted into three categories: blue, brown and green. An error rate of 5% was estimated for eye colour prediction, and skin colour prediction was 62% accurate in European population samples [101, 102].

Other pigmentation assays

A SNaPshot® assay of 37 pigment-associated SNPs was developed to further understand the intermediate eye colour prediction and contained all six IrisPlex SNPs. The eye colour categories were divided into light and dark blue and brown colours, and the intermediate eye colour had a sub-category of green–hazel colour. The rs12913832–rs1129038 combination was able to identify light, blue, inter-light and green–hazel categories with sensitivities of 96.5, 98.5, 88.9 and 75.3%, respectively. There were four additional *HERC2* gene SNPs to improve the distinction of eye colour. The 13 SNPs used in the assay for intermediate eye colour were unable to provide a clear resolution, and it was noted that more informative prediction markers were required. This study used the Bayesian Snipper classifier for predictions of eye colours. The study also emphasised the need for a uniform procedure for eye colour phenotype documentation to reduce errors associated with the human perception of eye colour [103].

A 12-plex SNaPshot® assay was developed for eye and hair colour prediction in the Slovenian population and was published earlier than HIrisPlex. The results revealed a significant association of five SNPs out of 12 with eye and hair colour, and all five SNPs are included in the HIrisPlex assay. The optimal sensitivity of the assay was 1 ng, but if polymerase concentration was increased by five times, then the assay produced full profiles at 62 pg. Two prediction models (MLR and Bayesian network models) were developed based on the five most strongly associated SNPs for eye and hair colour prediction. The comparison between the two models showed that MLR was somewhat better than the Bayesian network model in making accurate predictions [104].

Combined ancestry and phenotypic SNaPshot® assays

SNaPshot® assays containing a combination of ancestry and phenotypic SNPs have been developed. These assays help to infer BGA and EVCs together depending on the SNPs included in the multiplex. One example is a 10-plex SBE assay that was developed on 27 modern human samples and then tested on 25 skeletal remains. The panel was selected from six candidate genes and comprised of four ancestry and eight phenotypic SNPs (a few SNPs overlapped for both ancestry and phenotype predictions). The probability estimates for modern human samples determined using STRUCTURE were mostly greater than 80% for inferring BGA except two Asian samples which indicated probabilities of approximately 70%. When tested on 25 degraded ancient DNA samples, the assay revealed that most were derived from European origins, one had equal contributions from European and Asian origins and two samples had Asian origins [105].

A 32-plex assay designed to complement the SNP*for*ID 34-plex was developed for more distinct Eurasian ancestry inference, and it comprised of 22 ancestry SNPs and 10 phenotypic SNPS for eye, hair and skin colour prediction. STRUCTURE analysis revealed this assay alone did not optimally differentiate South Asians from Europeans but, when combined with the 34-plex [34] and Eurasiaplex [36] assays, yielded better ancestry inference. The IrisPlex MLR prediction model could not adequately predict intermediate eye colour when tested on Turkish population samples, but Snipper's likelihood values were more indicative. However, the need for more informative intermediate eye colour predictive markers remained unchanged [106].

More recently, a SNaPshot® assay with 50 SNPs for inferring BGA and phenotypic traits in the US population was developed. The assay was designed in three multiplexes comprised of 32 AISNPs and 18 PISNPs. The ancestry inference

made using the Snipper model revealed 77% accuracy with 21.6% of samples inconclusive and 1.4% misclassified. Prediction using the published IrisPlex MLR model was made in two sets: Europeans and non-Europeans. At 0.7 thresholds, Europeans were predicted with 81% accuracy. The inaccurate predictions were mostly for subjects having intermediate eye colour that was misclassified either as blue or brown. The non-European set was mostly comprised of brown-eyed people, and hence, 99% accuracy was achieved with two intermediate eye colour samples not predicted correctly. The Bayesian Snipper model offered more flexibility than the regression model in cases with missing data [107].

Non-human SNaPshot® assays

SNaPshot® has also been applied to non-human forensic DNA analysis, and a few examples are listed here.

Forensic entomology

SNaPshot assays relevant to the forensic entomology field have been developed. For example, a 6-plex blowfly species identification assay differentiating seven common Calliphoridae blowflies found in the UK has been developed [108]. These blowflies are generally the first to populate cadavers, and species identification could assist in determining time of death. The SNPs are from the cytochrome oxidase I gene, and distinctive haplotypes were identified for each species.

Microbial forensics

A SNaPshot® multiplex assay of five SNP species-specific primers has been designed for the species identification of *Lactobacillus casei* group based on the conserved regions of the *dnaK* gene [109]. The assay consisted of group-specific and species-specific primers and was shown to successfully assign all 63 strains to *L. casei* group and explicitly differentiated all *L. casei* strains from *Lactobacillus paracasei* and *Lactobacillus rhamnosus* simultaneously [109].

Wildlife forensics

The prevention of wildlife trafficking and protecting endangered species is a major focus of wildlife forensics research. A mtSNP SNaPshot® assay designed to identify 11 tiger species and sub-species is one such example. Five SNPs were species specific and another six were sub-species specific with three primer pairs designed to amplify all 11 SNPs. The SBE reaction was performed using 11 SNP-specific primers. The method was 100% accurate when used to identify 15 tigers with a sensitivity of 0.26 pg. A specificity test shows this assay's potential for the identification of other big cat species (closely related to *Panthera*) in addition to tiger species [110].

Discussion and conclusions

STRs are considered to be the gold standard for human identification but are less suited to trace and damaged DNA than SNPs due to their long repeat sequences, resulting in larger PCR amplicons. Furthermore, SNPs can provide identity, lineage, ancestry and phenotype information. The SNaPshot® minisequencing assay is a versatile forensic SNP genotyping tool which can be easily integrated into operational forensic laboratories without any investments in additional equipment. The custom multiplex assays described here can be widely applied for forensic human DNA SNP analysis. The SNP*for*ID 52-plex IISNP assay could be used as a supplementary identity assay alongside conventional proprietary STR assays, especially where degraded and low amounts of DNA are involved [54]. SNP*for*ID 52-plex profiles have been obtained from highly degraded and complex samples including bones, teeth, crime scene samples and a decomposed and charred femur where current STR profiling systems failed or only produced partial profiles [111, 112]. The 52-plex assay can also be a useful complementary tool to STR profiling for resolving paternity cases [113]. Validation studies exist which demonstrate its applicability to difficult forensic casework samples [114].

Forensic DNA phenotyping (FDP), also known as molecular photofitting, refers to the process of predicting the BGA and EVCs of a donor of an evidentiary DNA sample. It can lead investigators to narrow a pool of suspect(s) in cases when STR profiling is uninformative. In these cases, investigators may alternatively use eyewitness statements which are known to be unreliable [115]. FDP has gained attention in the forensic community with the increasing discovery of potential markers and genes associated with many physical traits [116]. BGA and EVC information can be used as a molecular 'silent witness' [117]. With the development of more DNA-based intelligence assays providing information for more physical traits (including facial morphology), we are moving towards more accurate 'molecular photofits'. SNaPshot® lineage, ancestry and phenotypic assays are potential FDP tools capable of providing DNA intelligence information that would certainly help investigators to focus their resources more effectively and efficiently.

The number and types of SNaPshot® assays available enable users to adopt a hierarchical approach to the analysis of samples. The lineage-informative assays such as the 28-Y-LISNP [66] and 36-mt LISNP [75] assays can indicate major continental origins. If the sample is of European origin, the 37-Y-LISNP [67] assay could then be used to infer the specific European haplogroup. There is a range of mitochondrial paternal assays that can be selected to further differentiate European lineages such as the 22-plex assay differentiating nine major European clades, the 16-plex assay separating West European Caucasian clades [77], the 11-plex assay differentiating identical European mitotypes [71] and the 25-

mtSNP assay extricating clades H1 to H15 [79]. The ability to separate specific clades or sub-clades could be of significant importance in a mass disaster victim identification (DVI) case. Alternatively, a screening tool based on a subset of informative control region substitution sites can assist in eliminating a large proportion of samples from an investigation ahead of a more detailed sequence-based analysis. Such a tool has been established by a group at the Netherlands Forensic Institute (NFI) and has been evaluated for operational value by the European DNA Profiling (EDNAP) group in 2016 ([118], submitted manuscript).

The autosomal ancestry-informative SNP assays could also be applied to provide investigators with BGA information. The SNP*for*ID 34-plex [34, 42] SNaPshot® assay is a validated tool that can differentiate between three major world populations: Asian, African and European. The Global AIMs Nano 31-plex assay can be used to differentiate between African, East Asian, European, Oceanian and Native American populations [89]. There are also tools available to deconvolute admixed samples, which can complement the 34-plex assay. Eurasiaplex [36] and EurEas_Gplex [88] can urther assist in offering higher-resolution differentiation of Europeans and Asians (East Asians). Pacifiplex could play a critical role in differentiating Oceanian populations (such as Australian Aboriginals and Papua New Guinea) from global populations [37] which is useful for the analysis of samples in the Asia-Pacific region. These ancestry tools were applied in providing investigator leads in solving some high-profile cases. One such example is the 11-M bombings where the SNP*for*ID 34-plex assay confirmed the North African origin of an evidentiary sample that led investigators to a perpetrator [86]. The 34-plex assay was also employed in Operation Minstead, Britain's largest investigation, and provided investigators with evidence that the suspect was most likely to have admixed African origins from the Caribbean or mainland America [119]. Similarly, it was employed in the investigation of a murder in Madrid to confirm that the suspect was Moroccan which enabled police to narrow a pool of suspects to a few from many thousands [120]. Such assays could potentially be applied to cases of illegal trafficking of organ transplants [121]. Acceptance of the SNP*for*ID 34-plex assay by the forensic community was demonstrated in a global trial of binary AIMs assays [122].

In addition to ancestry prediction, intelligence can be generated using PISNP SNaPshot® assays to infer the EVC of the donor of an evidentiary DNA sample. The HIrisPlex system could be utilised to obtain eye and hair colour information, not only from pristine human DNA samples but also from ancient DNA, provided that prediction accuracy in non-human European populations is characterised [96, 123]. Elucidation of skin tone is possible using the 8-plex skin colour prediction tool [101, 102] with a 'HIrisPlex-S' system, capable of predicting eye, hair and skin colours, in development (Manfred Kayser and Susan Walsh, personal communication). There are numerous other EVCs of potential forensic value with associated SNPs such as male pattern baldness [124], hair texture [125], facial characteristics [126], fingerprint patterns [127] and age estimation [128]. Future SNaPshot® assays may incorporate some or all of these.

The detection of mixtures using bi-allelic SNPs remains challenging due to their low polymorphic nature consisting of only two alleles [96, 97]. For example, the mixture of two single-source samples with homozygote and heterozygote genotypes for a bi-allelic SNP would combine to generate a heterozygote genotype, indistinguishable from the original heterozygote contributor. Even if a mixture is suspected, deconvoluting the mixture may not be possible. This challenge could be overcome by including tri- or tetra-allelic SNPs in the assays, as for the SNP*for*ID 34-plex assay which contains two tri-allelic SNPs [34]. A Global ancestry-informative marker set developed by EUROFORGEN was purpose built with six tri-allelic SNPs to help identify mixed-source samples [129]. Further utilisation of tri-allelic SNPs associated with FDP will only improve the utility of SNaPshot® assays [58].

In recent times, MPS has been gaining popularity in the forensic community due to its ability to type large batteries of markers in multiple samples simultaneously [21, 130]. This technology has demonstrated potential to type identity, BGA and EVC markers together and hence can provide identity and FDP information in a single run [131]. However, SNaPshot® is a low-cost and time-efficient alternative to MPS for smaller-scale genotyping requirements and is ideal for laboratories that do not have the resources to consider MPS. A customised approach to MPS analysis of BGA and EVCs using the PCR products from existing SNaPshot® multiplexes has been demonstrated [21, 130]. Thus, even for labs that may adopt MPS, SNaPshot® assays remain useful and provide a flexible, modular approach to FDP (or identity) where population reference databases already exist for published SNaPshot® assays. This approach offers a reduction in costs associated with commercial panels.

In conclusion, SNaPshot® is an easily integrable and cost-effective SNP typing option for forensic laboratories with readily available forensic human and non-human DNA assays.

Acknowledgements The authors gratefully acknowledge the funding from the Australian Research Council (Linkage Project 110100121: 'From genotype to phenotype: molecular photofitting for criminal investigations').

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interest.

References

1. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. Nat Rev Genet 5:739–51
2. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. Am J Hum Genet 90:7–24
3. Choudhury A, Hazelhurst S, Meintjes A et al (2014) Population-specific common SNPs reflect demographic histories and highlight regions of genomic plasticity with functional relevance. BMC Genomics 15:1
4. Gill P, Sparkes R, Tully G (2001) DNA profiling in forensic science. Nature Publishing Group
5. Prinz M, Caragine T, Shaler R (2003) DNA testing as the primary tool for the victim identification effort after the World Trade Center terrorist attack. Proceedings of the 20th Congress of the International Society of Forensic Genetics
6. Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Sci Med Pathol 3:200–5. doi:10.1007/s12024-007-0018-1
7. Budowle B, vanDaal A (2008) Forensically relevant SNP classes. Biotechniques 44:603–10
8. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12:179–92
9. Gyllensten UB, Erlich HA (1988) Generation of single-stranded DNA by the polymerase chain reaction and its application to direct sequencing of the HLA-DQA locus. Proc Natl Acad Sci 85:7652–6
10. Walsh PS, Fildes N, Louie AS, Higuchi R (1991) Report of the blind trial of the Cetus Amplitype HLA DQ alpha forensic deoxyribonucleic acid (DNA) amplification and typing kit. J Forensic Sci 36:1551–6
11. Herrin G, Fildes N, Reynolds R (1994) Evaluation of the AmpliType PM DNA test system on forensic case samples. J Forensic Sci 39:1247
12. Rascati RJ (2003) DNA profiling by multiplex PCR amplification and genotype determination by reverse dot-blot hybridization to sequence-specific oligonucleotide probes: Amplitype® PM & DQA1 amplification and analysis. In: O'Donnell MA (ed) Forensic DNA analysis. Association for Biology Laboratory Education (ABLE)., pp 173–90
13. Primorac D, Anđelinović Š, Definis-Gojanović M, Drmić-Hofman I (1996) Identification of war victims from mass graves in Croatia, Bosnia and Hercegovina by the use of standard forensic methods and DNA typing. J Forensic Sci 41:891–4
14. Baird ML (1998) Use of the AmpliType PM+ HLA DQA1 PCR amplification and typing kits for identity testing. Forensic DNA Profiling Protocols. Springer. pp. 261–77
15. Edwards A, Civitello A, Hammond HA, Caskey CT (1991) DNA typing and genetic mapping with trimeric and tetrameric tandem repeats. Am J Hum Genet 49:746
16. Mehta B, Daniel R, McNevin D (2013) High resolution melting (HRM) of forensically informative SNPs. Forensic Sci Int 4:e376–e7
17. Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35:3036–43. doi:10.1002/clps.201400089
18. Reed GH, Kent JO, Wittwer CT (2007) High resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics 8:597–608
19. Sobrino B, Brion M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Sci Int 154:181–94
20. Krjutškov K, Viltrop T, Palta P et al (2009) Evaluation of the 124-plex SNP typing microarray for forensic testing. Forensic Sci Int Genet 4:43–8
21. Daniel R, Santos C, Phillips C et al (2015) A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci Int Genet 14:50–60
22. Wang Q, Fu L, Zhang X et al (2016) Expansion of a SNaPshot assay to a 55-SNP multiplex: assay enhancements, validation, and power in forensic science. Electrophoresis 37:1310
23. Shumaker JM, Metspalu A, Caskey CT (1996) Mutation detection by solid phase primer extension. Hum Mutat 7:346–54
24. Braun A, Little DP, Köster H (1997) Detecting CFTR gene mutations by using primer oligo base extension and mass spectrometry. Clin Chem 43:1151–8
25. Haff LA, Smirnov IP (1997) Multiplex genotyping of PCR products with MassTag-labeled primers. Nucleic Acid Res 25:3749–50
26. Syvänen A-C (1999) From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms. Hum Mutat 13:1–10
27. Sokolov BP (1990) Primer extension technique for the detection of single nucleotide in genomic DNA. Nucleic Acids Res 18:3671
28. Syvänen A-C, Aalto-Setälä K, Harju L, Kontula K, Söderlund H (1990) A primer-guided nucleotide incorporation assay in the genotyping of apolipoprotein E. Genomics 8:684–92
29. Tully G, Sullivan KM, Nixon P, Stones RE, Gill P (1996) Rapid detection of mitochondrial sequence polymorphisms using multiplex solid-phase fluorescent minisequencing. Genomics 34:107–13
30. Morley J, Bark J, Evans C, Perry J, Hewitt C, Tully G (1999) Validation of mitochondrial DNA minisequencing for forensic casework. Int J Legal Med 112:241–8
31. Grimes EA, Noake PJ, Dixon L, Urquhart A (2001) Sequence polymorphism in the human melanocortin 1 receptor gene as an indicator of the red hair phenotype. Forensic Sci Int 122:124–9
32. Applied Biosystems. ABI PRISM® SNaPshot™ Multiplex Kit. P/N 4323357 Rev. B ed. Thermo Fisher Scientific. pp. 1–42
33. Podini D, Vallone PM (2009) SNP genotyping using multiplex single base primer extension assays. Single nucleotide polymorphisms. Springer. pp 379–91
34. Fondevila M, Phillips C, Santos C et al (2013) Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet 7:63–74
35. Sanchez JJ, Phillips C, Børsting C et al (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis 27:1713–24
36. Phillips C, Aradas AF, Kriegel AK et al (2013) Eurasiaplex: a forensic SNP assay for differentiating European and South Asian ancestries. Forensic Sci Int Genet 7:359–66
37. Santos C, Phillips C, Fondevila M et al (2016) Pacifiplex: an ancestry-informative SNP panel centred on Australia and the Pacific region. Forensic Sci Int Genet 20:71–80
38. Untergasser A, Cutcutache I, Koressaar T et al (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115
39. Vallone PM, Butler JM (2004) AutoDimer: a screening tool for primer-dimer and hairpin structures. Biotechniques 37:226–31
40. Owczarzy R, Tataurov AV, Wu Y et al (2008) IDT SciTools: a suite for analysis and design of nucleic acid oligomers. Nucleic Acids Res 36:W163–W9
41. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinf 13:134
42. Phillips C, Salas A, Sánchez JJ et al (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet 1:273–80

43. Daniel R (2009) A new era in forensic intelligence: SNPs and the inference of biogeographical ancestry. University of Technology Sydney
44. McNevin D, Bate A, Daniel R, Walsh SJ (2011) A preliminary mitochondrial DNA SNP genotyping assay for inferring genealogy. Aust J Forensic Sci 43:39–51
45. Bouakaze C, Keyser C, Amory S, Crubezy E, Ludes B (2007) First successful assay of Y-SNP typing by SNaPshot minisequencing on ancient DNA. Int J Legal Med 121:493–9
46. Forensic S (1996) The evaluation of forensic DNA evidence. National Academies Press (US)
47. Nei M (1977) F-statistics and analysis of gene diversity in subdivided populations. Ann Hum Genet 41:225–33
48. Doi Y, Yamamoto Y, Inagaki S, Shigeta Y, Miyaishi S, Ishizu H (2004) A new method for ABO genotyping using a multiplex single-base primer extension reaction and its application to forensic casework samples. Legal Med 6:213–23
49. Palacajornsuk P, Halter C, Isakova V et al (2009) Detection of blood group genes using multiplex SNaPshot method. Transfusion 49:740–9
50. Lee HY, Park MJ, Yoo J-E, Chung U, Han G-R, Shin K-J (2005) Selection of twenty-four highly informative SNP markers for human identification and paternity analysis in Koreans. Forensic Sci Int 148:107–12
51. Musgrave-Brown E, Ballard D, Balogh K et al (2007) Forensic validation of the SNPforID 52-plex assay. Forensic Sci Int Genet 1:186–90
52. Børsting C, Rockenbauer E, Morling N (2009) Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard. Forensic Sci Int Genet 4:34–42
53. Borsting C, Sanchez JJ, Hansen HE, Hansen AJ, Bruun HQ, Morling N (2008) Performance of the SNPforID 52 SNP-plex assay in paternity testing. Forensic Sci Int Genet 2:292–300
54. Schwark T, Meyer P, Harder M, Modrow J-H, von Wurmb-Schwark N (2012) The SNPforID assay as a supplementary method in kinship and trace analysis. Transfus Med Hemother 39:187–93
55. Barbaro A, Phillips C, Fondevila M, Carracedo Á, Lareu M (2009) Population data of 52 autosomal SNPs in Italian population. Forensic Sci Int Genet Suppl Ser 2:351–2
56. Børsting C, Mogensen HS, Morling N (2013) Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. Forensic Sci Int Genet 7:345–52
57. Lou C, Cong B, Li S et al (2011) A SNaPshot assay for genotyping 44 individual identification single nucleotide polymorphisms. Electrophoresis 32:368–78
58. Westen AA, Matai AS, Laros JF et al (2009) Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. Forensic Sci Int Genet 3:233–41
59. Li Z, Yan J, Tang D et al (2013) Validation of a multiplex system with 20 tri-allelic SNP loci for forensic identification purposes. Forensic Sci Int Genet Suppl Ser 4:e324–e5
60. Oki T, Hayashi T, Ota M, Asamura H (2012) Development of multiplex assay with 16 SNPs on X chromosome for degraded samples. Legal Med 14:11–6
61. Freire-Aradas A, Fondevila M, Kriegel A-K et al (2012) A new SNP assay for identification of highly degraded human DNA. Forensic Sci Int Genet 6:341–9
62. Sanchez JJ, Borsting C, Hallenberg C, Buchard A, Hernandez A, Morling N (2003) Multiplex PCR and minisequencing of SNPs-a model with 35 Y chromosome SNPs. Forensic Sci Int 137:74–84
63. Vallone PM, Butler JM (2004) Y-SNP typing of US African American and Caucasian samples using allele-specific hybridization and primer extension. J Forensic Sci 49:723–32
64. Lessig R, Edelmann J, Zoledziewska M, Dobosz T, Fahr K, Kostrzewa M (2004) SNP-genotyping on human Y-chromosome for forensic purposes: comparison of two different methods. International Congress Series. Elsevier. pp. 334–6
65. Brión M, Sanchez JJ, Balogh K et al (2005) Introduction of an single nucleodite polymorphism-based "major Y-chromosome haplogroup typing kit" suitable for predicting the geographical origin of male lineages. Electrophoresis 26:4411–20
66. van Oven M, Ralf A, Kayser M (2011) An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. Int J Legal Med 125:879–85
67. Onofri V, Alessandrini F, Turchi C, Pesaresi M, Buscemi L, Tagliabracci A (2006) Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic Sci Int 157:23–35
68. van Oven M, van den Tempel N, Kayser M (2012) A multiplex SNP assay for the dissection of human Y-chromosome haplogroup O representing the major paternal lineage in East and Southeast Asia. J Hum Genet 57:65–9
69. Park MJ, Lee HY, Kim NY, Lee EY, Yang WI, Shin K-J (2013) Y-SNP miniplexes for East Asian Y-chromosomal haplogroup determination in degraded DNA. Forensic Sci Int Genet 7:75–81
70. Coble M, Just R, OC JE et al (2004) Single nucleotide polymorphisms over the entire mtDNA genome that increase the power of forensic testing in Caucasians. Int J Leg Med 118:137–46
71. Vallone PM, Just RS, Coble MD, Butler JM, Parsons TJ (2004) A multiplex allele-specific primer extension assay for forensically informative SNPs distributed throughout the mitochondrial genome. Int J Legal Med 118:147–57
72. Ziętkiewicz E, Witt M, Daca P et al (2012) Current genetic methodologies in the identification of disaster victims and in forensic analysis. J Appl Genet 53:41–60
73. Nelson TM, Just RS, Loreille O, Schanfield MS, Podini D (2007) Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing. Croat Med J 48:0–472
74. Mosquera-Miguel A, Alvarez-Iglesias V, Cerezo M, Lareu M, Carracedo A, Salas A (2009) Testing the performance of mtSNP minisequencing in forensic samples. Forensic Sci Int Genet 3:261–4
75. van Oven M, Vermeulen M, Kayser M (2011) Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. Investig Genet 2:1–14
76. Paneto GG, Koehnemann S, Martins JA, Cicarelli RM, Pfeiffer H (2011) A single multiplex PCR and SNaPshot minisequencing reaction of 42 SNPs to classify admixture populations into mitochondrial DNA haplogroups. Mitochondrion 11:296–302
77. Brandstätter A, Parsons TJ, Parson W (2003) Rapid screening of mtDNA coding region SNPs for the identification of West European Caucasian haplogroups. Int J Legal Med 117:291–8
78. Quintáns B, Alvarez-Iglesias V, Salas A, Phillips C, Lareu M, Carracedo A (2004) Typing of mitochondrial DNA coding region SNPs of forensic and anthropological interest using SNaPshot minisequencing. Forensic Sci Int 140:251–7
79. Grignani P, Peloso G, Achilli A et al (2006) Subtyping mtDNA haplogroup H by SNaPshot minisequencing and its application in forensic individual identification. Int J Legal Med 120:151–6
80. Köhnemann S, Sibbing U, Pfeiffer H, Hohoff C (2008) A rapid mtDNA assay of 22 SNPs in one multiplex reaction increases the power of forensic testing in European Caucasians. Int J Legal Med 122:517–23
81. Endicott P, Metspalu M, Stringer C, Macaulay V, Cooper A, Sanchez JJ (2006) Multiplexed SNP typing of ancient DNA clarifies the origin of Andaman mtDNA haplogroups amongst South Asian tribal populations. PLoS One 1:e81
82. Álvarez-Iglesias V, Jaime J, Carracedo A, Salas A (2007) Coding region mitochondrial DNA SNPs: targeting East Asian and Native American haplogroups. Forensic Sci Int Genet 1:44–55

83. Hu C-T, Yan J-W, Chen F et al (2015) Genetic analysis of 15 mtDNA SNP loci in Chinese Yi ethnic group using SNaPshot minisequencing. Gene
84. Coutinho A, Valverde G, Fehren-Schmitz L et al (2014) AmericaPlex26: a SNaPshot multiplex system for genotyping the main human mitochondrial founder lineages of the Americas. PLoS One 26:e93292
85. The Snipper 2.0: Binary AIM classification of individuals. University of Santiago de Compostela, Spain
86. Phillips C, Prieto L, Fondevila M et al (2009) Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS One 4:e6583
87. Phillips C, Fondevila M, Vallone PM et al (2011) Characterization of US population samples using a 34plex ancestry informative SNP multiplex. Forensic Sci Int Genet Suppl Ser 3:e182–e3
88. Daca-Roszak P, Pfeifer A, Żebracka-Gala J, Jarząb B, Witt M, Ziętkiewicz E (2016) EurEAs_Gplex—a new SNaPshot assay for continental population discrimination and gender identification. Forensic Sci Int Genet 20:89–100
89. de la Puente M, Santos C, Fondevila M et al (2016) The Global AIMs Nano set: a 31-plex SNaPshot assay of ancestry-informative SNPs. Forensic Sci Int Genet 22:81–8
90. Daniel R, Sanchez JJ, Nassif NT, Hernandez A, Walsh SJ (2009) Partial forensic validation of a 16plex SNP assay for the inference of biogeographical ancestry. Forensic Sci Int Genet Suppl Ser 2: 477–8
91. Kosoy R, Nassir R, Tian C et al (2009) Ancestry informative marker sets for determining continental origin and admixture proportions in common populations in America. Hum Mutat 30:69–78
92. Silva M, Zuccherato L, Soares-Souza G et al (2010) Development of two multiplex mini-sequencing panels of ancestry informative SNPs for studies in Latin Americans: an application to populations of the State of Minas Gerais (Brazil). Genet Mol Res 9:2069–85
93. Lins TC, Vieira RG, Abreu BS, Grattapaglia D, Pereira RW (2010) Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. Am J Hum Biol 22:187–92
94. Corach D, Lao O, Bobillo C et al (2010) Inferring continental ancestry of Argentineans from autosomal, Y-chromosomal and mitochondrial DNA. Ann Hum Genet 74:65–76
95. Walsh S, Liu F, Ballantyne KN, Mv O, Lao O, Kayser M (2011) IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet 5:170–80
96. Walsh S, Liu F, Wollstein A et al (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Sci Int Genet 7:98–115
97. Walsh S, Lindenbergh A, Zuniga SB et al (2011) Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet 5:464–71
98. Kastelic V, Pośpiech E, Draus-Barini J, Branicki W, Drobnič K (2013) Prediction of eye color in the Slovenian population using the IrisPlex SNPs. Croat Med J 54:381–6
99. Dembinski GM, Picard CJ (2014) Evaluation of the IrisPlex DNA-based eye color prediction assay in a United States population. Forensic Sci Int Genet 9:111–7. 101
100. Draus-Barini J, Walsh S, Pośpiech E et al (2013) Bona fide colour: DNA prediction of human eye and hair colour from ancient and contemporary skeletal remains. Investig Genet 4:1–15
101. Wurmbach E (2013) DNA assay development and validation for pigment-related features to assist in the identification of missing persons and human remains
102. Hart KL, Kimura SL, Mushailov V, Budimlija ZM, Prinz M, Wurmbach E (2013) Improved eye-and skin-color prediction based on 8 SNPs. Croat Med J 54:248–56
103. Ruiz Y, Phillips C, Gomez-Tato A et al (2013) Further development of forensic eye color predictive tests. Forensic Sci Int Genet 7:28–40
104. Kastelic V, Drobnič K (2012) A single-nucleotide polymorphism (SNP) multiplex system: the association of five SNPs with human eye and hair color in the Slovenian population and comparison using a Bayesian network and logistic regression model. Croat Med J 53:401–8
105. Bouakaze C, Keyser C, Crubezy E, Montagnon D, Ludes B (2009) Pigment phenotype and biogeographical ancestry form ancient skeletal remains: inferences from multiplexed autosomal SNP. Int J Leg Med 123:315–25
106. Bulbul O, Filoglu G, Altuncul H et al (2011) A SNP multiplex for the simultaneous prediction of biogeographic ancestry and pigmentation type. Forensic Sci Int Genet Suppl Ser 3:e500–e1
107. Gettings KB, Lai R, Johnson JL et al (2014) A 50-SNP assay for biogeographic ancestry and phenotype prediction in the US population. Forensic Sci Int Genet 8:101–8
108. Smith J, Godfrey H (2011) A SNaPshot™ assay for the identification of forensically important blowflies. Forensic Sci Int Genet Suppl Ser 3:e479–e80
109. Huang C-H, Chang M-T, Huang M-C, Lee F-L (2011) Application of the SNaPshot minisequencing assay to species identification in the Lactobacillus casei group. Mol Cell Probes 25:153–7
110. Kitpipit T, Tobe SS, Kitchener AC, Gill P, Linacre A (2012) The development and validation of a single SNaPshot multiplex for tiger species and subspecies identification—implications for forensic purposes. Forensic Sci Int Genet 6:250–7
111. Dario P, Oliveira A, Ribeiro T et al (2015) SNPforID 52-plex in casework samples: "cracking" bones and other difficult samples. Forensic Sci Int Genet Suppl Ser 5:e118–e20
112. Fondevila M, Phillips C, Naveran N et al (2008) Case report: identification of skeletal remains using short-amplicon marker analysis of severely degraded DNA extracted from a decomposed and charred femur. Forensic Sci Int Genet 2:212–8
113. Phillips C, Fondevila M, García-Magariños M et al (2008) Resolving relationship tests that show ambiguous STR results using autosomal SNPs as supplementary markers. Forensic Sci Int Genet 2:198–204
114. Pontes ML, Medeiros R (2015) Autosomal SNPs in different forensic applications. Aust J Forensic Sci 48:1–9
115. Phillips C (2015) Forensic genetic analysis of bio-geographical ancestry. Forensic Sci Int Genet 18:49–65
116. Pulker H, Lareu MV, Phillips C, Carracedo A (2007) Finding genes that underlie physical traits of forensic interest using genetic tools. Forensic Sci Int Genet 1:100–4
117. Butler K, Peck M, Hart J, Schanfield M, Podini D (2011) Molecular "eyewitness": forensic prediction of phenotype and ancestry. Forensic Sci Int Genet Suppl Ser 3:e498–e9
118. Titia Sijen NECW, Baca K, Ballard D, Balsa F, Bogus M, Borsting C, Brisighelli F, Cervenákova J, Chaitanya L, Decroyer V, Desmyter S, van der Gaag K, Gettings K, Haas C, Heinrich J, João Anjos M, Kal A, Kiesler K, Kúdelová A, Mosquera A, Noel F, Parson W, Pereira V, Phillips C, Schneider PM, Syndercombe-Court D, Turanska M, Vidaki A, Woliński P, Zatkalíková L (2016) A collaborative EDNAP exercise on the use of a SNaPshot™ tool for typing the mtDNA control region
119. McShane J (2011) The Night Stalker—the true story of Delroy Grant, Britain's most shocking serial sex attacker. Kindle Edition ed. John Blake
120. Patricia Ortega Dolz, Barroso FJ. Madrid teen's suspected murderer arrested in France 18 years after crime. ELPAIS
121. Severini S, Carnevali E, Margiotta G, Garcia-González M, Carracedo Á (2015) Use of ancestry-informative markers as a scientific tool to combat the illegal traffic in human kidneys. Forensic Sci Int Genet Suppl Ser 5:e302–e4

122. Santos C, Fondevila M, Ballard D et al (2015) Forensic ancestry analysis with two capillary electrophoresis ancestry informative marker (AIM) panels: results of a collaborative EDNAP exercise. Forensic Sci Int Genet 19:56–67
123. Walsh S, Chaitanya L, Clarisse L et al (2014) Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet 9:150–61
124. Hillmer AM, Brockschmidt FF, Hanneken S et al (2008) Susceptibility variants for male-pattern baldness on chromosome 20p11. Nat Genet 40:1279–81. doi:10.1038/ng.228
125. Medland SE, Nyholt DR, Painter JN et al (2009) Common variants in the trichohyalin gene are associated with straight hair in Europeans. Am J Hum Genet 85:750–5
126. Fagertun J, Wolffhechel K, Pers TH et al (2015) Predicting facial characteristics from complex polygenic variations. Forensic Sci Int Genet 19:263–8
127. Ho YY, Evans DM, Montgomery GW et al (2015) Genetic variant influence on whorls in fingerprint patterns. J Investig Dermatol
128. Bekaert B, Kamalandua A, Zapico S, Van de Voorde W, Decorte R (2015) A selective set of DNA-methylation markers for age determination of blood, teeth and buccal samples. Forensic Science International: Genetics Supplement Series
129. Phillips C, Parson W, Lundsberg B et al (2014) Building a forensic ancestry panel from the ground up: the EUROFORGEN Global AIM-SNP set. Forensic Sci Int Genet 11:13–25
130. Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis
131. Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina® Beta Version ForenSeq™ DNA Signature Prep Kit for use in genetic profiling. Forensic Sci Int Genet 20:20–9

Chapter 3: Low- and medium-throughput genotyping tools: High resolution melting and single base extension (SNaPshot™)

3.1.1 Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35 (21-22):3036–304

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 3.1.1

Declaration by candidate

In the case of Chapter 3.1.1, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Contributed in research conduction, data analysis and writing of a few sections of the published manuscript	20

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student UC Y/N
Samantha Venables	Proposal of research, experimental work, data analysis and writing of the entire manuscript	50	N
Runa Daniel	Revising and editing of the manuscript, critical feedback for the research plan and data analysis	10	N
Dennis McNevin	Revising and editing of the manuscript, critical feedback for the research plan and data analysis	10	N
Roland van Oorschot	Critical feedback on the manuscript	5	N
Simon Walsh	Critical feedback on the manuscript	5	N

Candidate's Signature	S M Mehta	Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	Dennis McN
9/4/2018	Roland van Oorschot	

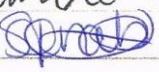
[* Please insert additional rows as needed.]

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	**Name of the co-author**	**Signature**
4/4/18	ALANA DANIEL	
6/4/2018	SAMANTHA VENABLES	

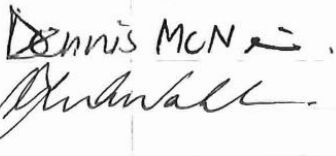
[* Please insert additional rows as needed.]

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	**Name of the co-author**	**Signature**
04/04/2018	Dennis McNevin	
06/04/2018	Simon WALSH	

[* Please insert additional rows as needed.]

Samantha J. Venables[1]
Bhavik Mehta[1]
Runa Daniel[2]
Simon J. Walsh[3]
Roland A. H. van Oorschot[2]
Dennis McNevin[1]

[1]National Centre for Forensic Studies, Faculty of Education, Science, Technology and Mathematics, University of Canberra, Canberra, Australia
[2]Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Australia
[3]Forensics, Australian Federal Police, Australia

Received February 19, 2014
Revised August 2, 2014
Accepted August 12, 2014

Research Article

Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework

High resolution melting (HRM) analysis is a simple, cost effective, closed tube SNP genotyping technique with high throughput potential. The effectiveness of HRM for forensic SNP genotyping was assessed with five commercially available HRM kits evaluated on the ViiA™ 7 Real Time PCR instrument. Four kits performed satisfactorily against forensically relevant criteria. One was further assessed to determine the sensitivity, reproducibility, and accuracy of HRM SNP genotyping. The manufacturer's protocol using 0.5 ng input DNA and 45 PCR cycles produced accurate and reproducible results for 17 of the 19 SNPs examined. Problematic SNPs had GC rich flanking regions which introduced additional melting domains into the melting curve (rs1800407) or included homozygotes that were difficult to distinguish reliably (rs16891982; a G to C SNP). A proof of concept multiplexing experiment revealed that multiplexing a small number of SNPs may be possible after further investigation. HRM enables genotyping of a number of SNPs in a large number of samples without extensive optimization. However, it requires more genomic DNA as template in comparison to SNaPshot®. Furthermore, suitably modifying pre-existing forensic intelligence SNP panels for HRM analysis may pose difficulties due to the properties of some SNPs.

Keywords:
Forensic DNA analysis / High resolution melting (HRM) analysis / SNP genotyping
DOI 10.1002/elps.201400089

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Significant research efforts in forensic genetics have been directed toward extending the use of DNA obtained from biological evidence beyond identification and toward the provision of intelligence for investigative purposes. This new era in forensic DNA analysis involves the use of DNA polymorphisms such as SNPs and provides a powerful compliment to traditional forensic identity testing [1–3].

SNP genotyping is beneficial in cases where traditional DNA profiling using conventional STR forensic identity markers is unsuccessful or uninformative, that is when DNA is degraded, no database matches are obtained, when there are no suspects or a large pool of suspects must be reduced. The ability to infer the phenotype of an unknown contributor, specifically biogeographical ancestry (BGA: e.g. [4–6]) and externally visible characteristics, such as hair color (e.g. [7]) and iris (eye) color (e.g. [8–10]), from their SNP genotype has immense value in forensic applications [1–3].

Forensic SNP assays developed for identification [11, 12] and predicting BGA [5, 6] and externally visible characteristics [7,9,10] commonly utilize the SNaPshot® assay (Life Technologies, Cat. # 4323159), a single base extension method [13]. The advantages of SNaPshot® for forensic SNP genotyping include its high multiplex capability (up to 30–40 SNPs) [5,11] and its reliance on instrumentation currently used in forensic laboratories. However, significant optimization is often required when multiplexing large numbers of SNPs [14, 15]. An additional limitation of SNaPshot® is that the workflow is labor and time intensive due to tube-to-tube transfers and purifications. This represents a higher contamination risk compared to other closed tube genotyping methods.

An alternative method is high resolution melting (HRM) analysis. Requiring only a generic, saturating dsDNA specific dye and two target specific primers per locus, HRM

Correspondence: Dr. Dennis McNevin, National Centre for Forensic Studies, Faculty of Education, Science, Technology and Mathematics, University of Canberra, Locked Bag 1, ACT 2601, Australia
E-mail: Dennis.McNevin@canberra.edu.au
Fax: +61-2-6201-2461

Abbreviations: BGA, biogeographical ancestry; **HRM**, high resolution melting

analysis enables the amplification and detection of the target sequence in a single closed tube assay [16–19]. Briefly, the dsDNA dye and SNP specific primers are included in the PCR reaction where the SNP target (40–100 bp) is amplified. The dsDNA dye interacts with the amplified SNP target and fluorescence is continually monitored as the PCR products are gradually heated from 60 to 95°C. As the PCR product transitions from the double stranded to single stranded conformation, a rapid loss of fluorescence is observed. The temperature at which 50% of the DNA exists in the single stranded conformation is termed the apparent "melting temperature" (T_m) [16, 20, 21]. This melting temperature and the morphology of the PCR amplicon melting profile are used to identify the presence of sequence variants. The melting profile is dependent on GC content, GC distribution, length, and sequence of the PCR product [16, 22]. HRM analysis is more sensitive for smaller fragments (e.g. 40–100 bp for SNP genotyping) as increasing the size of PCR products can reduce temperature differences between alternate genotypes and thus increase difficulty in accurately distinguishing between different variants [17, 23–26].

Most SNPs are easily genotyped using HRM analysis, as variations in the melting temperatures and melting profiles of the three possible genotypes allow the homozygous and heterozygous variants to be definitively identified. Alternate homozygotes usually vary by approximately 1°C, while heterozygotes display an intermediate melting temperature and an altered melting profile [19]. Homozygote profiles display sharp symmetric melting transitions, with the alternate homozygote genotypes differing by their melting temperature [27]. The melting profile of a heterozygote differs in shape from a homozygote in that it generally contains two melting domains. Prior to melting, a PCR product with a heterozygote genotype forms four duplexes, two homoduplexes, and two mismatched heteroduplexes. Since each duplex has a characteristic melting temperature, the sum of all melting transitions is observed in HRM analysis [28], which alters the shape of the melting profile of a heterozygote compared to the homozygotes [17].

The melting profiles of SNP variants can be represented in two ways: (i) a difference curve plots the difference in fluorescence (relative to an arbitrary reference variant) for each amplicon at temperature increments; (ii) the negative second derivative of the fluorescence with respect to temperature ($-dF/dT$) has a peak at the point of inflection where 50% of the amplicons consist of dsDNA (T_m). The melting profile may display more than one melting domain (hence, more than one T_m), particularly when the SNP genotype is heterozygous.

A major advantage of HRM analysis for SNP genotyping is that the acquisition of the melting profile occurs within the instrument immediately following PCR [16]. The elimination of post-PCR sample manipulation reduces the time and labor required for genotyping and reduces the risk of carryover PCR contamination [29]. Additionally, the cost of HRM analysis is competitive, with HRM master mixes retailing for USD $0.63 to $0.90 per reaction (at the time of writing). The major limitation of HRM analysis for SNP genotyping is the limited multiplex capability inherent in the method given that product detection occurs in a limited temperature range (60–95°C). However, Seipp et al. [30] developed a HRM multiplex which reliably genotyped four SNPs simultaneously by tailing PCR primers to artificially separate the melting temperatures of the PCR products.

The aim of this study was to assess the applicability of HRM analysis to SNP genotyping in forensic applications. After initial assessment of five commercial HRM kits, the best performing kit was further assessed by forensically relevant criteria. A large sample set was genotyped using 12 SNPs to assess intra- and inter-plate temperature variability and genotype variant calling concordance. Multiplexing capability was investigated using two SNPs.

2 Materials and methods

2.1 Collection of DNA samples

Buccal (inner cheek epithelial) cells were collected from volunteers via cotton-tipped buccal swabs (Interpath, Cat # 8150CIS) with ethics approval from the University of Canberra Committee for Ethics in Human Research (Project numbers 09–127 and 11–119).

2.2 SNP selection and primer design

A literature search identified 13 SNPs for a population study to determine BGA [31]. Symmetrical (G/C and A/T) SNPs were avoided as the difference in melting temperatures for the different homozygous genotypes is expected to be less than 0.4°C [17]. Three of these 13 SNPs were used in the initial HRM assessment and optimization, as were the six phenotypically informative SNPs in the "IrisPlex" assay [9, 10] (Supporting Information Table 1).

Primers were designed using Primer 3 [32] to produce short (35–60 bp) PCR products (Supporting Information Table 1) and so that the SNP position was the only base of the PCR product not included in the primer binding regions, such that: Amplicon Length (bp) = Length of Forward Primer + Length of Reverse Primer + 1

Primers were assessed for the formation of secondary structures such as hairpins and primer dimer using AutoDimer [33] and target specificity using Primer-BLAST [34].

2.3 SNP variant assignment

SNP variants were identified by the ViiA 7 software v1.2 using default temperature windows (premelt (dsDNA) and postmelt (ssDNA)) for data normalization. Identification of individual genotypes for biallelic SNPs (without using genotype controls) was conducted by examination of the melting temperatures and the morphology of the melting profile

$(-dF/dt)$. The majority of SNPs assessed in this study produced simple melting profiles that were straightforward to genotype. In these cases, the heterozygote genotype was indicated by a second melting domain (peak) in the melting profile and homozygote genotypes were distinguished from each other on the basis of their melting temperature, since a GC pair in dsDNA requires more thermal energy to destabilize than an AT pair.

2.4 HRM assessment and optimization

2.4.1 Assessment of HRM kit performance

SNP rs733559 was selected for the initial assessment of HRM kit sensitivity and performance. A phenol-chloroform/ethanol precipitation method (as described in [35] except that DTT was omitted) was used to extract DNA from buccal swabs from four individuals. DNA was then quantified using the Quantifiler® Human DNA Quantification Kit (Life Technologies) following the manufacturer's protocol. PCR and HRM were performed on these samples and control DNA 9947A (Promega) according to the manufacturer's protocols, except that 45 PCR cycles were used for each kit (Supporting Information Table 2). The five HRM kits were tested using DNA template amounts of 0.1, 0.2, and 0.5 ng. Where permitted, two further analyses were performed. First, DNA template at a concentration of 0.1 ng/μL was added to maximum volume permitted by each kit (up to 1.0 ng DNA/reaction for the SensiMix kit). Second, 1 μL of DNA template at a much higher concentration (5–10 ng/μL depending on the sample) was also tested. These additional analyses allowed the amplification of increased amounts of DNA template, ensuring that each kit was tested within the recommended DNA template range specified by the manufacturer (Supporting Information Table 2).

HRM kits were assessed for suitability to forensic applications against the following criteria: successful SNP amplification using a minimal amount of input DNA; observable differentiation of genotypes to distinguish variants; and minimal variation between the melting temperature and melting curve morphology for different individuals with the same apparent genotype. The MeltDoctor™ HRM Master Mix (Life Technologies Cat #4415440) best satisfied these criteria and was therefore further evaluated (Supporting Information Table 3).

2.4.2 The effect of extraction method on MeltDoctor SNP genotyping

DNA was extracted from buccal samples from five individuals using three different extraction methods (phenol-chloroform with ethanol precipitation [35] as described earlier, QIAamp DNA mini kit (Qiagen), and Isolate II (Bioline) according to manufacturer's protocols). DNA was then quantified using the Quantifiler® Human DNA Quantification Kit (Life Technologies) following the manufacturer's protocol. The DNA extracts were diluted appropriately and subjected to HRM SNP genotyping with the MeltDoctor HRM Master Mix using the recommended protocol (with 0.5 ng of input DNA and 45 PCR cycles) as shown in Supporting Information Table 2a.

2.4.3 Sensitivity and reproducibility of MeltDoctor SNP genotyping

Further assessment of MeltDoctor HRM Master Mix sensitivity was performed on DNA from five individuals using the six SNPs contained in the "IrisPlex" assay developed by Walsh et al. [9, 10]. DNA input amounts of 10, 1, 0.5, and 0.1 ng were tested using the protocol outlined in Supporting Information Table 2a, extending the work recently published by Mehta et al. [36]. To assess the reproducibility of this protocol, 11 samples were genotyped for SNP rs733559 in triplicate for each of three DNA input amounts (0.1, 0.2, and 0.3 ng) according to the manufacturer's instructions with 45 PCR cycles (Supporting Information Table 2a).

2.4.4 SNP genotyping accuracy

DNA sequencing of the HRM PCR products was not possible due to the small (<50 bp) fragment size and the position of the SNPs adjacent to the primer binding sites. Electropherogram quality in the vicinity of primers is always poor, making base calling unreliable (e.g. see, http://www.udel.edu/dnasequence/Site/Interpreting_Electropherograms.html). Thus, restriction enzyme digestion of the HRM PCR product and electrophoretic separation of DNA fragments was used to confirm the accuracy of the HRM genotyping method. NEBcutter v2.0 [37] was used to assess whether the SNPs under study created a recognition site for a restriction enzyme. Specifically, the presence of one SNP allele should create a recognition site while the presence of the other SNP allele does not, allowing for elucidation of the SNP genotype from the banding pattern observed. Of the SNPs assessed, only three SNPs met the above criteria: rs733559 (enzyme *RsaI* with the restriction site GTAC; the presence of the C allele enables digestion of the PCR product); and rs310850 and rs3892905 (enzyme *MluCI* with the restriction site AATT; the presence of the A allele enables digestion of the PCR product). These three SNPs were used to confirm that the HRM analysis using the ViiA 7™ instrument produced accurate genotypes.

The 11 DNA samples used for analysis of SNP rs733559 during the reproducibility testing (Section 2.4.3) and 9947A control DNA were also subjected to HRM genotyping in duplicate using rs310850 and rs3892905. The PCR reactions and thermal cycling protocols were performed as listed in Supporting Information Table 2a with 0.3 ng DNA template. Restriction digest reactions were then conducted using 15 μL of HRM PCR product and 5U of restriction enzyme according to the manufacturer's recommended protocol. *RsaI* reactions were left to digest at 37°C for 3 h and inactivated

at 65°C for 20 min, while *MluCI* reactions were digested for 1 h at 37°C and were inactivated by heating at 80°C for 20 min. The digestion products and undigested controls were electrophoresed on the 2100 Bioanalyser with the DNA 1000 kit (Agilent Technologies), following the manufacturer's recommended protocol.

2.5 Extended trial of the optimized HRM method

Genotyping using 12 SNPs (Supporting Information Table 1) was conducted on DNA obtained from 1284 individuals using the optimized method described above. Each HRM analysis included three negative (no template) controls and three positive genotype controls. The position of the negative controls varied periodically to monitor contamination across the plate. The positive controls contained DNA representing the three possible genotypes for SNPs rs310850 or rs3892905 to indicate that the PCR and HRM analysis produced the expected genotypes from these samples with each run. The final reaction volume in each well was 20 μL.

2.5.1 Assessing intra-run and inter-run temperature variability

One-way ANOVAs were performed using IBM® SPSS® Statistics v19 [38] to assess the melting temperature variations within a single run (96-well plate) and between multiple runs.

3 Results

3.1 HRM assessment and optimization

3.1.1 HRM kit performance

Four of the five HRM kits successfully amplified SNP rs733559 across the entire range of DNA template amounts were investigated (Fig. 1). The fifth (SensiMix™ HRM kit) did not amplify rs733559 from any of the samples at any of the DNA template amounts tested.

Overall, the melting peaks detected for each of the samples group together regardless of DNA template amount. However, with two of the kits, KAPA™ HRM Fast PCR kit and SensiFast™ HRM kit, extraneous peaks appeared in the melting profiles when using lower DNA template input amounts in the reaction (Fig. 1). The additional peaks in the melting profile produced using the SensiFast™ HRM kit caused the morphology of the homozygote genotype (sample 1) to approximate the heterozygous genotype (sample 2), although a temperature shift was observed in sample 1 relative to sample 2. Given that extraneous peaks in the melting profile could possibly affect genotype determination, these kits were not considered further.

The remaining kits, MeltDoctor™ HRM MasterMix and Precision Melt Supermix, both performed similarly with minimal variation in the melting profiles for different individuals with the same genotype (Supporting Information Table 3). The principal difference between these two kits was that the MeltDoctor™ HRM MasterMix produced a larger difference between the melting temperatures of the two observed genotypes (Fig. 1) resulting in more accurate genotype calls. Based on these results, it was determined that the MeltDoctor™ HRM MasterMix was the most suitable HRM kit for SNP genotyping in this study.

Despite differences in the melting temperatures recorded for each genotype between different HRM chemistries, classification of the genotypes for SNP rs733559 were the same. That is, where one kit recognized a sample as a particular genotypic variant, the other kits were in concordance.

3.1.2 Effect of DNA extraction methods

HRM genotyping was not greatly affected by DNA extraction method (Supporting Information Fig. 1a–c). Amplification was successful for all methods assessed and the same genotype variant was detected for all SNPs tested. The peak $-\mathrm{d}F/\mathrm{d}T$ appeared to be consistently lower for the phenol-chloroform method, indicating a transition between dsDNA and ssDNA that is less well defined than for the other methods, although this did not impact on the genotype assignment.

Two SNPs (rs1800407 and rs16891982) proved difficult to genotype consistently. The melting profile produced for SNP rs1800407 has more melting domains (peaks) than expected from in silico prediction tools such as uMelt HETS [39, 40] (Supporting Information Fig. 2). Furthermore, the alternate homozygote genotypes of symmetrical SNP rs16891982 (i.e. GG and CC) share very similar melting temperatures (data not shown).

3.1.3 Sensitivity and reproducibility

The sensitivity testing conducted on the six "IrisPlex" SNPs [9, 10] showed that the MeltDoctor™ HRM Master Mix detected DNA input amounts of 0.1 ng for all six SNPs, although genotype calling was inconsistent at this DNA input amount. As expected, increased DNA input amounts were associated with higher values for the peak $-\mathrm{d}F/\mathrm{d}T$ (data not shown).

The amount of DNA template required for reproducible SNP genotyping results was then determined using triplicate reactions of various DNA input amounts. Nonconcordant genotypes were observed in the triplicate reactions for two of the 11 samples tested when using 0.1 ng DNA template. Nonconcordant genotypes were only observed in a single sample when 0.2 ng DNA template was analyzed. Genotypes of all 11 samples were fully concordant across the triplicate reactions using 0.3 ng DNA template (data not shown), which was deemed to be the minimum amount of DNA template required for genotyping.

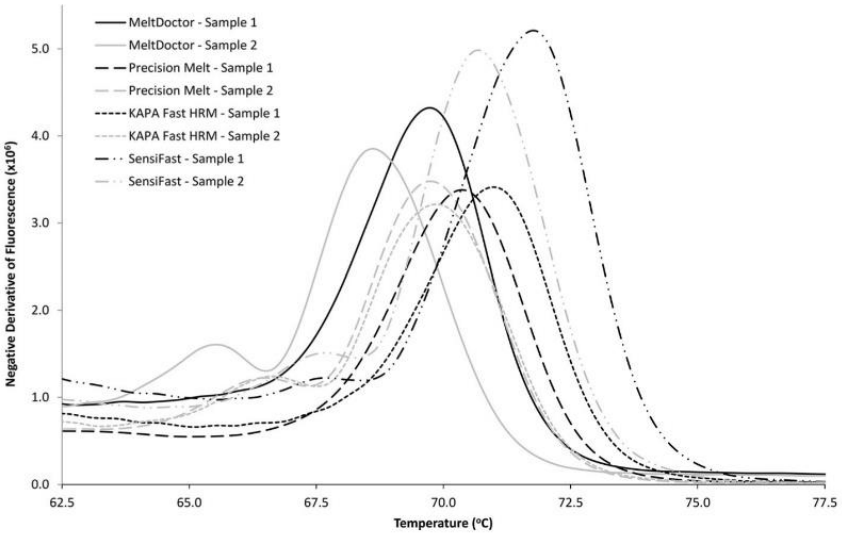


Figure 1. Kit comparison results for the 0.5 ng DNA template amount (rs733559). Sample 1 has a homozygous genotype while sample 2 has a heterozygous genotype. Results are not shown for the SensiMix HRM kit which failed to amplify the SNP target.

3.1.4 Accuracy

Full concordance was observed between the HRM genotypes and the restriction enzyme genotypes for all three SNPs assessed (rs733559, rs310850, rs3892905), thus providing confirmation that the HRM SNP genotyping method produces accurate SNP genotypes (Supporting Information Table 4). To illustrate this, SNP rs733559 genotyping results for four samples are shown from the HRM method (Fig. 2) and the restriction digest confirmation method (Fig. 3). Samples 2, 3, and 4 are heterozygotes (CT), while sample 1 is a CC homozygote. Figure 2 shows that samples 2–4 display two melting domains (peaks) and sample 1 has a single peak at a slightly elevated melting temperature (further right on the horizontal-axis) as expected. Meanwhile, in Fig. 3., sample 1 displays one short band (approx. 20 bp) as the presence of the C allele in both DNA strands has enabled the restriction enzyme to digest all of the PCR product. Samples 2–4 display two bands, one at approximately 50 bp, which indicates the presence of a T allele (which the restriction enzyme cannot digest) and one band approximately 20 bp in length representing the C allele.

3.2 Extended trial of the optimized HRM method

3.2.1 Intrarun melting temperature variability

A one-way ANOVA was performed to assess the temperature variation within the HRM genotyping runs for each of the 12 SNPs tested in the population samples. The null hypothesis was that there is no difference between the mean melting temperatures observed for the three possible genotypes for a particular SNP within a 96-well plate. ANOVA results showed that the null hypothesis was almost always rejected (153 of 155

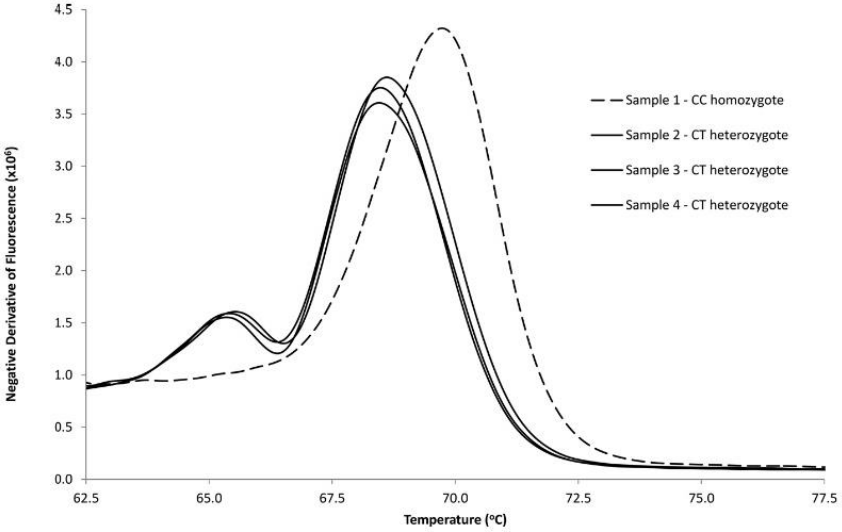


Figure 2. HRM genotyping results for SNP rs733559 in four samples analysed during the assessment and optimization experiments. Sample 1 is a "CC" homozygote, while samples 2–4 are "CT" heterozygotes.

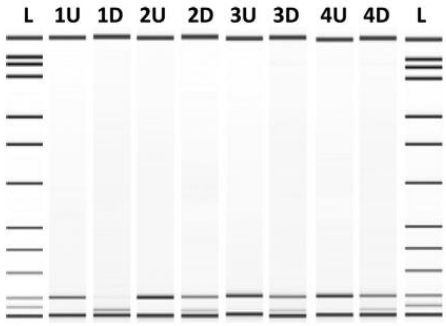


Figure 3. Restriction enzyme digest confirmed the genotypes of SNP rs733559 in the same four samples. Each sample is shown in undigested (U) and digested (D) forms. The ladder contains the following size bands (bp): 1500, 1000, 850, 700, 500, 400, 300, 200, 150, 100, 50, 25, and 15. Bands at 15 bp and 1500 bp are system markers present in each lane. Sample 1 is a "CC" homozygote and samples 2–4 are "CT" heterozygotes.

comparisons; p-value = 0.000). Therefore, the melting temperatures of different genotypes were significantly different except when SNPs rs10494531 and 10496623 were genotyped on one plate, however these genotypes were resolved by the software.

3.2.2 Inter-run melting temperature variability

A one-way ANOVA was performed to assess the temperature variation observed for each of the possible genotypes between the HRM genotyping runs for each of the 12 SNPs tested in the population samples. The null hypothesis was that there is no difference between the mean melting temperatures observed for a particular SNP genotype across all of the HRM genotyping runs. In most cases, the null hypothesis was rejected (p-value <0.05) and therefore the melting temperatures for a particular genotype were significantly different between HRM genotyping runs. The three exceptions were rs10494531 G/T, rs2185785 G/G, and rs10496623 T/T.

3.2.3 Genotype variant calling concordance

Genotype variant calling was compared within a single run (where multiple samples were analyzed numerous times) and between multiple runs (using the PCR positive control results) and found to be fully concordant when using the optimized HRM protocol (0.5 ng input DNA and 45 PCR cycles).

3.3 Multiplex capability

Two SNPs, rs963170 (A/C) and rs10494531 (G/T), were identified as having melting temperatures that differ by approximately 5°C. These two SNPs were combined into a duplex that was tested (without further optimization) on three samples known to have different genotype combinations (sample 5: AA and GG with two expected melting domains; sample 6: AA and GT with three expected melting domains; and sample 7: AC and GG with three expected melting domains) in a proof of concept experiment to assess the multiplex capability of the HRM SNP genotyping method.

The duplex produced the expected results for two of the samples (5 and 7). However, sample 6 displayed two melting domains instead of the predicted three, due to the similarity in melting temperature of the rs963170 homozygote genotype and the secondary melting domain in the rs10494531 heterozygote genotype. However, using the default settings the ViiA 7™ Software v1.2 was able to distinguish the genotype variants produced by the duplex analysis for all three samples despite their visual similarity (data not shown).

4 Discussion

In addition to being sensitive, reproducible, and accurate, HRM has a number of other characteristics which confer suitability as a forensic SNP genotyping method. HRM SNP genotyping is rapid, with amplification and genotyping occurring in a single reaction in approximately 2 h. The closed tube nature of the method minimizes the risk of sample contamination [29]. HRM SNP genotyping is also amenable to high throughput analysis depending on instrument specifications. For example, the ViiA 7™ Real-Time PCR system used in this study has interchangeable 96- and 384-well blocks. Another advantage of HRM SNP genotyping is that minimal optimization is required to apply the method to different SNPs, particularly if primers were specifically designed to enable differentiation of different melting profiles and will anneal to the DNA template at the same temperature. All SNPs analyzed in this study were successfully amplified using the same protocol (Supporting Information Table 2a).

Some major limitations exist with this methodology for forensic applications. Successful HRM SNP genotyping is heavily dependent on a number of factors including the absence of complex melting transitions and unexpected melting domains (related to the GC content and distribution in the amplicon); difficulties associated with the differentiation of genotypes of symmetrical (G/C and A/T) SNPs; and, the availability of positive HRM genotype controls. The limited multiplex capability of HRM may result in significant depletion of evidentiary samples. Therefore, HRM may be more suited to genotyping population samples. Furthermore, HRM analysis also requires high resolution real time PCR instruments which may not be utilized in all forensic laboratories.

Two of the SNPs analyzed from the "IrisPlex" panel proved difficult to genotype. First, the alternate homozygote genotypes for the symmetrical SNP, rs16891982 (G/C), were difficult to distinguish from each other based on temperature. Second, the GC rich flanking regions of SNP rs1800407 produced multiple melting domains which caused genotyping complications without the presence of samples of known genotype for this SNP included in the run. There are a number of useful in silico tools available for the prediction of high-resolution melting curves (e.g. uMelt [39] and uMelt HETS [40]). However, we note that while uMelt HETS [40]

predicted small temperature differences for the alternate genotypes of the symmetrical SNP rs16891982, it was unable to predict the complex melting transitions that we observed for SNP rs1800407. Therefore, SNPs with these characteristics could be avoided during SNP selection when designing HRM SNP assays (e.g. the extended trial reported here, which avoided G/C and A/T SNPs). However, when restricted to the use of an existing panel of SNPs such as "IrisPlex" [9] or ancestry informative assays, this presents a real limitation of HRM SNP genotyping. SNP rs1800407 is the second best predictor of eye color in the "IrisPlex" panel [9] and simply excluding the SNP from a HRM assay would significantly reduce the forensic informativeness of the phenotype prediction.

The inclusion of positive controls for each genotype and each SNP is recommended to allow the unequivocal identification of genotypes. This is particularly important where a single homozygous genotype may be observed, or the melting temperatures of the alternate homozygotes are very similar, as it may be difficult to identify which homozygous genotype is present based on melting temperature alone (especially given the inter-run temperature variability shown in Section 3.2.2 which renders melting temperature alone as unreliable for genotyping purposes). Such controls may have to be synthesized or purchased as oligonucleotides. The inclusion of genotype controls may become particularly important for reliable genotyping if casework samples are to be processed using a HRM SNP multiplex.

The limited capacity for analyzing multiple SNPs simultaneously resulted in the depletion of a significant amount of DNA template (approximately 7 ng) to produce genotypes for the 12 SNPs analyzed in the extended trial of the optimized HRM method. In some forensic casework situations, such as the analysis of trace DNA, this amount of DNA template would not be available for use. The limited multiplex capacity of HRM arises because detection of the SNP targets must occur between 60°C and 95°C. The proof of concept experiment presented here (Section 3.3) demonstrated that even with melting temperature differences of 5°C between two SNPs, it can be difficult to discern alternate duplex genotypes. Further investigation into the development of HRM SNP genotyping multiplexes for forensic use is warranted, and of particular interest would be the empirical testing of artificially lengthened (tailed) PCR primer sets to exaggerate melting temperature differences between amplicons. Using this approach, Seipp et al. [30] demonstrated that four SNPs can be reliably genotyped using HRM.

In conclusion, results from this study indicate that HRM SNP genotyping is an appropriate method for the high throughput genotyping of numerous SNPs in a large number of population samples, particularly where ample DNA template is available. However, a more extensive validation including, but not limited to, mixture studies and the analysis of trace and highly degraded DNA samples, would be required prior to considering the use of HRM in forensic SNP analysis.

Samples for the population study were provided to S.J. Walsh (for analysis by S.J. Venables using SNPs chosen by S.J. Venables, D. McNevin and R. Daniel) by Professor Herawati Sudoyo from the Eijkman Institute for Medical Research, Jakarta, Indonesia. The authors gratefully acknowledge funding from the Australian Research Council (Linkage Project LP110100121) and for Australian Postgraduate Award stipends (to S.J. Venables and B. Mehta). The Australian Federal Police provided funding for consumables.

The authors have declared no conflict of interest.

5 References

[1] Gill, P., Werrett, D. J., Budowle, B., Guerrieri, R., *Sci. Justice* 2004, *44*, 51–53.

[2] Daniel, R., Walsh, S. J., *Aust. J. Forensic Sci.* 2006, *38*, 59–74.

[3] Butler, J. M., Coble, M. D., Vallone, P. M., *Forensic Sci. Med. Pathol.* 2007, *3*, 200–205.

[4] Frudakis, T., Venkateswarlu, K., Thomas, M., Gaskin, Z., Ginjupalli, S., Gunturi, S., Ponnuswamy, V., Natarajan, S., Nachimuthu, P. K., *J. Forensic Sci.* 2003, *48*, 771–782.

[5] Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., Carracedo, Á., *Forensic Sci. Int. Genet.*, 2007, *1*, 273–280.

[6] Phillips, C., Aradas, A. F., Kriegel, A., Fondevila, M., Bulbul, O., Santos, C., Rech, F. S., Carceles, M., Carracedo, Á., Schneider, P., *Forensic Sci. Int. Genet.* 2013, *7*, 359–366.

[7] Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., Kayser, M., *Forensic Sci. Int. Genet.* 2013, *7*, 98–115.

[8] Frudakis, T., Thomas, M., Gaskin, Z., Venkateswarlu, K., Chandra, K. S., Ginjupalli, S., Gunturi, S., Natrajan, S., Ponnuswamy, V. K., Ponnuswamy, K., *Genetics* 2003, *165*, 2071–2083.

[9] Walsh, S., Liu, F., Ballantyne, K. N., van Oven, M., Lao, O., Kayser, M., *Forensic Sci. Int. Genet.* 2011, *5*, 170–180.

[10] Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijen, T., de Knijff, P., Kayser, M., Ballantyne, K. N., *Forensic Sci. Int. Genet.* 2011, *5*, 464–471.

[11] Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., Morling, N., *Electrophoresis* 2006, *27*, 1713–1724.

[12] Kidd, K. K., Pakstis, A. J., Speed, W. C., Grigorenko, E. L., Kajuna, S. L., Karoma, N. J., Kungulilo, S., Kim, J. J., Lu, R. B., Odunsi, A., Okonofua, F., Parnas, J., Schulz, L. O., Zhukova, O. V., Kidd, J. R., *Forensic Sci. Int. Genet.* 2006, *164*, 20–32.

[13] Tully, G., Sullivan, K. M., Nixon, P., Stones, R. E., Gill, P., *Genomics* 1996, *34*, 107–113.

[14] Sobrino, B., Brion, M., Carracedo, A., *Forensic Sci. Int. Genet.* 2005, *154*, 181–194.

[15] Vallone, P. M., Butler, J. M., *J. Forensic Sci.* 2004, *49*, 723–732.

[16] Ririe, K. M., Rasmussen, R. P., Wittwer, C. T., *Anal. Biochem.* 1997, *245*, 154–160.

[17] Liew, M., Pryor, R., Palais, R., Meadows, C., Erali, M., Lyon, E., Wittwer, C., *Clin. Chem.* 2004, *50*, 1156–1164.

[18] Seipp, M., Durtschi, J., Voelkerding, K., Wittwer, C., *J. Bio. Mol. Tech.* 2009, *20*, 160–164.

[19] Montgomery, J. L., Sanford, L. N., Wittwer, C. T., *Expert Rev. Mol. Diagn.* 2010, *10*, 219–240.

[20] Akey, J. M., Sosnoski, D., Parra, E., Dios, S., Hiester, K., Su, B., Bonilla, C., Jin, L., Shriver, M. D., *Biotechniques* 2001, *30*, 358–362, 364, 366–367.

[21] Giglio, S., Monis, P. T., Saint, C. P., *Nucleic Acids Res.* 2003, *31*, e136.

[22] Dobrowolski, S. F., Gray, J., Miller, T., Sears, M., *Hum. Mutat.* 2009, *30*, 891–898.

[23] Wittwer, C. T., Reed, G. H., Gundry, C. N., Vandersteen, J. G., Pryor, R. J., *Clin. Chem.* 2003, *49*, 853–860.

[24] Erali, M., Voelkerding, K. V., Wittwer, C. T., *Exp. Mol. Pathol.* 2008, *85*, 50–58.

[25] Gundry, C. N., Dobrowolski, S. F., Martin, Y. R., Robbins, T. C., Nay, L. M., Boyd, N., Coyne, T., Wall, M. D., Wittwer, C. T., Teng, D. H. F., *Nucleic Acids Res.* 2008, *36*, 3401–3408.

[26] Vossen, R. H. A. M., Aten, E., Roos, A., den Dunnen, J. T., *Hum. Mutat.* 2009, *30*, 860–866.

[27] Seipp, M. T., Durtschi, J. D., Liew, M. A., Williams, J., Damjanovich, K., Pont-Kingdon, G., Lyon, E., Voelkerding, K. V., Wittwer, C. T., *J. Mol. Diagn.* 2007, *9*, 284–289.

[28] Reed, G. H., Wittwer, C. T., *Clin. Chem.* 2004, *50*, 1748–1754.

[29] Gundry, C. N., Vandersteen, J. G., Reed, G. H., Pryor, R. J., Chen, J., Wittwer, C. T., *Clin. Chem.* 2003, *49*, 396–406.

[30] Seipp, M. T., Pattison, D., Durtschi, J. D., Jama, M., Voelkerding, K. V., Wittwer, C. T., *Clin. Chem.* 2008, *54*, 108–115.

[31] Venables, S. J., *A Population Genetic Analysis of Forensic Identity and Ancestry Markers in Indonesia*. PhD Thesis, University of Canberra 2012, p. 239.

[32] Rozen, S., Skaletsky, H. J., in: Misener, S. and Krawetz, S. A., (Eds.), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*, Humana Press, Totowa, NJ 2000, pp. 365–386.

[33] Vallone, P. M., Butler, J. M., *Biotechniques* 2004, *37*, 226–231.

[34] Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T., *BMC Bioinformatics* 2012, *13*, 134.

[35] Allen-Hall, A., McNevin, D., *Forensic Sci. Int. Genet.* 2012, *6*, 653–657.

[36] Mehta, B., Daniel, R., McNevin, D., *Forensic Sci. Int. Genet.* 2013, *4*, e376–e377.

[37] Vincze, T., Posfai, J., Roberts, R. J., *Nucleic Acids Res.* 2003, *31*, 3688–3691.

[38] IBM Corporation, *IBM SPSS Statistics for Windows, Version 19*, 2010, IBM Corporation, Armonk, New York.

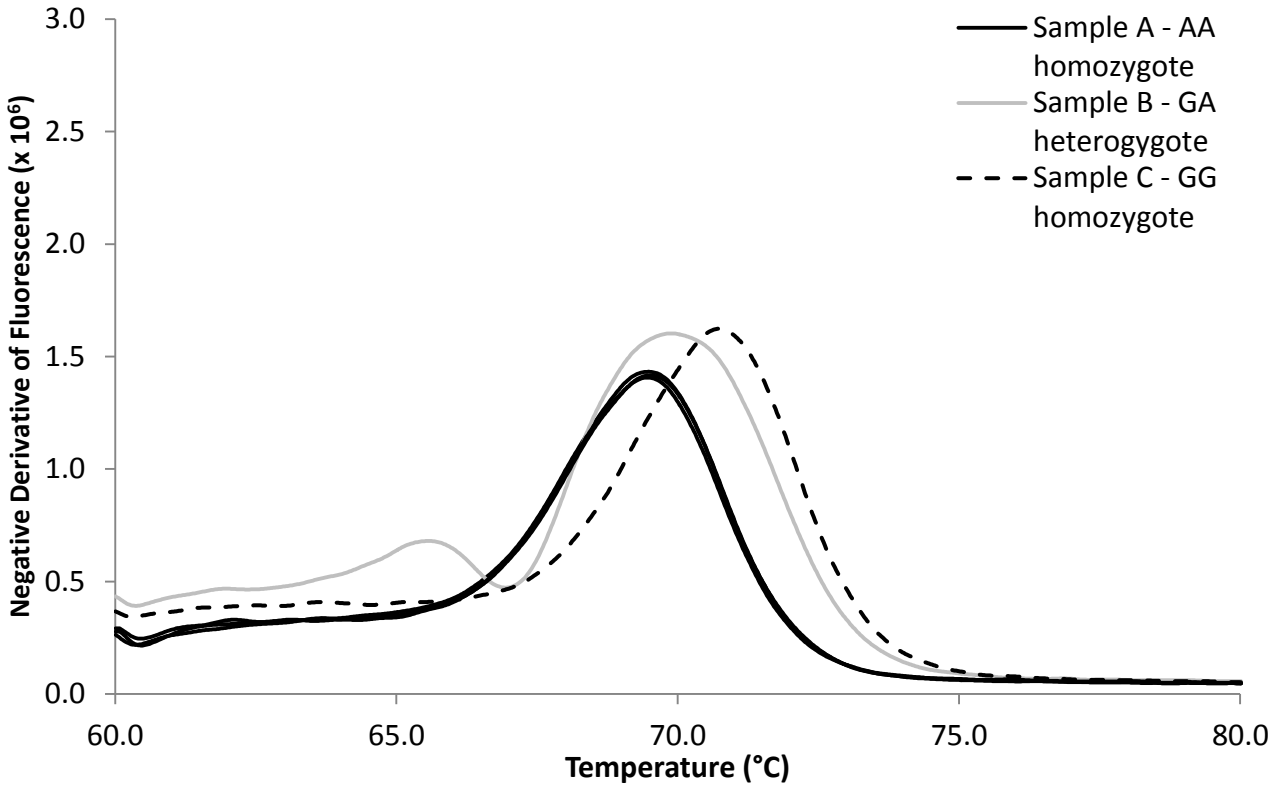
[39] Dwight, Z., Palais, R., Wittwer, C. T., *Bioinformatics* 2011, *27*, 1019–1020.

[40] Dwight, Z. L., Palais, R., Kent, J., Wittwer, C. T., *Hum. Mutat.* 2014, *35*, 278–282.

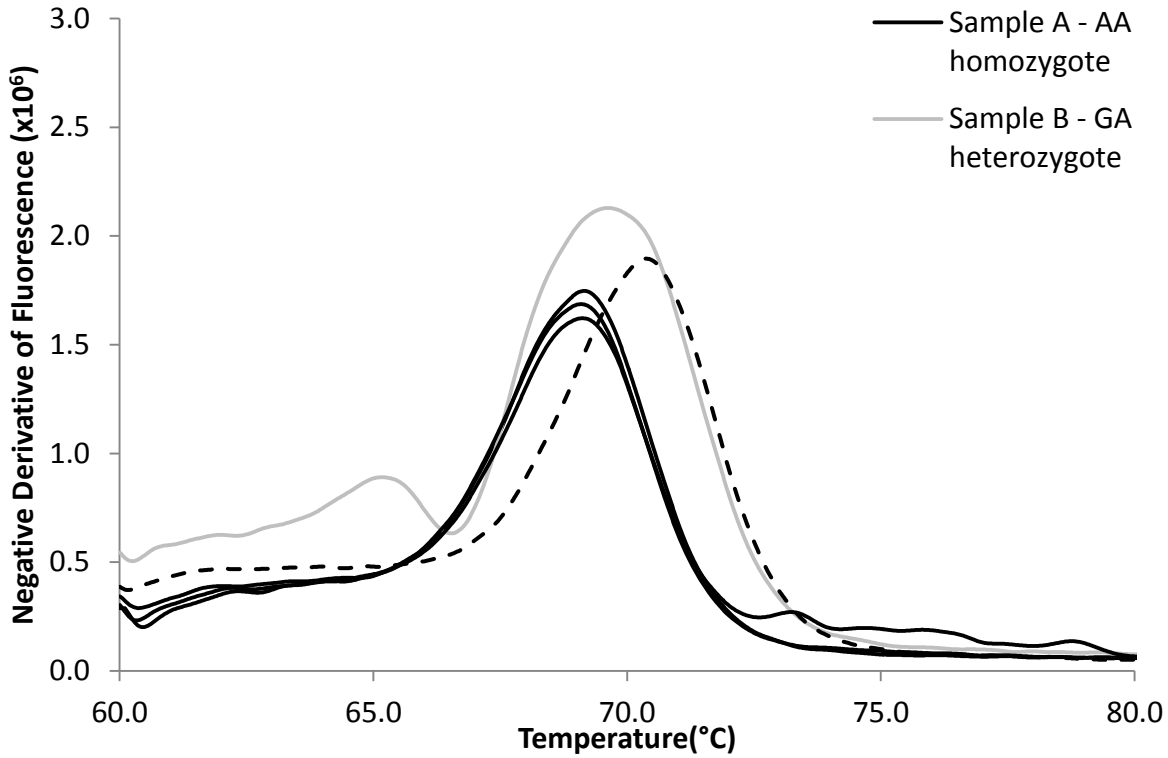
Supplementary figures

Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35 (21–22):3036–3043

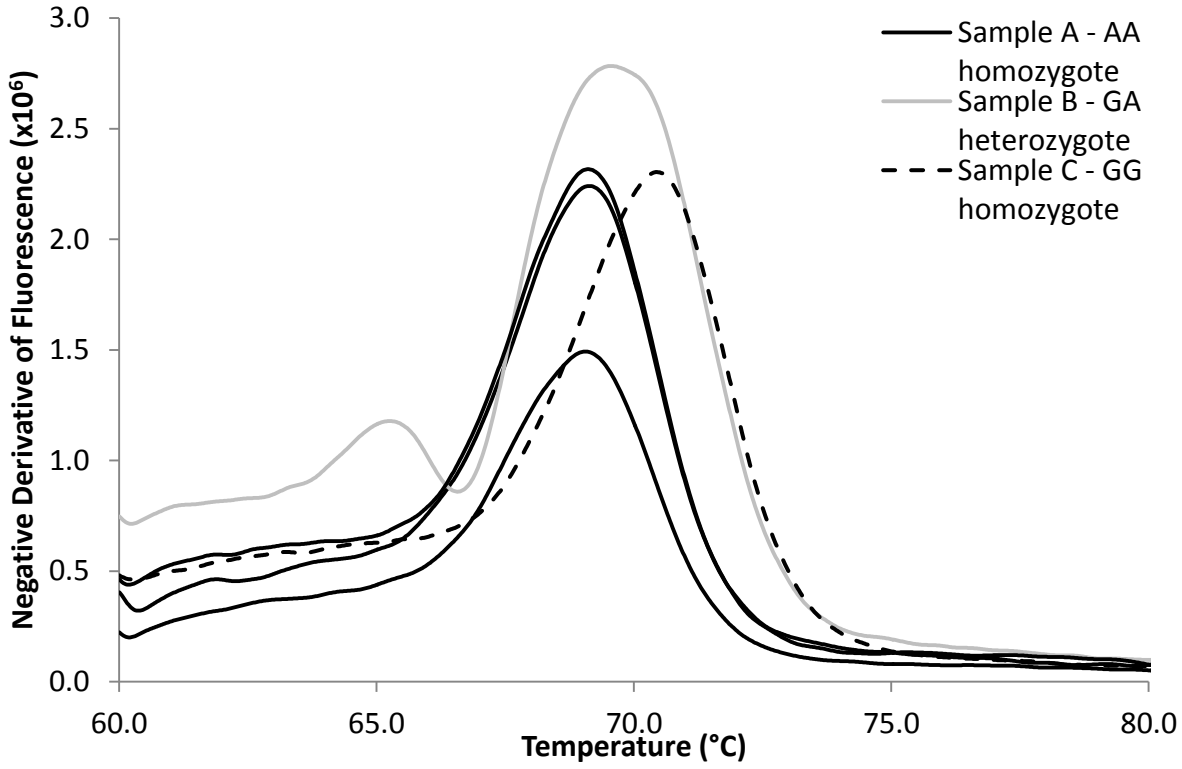
Supp. Info. Figure 1 a)



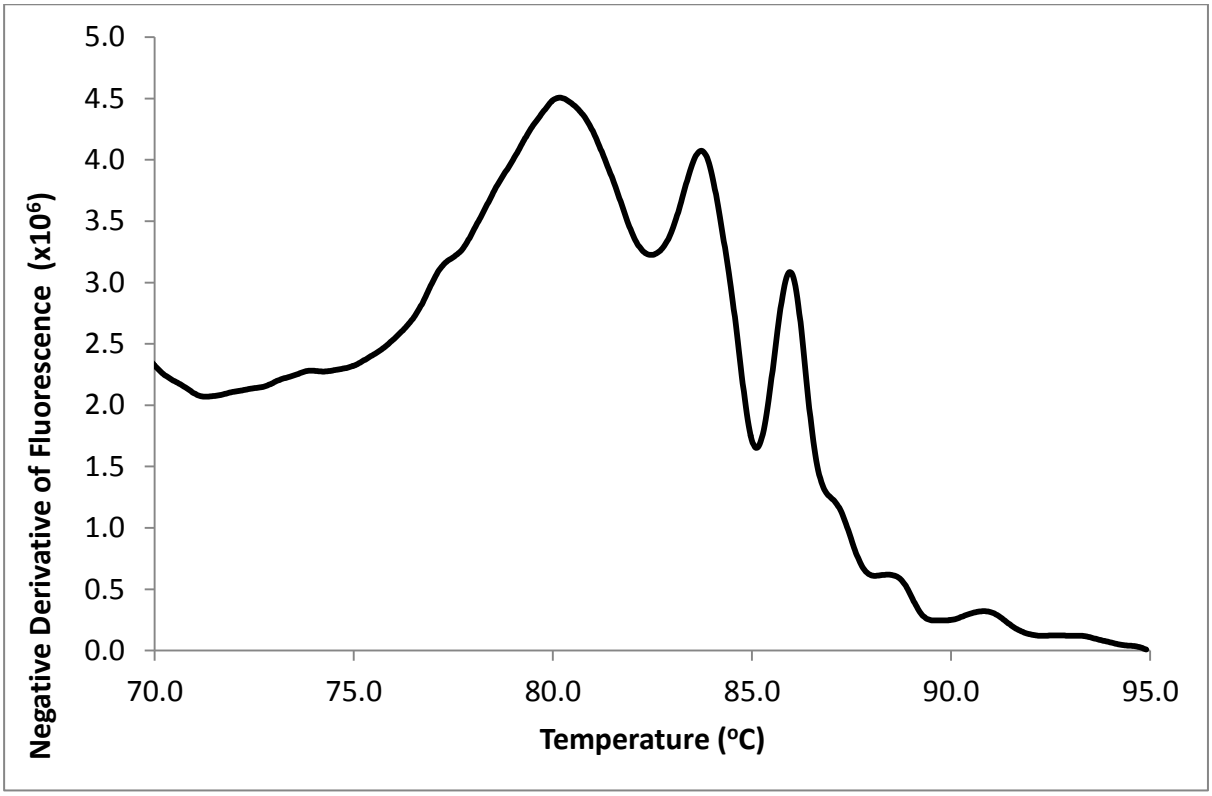
Supp. Info. Figure 1 b)



Supp. Info. Figure 1 c)



Supp. Info. Figure 2



Supplementary tables

Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D (2014) Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis 35 (21–22):3036–3043

Supplementary Table 1

dbSNP ID	Forward Primer	Tm (°C)	%GC	Reverse Primer	Tm (°C)	%GC	Amplicon Length	Study
rs733559	AGGAGCAGATTAAAATGAGGTA	55.0	36.4	GATCTGATACTTTACCTTCCACAT	55.6	37.5	47	1
rs12913832	CGAGGCCAGTTTCATTTGAGCATTAA	56.0	42.0	ATGATGATAGCGTGCAGAACTTGACA	56.0	42.0	53	1
rs12896399	CTTTGTTCTTTAGGTCAGTATATTTTGGG	56.0	34.0	GAAGGTTAATCTGCTGTGACAAAGAGA	57.0	41.0	57	1
rs1393350	CCTCAGTCCCTTCTCTGCAAC	56.0	57.0	AGATTATCATTTGTAAAAGACCACACAGATTT	55.0	28.0	54	1
rs12203592	CCACTTTGGTGGGTAAAAGAAGG	55.0	48.0	CACCAAAAGTACCACAGGGGAATTT	56.0	44.0	49	1
rs1800407	CAGGCATACCGGCTCTCCCC	58.0	68.0	ATGGCCCACACCCGTCCC	57.0	72.0	38	1
rs16891982	TGAGGAAAACACGGAGTTGATGCA	56.0	46.0	CGAGGTTGGATGTTGGGGCTT	56.0	57.0	46	1
rs310850	TTGTGTTTTGTTCAGCTGTTTA	56.2	31.8	CAGGCTTTTCCTAGAGCAA	55.9	47.4	42	1 and 2
rs3892905	CATGTTCATAAAGGTGATCCA	56.0	38.1	GTGGACCCCCTAAAGTTAA	54.1	47.4	41	1 and 2
rs10494531	TGTGCAGACTCGGCTTT	56.3	52.9	AATCACAGCCTGGGGTAA	56.4	50.0	36	2
rs723937	GAAAAATGTAATTTGCTAGGTCA	55.4	30.4	TCCAGACATTCCTACTCAATG	55.2	42.7	45	2
rs963170	TTTCCTGCTTCCCTTTTTC	57.0	42.1	GCTAAATTGTTGTGAATTAATCTGA	56.8	28.0	45	2
rs10505351	GCCCTCGTAGTAAAAAGATGA	56.2	42.9	TTGGTTTCAGATTTTGACTCTG	57.0	36.4	44	2
rs2185785	ACCTGAAGGCCTAGAACTTATT	55.5	40.9	CATACCCGCTTGCTTCTTA	56.6	47.4	42	2
rs1421883	GCAGTATGATTGGAGTGATCT	54.3	42.9	TGTTTTTTTACTGAGCTTAGAATG	55.1	29.2	46	2
rs724404	AACACGAAGGGTGGGAT	56.1	52.9	TGCTAAAAATGCAGACAACTG	56.7	38.1	39	2
rs2327046	AGGTATGGTGCCTCACAC	54.0	55.6	CCATAGGTACAGAATTCCTACTT	53.6	39.1	42	2
rs10485226	ACTCTTAAGACGTTAACATTTTTTAG	54.2	26.9	CTCACTCAGGACTTCCTTTG	55.0	50.0	47	2
rs10496623	TGAGCCAACTTGTCACACT	55.1	47.4	ACTAACAGCCATCCTATCTAAGAG	55.3	41.7	44	2

Supplementary Table 2

PCR set up **PCR conditions**

a) MeltDoctor® HRM Master Mix (Life Technologies – SYTO-9 dye)

Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	10	1 ×	Activation	95	10 mins
FWD Primer	10 µM	0.6	0.3 µM	45 × Denaturation	95	15 sec
REV Primer	10 µM	0.6	0.3 µM	Annealing/Extension	60	60 sec
dH_2O		7.8		HRM	95	10 sec
					60	60 sec
Total Master Mix		19			60-95	0.025°C/sec
DNA (suggest 200 pg to 200 ng)		1			95	15 sec
Total Volume		20			60	15 sec

b) Precision Melt Supermix (BioRad – EvaGreen dye)

Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	10	1 ×	Activation	95	2 mins
FWD Primer	10 µM	0.4	0.2 µM	45 × Denaturation	95	10 sec
REV Primer	10 µM	0.4	0.2 µM	Annealing/Extension	60	30 sec
dH_2O		4.2		HRM	95	30 sec
Total Master Mix		15			60	60 sec
DNA (suggest 1 ng to 50ng)		5			60-95	10 sec/step
Total Volume		20				0.2°C/step

c) KAPA HRM Fast PCR kit (Geneworks – EvaGreen dye)

Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	10	1 ×	Activation	95	2 mins
FWD Primer	10 µM	0.4	0.2 µM	45 × Denaturation	95	5 sec
REV Primer	10 µM	0.4	0.2 µM	Annealing/Extension	60	30 sec
$MgCl_2$	25 mM	2	2.5 mM	HRM	95	60 sec
dH_2O		6.2			60	60 sec
Total Master Mix		19			60-95	2 sec/step
DNA (suggest 100 pg to 20 ng)		1				0.2°C/step
Total Volume		20				

d) SensiMix (BioLine – EvaGreen dye)

Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	12.5	1 ×	Activation	95	10 mins
FWD Primer	10 µM	0.5	0.3 µM	45 × Denaturation	95	15 sec
REV Primer	10 µM	0.5	0.3 µM	Annealing	60	10 sec
$MgCl_2$	50 mM	0.5	2.5 mM	Extension	72	10 sec
EvaGreen	25 ×	1	1 ×	HRM	95	10 sec
dH_2O		5			75	60 sec
Total Master Mix		20			75-95	5 sec/step
DNA (suggest 1 ng to 100 ng)		5				0.1°C/step
Total Volume		25				

e) SensiFast (BioLine – EvaGreen dye)

Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	10	1 ×	Activation	95	3 mins
FWD Primer	10 µM	0.8	0.4 µM	45 × Denaturation	95	5 sec
REV Primer	10 µM	0.8	0.4 µM	Annealing/Extension	60	30 sec
dH_2O		4.4		HRM	95	10 sec
				(as per instrument)	60	60 sec
Total Master Mix		16			60-95	0.025°C/sec
DNA (suggest 1 ng to 1 ug)		4			95	15 sec
Total Volume		20			60	15 sec

Supplementary Table 3

	SensiMix™ HRM	**SensiFast™ HRM**	**KAPA™ HRM Fast PCR**	**Precision Melt Supermix**	**MeltDoctor™ HRM MasterMix**
Manufacturer	Bioline	Bioline	KAPA Biosystems	BioRad	Life Technologies
Amplify rs733559 using minimal DNA?	×	✓	✓	✓	✓
Differentiation of genotypes?		✓	✓	✓	✓✓[a]
No extraneous peaks at lower DNA input amounts?		×	×	✓	✓
Minimal variation between individuals of same genotype?				✓	✓

[a] MeltDoctor™ HRM MasterMix produced larger temperature differences between the 2 genotypes observed (see Figure 1) compared to Precision Melt Supermix

Supplementary Table 4

Sample ID	rs733559[a]	rs310850[a]	rs3892905[a]	rs12913832[b]	rs12896399[b]	rs1393350[b]	rs12203592[b]	rs1800407[c]	rs16891987[c]
1	CC								
2	CT								
3	CT								
4	CT								
5	CT	GG	GG						
6	CC	AG	AG						
7	CC	AG	AA						
8	CT	AG	AA						
9	CT	AG	AG						
11	CT	AG	AA						
12	CT	AA	AA						
13	CC	AA	AA						
15	CC	AG	AA						
16	CC	AG	AG						
17	CT	AG	AG						
9947A	CC	GG	AA						
A				GG	GG	GG	CC	--	--
B				AG	TT	GG	CC	--	--
C				AA	GT	AG	CC	--	--
D				GG	GG	GG	CC	--	--
E				GG	GT	GG	CC	--	--

[a] HRM genotypes for these SNPs were confirmed using restriction enzyme digests and capillary electrophoresis

[b] HRM genotypes for these SNPs were confirmed using a SNaPshot assay (IrisPlex)

[c] HRM genotyping for these SNPs was inconclusive (indicated by "--").

3.1.2 Mehta B, Daniel R, McNevin D (2013) High resolution melting (HRM) of forensically informative SNPs. Forensic Science International: Genetics Supplement Series 4 (1):e376–e377

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 3.1.2

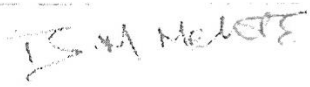
Declaration by candidate

In the case of Chapter 3.1.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Contributed to experimental work, data analysis and writing of the published manuscript	70

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student UC Y/N
Runa Daniel	Proposal of research, revision and editing of the manuscript, critical feedback on the research plan and data analysis	15	N
Dennis McNevin	Proposal of research, revision and editing of the manuscript, critical feedback on the research plan and data analysis	15	N

Candidate's Signature		Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or

interpretation, of at least that part of the publication in their field of expertise;

3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of co-author	Signature
04/04/2018	Dennis McNevin	Dennis McN.
5/4/18	RUNA DANIEL	Runa Daniel

[* Please insert additional rows as needed.]

[Article DOI: https://doi.org/10.1016/j.fsigss.2013.10.191]

Forensic Science International: Genetics Supplement Series 4 (2013) e376–e377

Contents lists available at ScienceDirect

Forensic Science International: Genetics Supplement Series

journal homepage: www.elsevier.com/locate/FSIGSS

High resolution melting (HRM) of forensically informative SNPs

Bhavik Mehta [a,*], Runa Daniel [b], Dennis McNevin [a]

[a] National Centre for Forensic Studies, University of Canberra, Canberra, Australia
[b] Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Australia

ARTICLE INFO

Article history:
Received 30 August 2013
Accepted 2 October 2013

Keywords:
High resolution melt
HRM
SNP genotyping
SNPs
IrisPlex
SNaPshot

ABSTRACT

The SNaPshot® assay is commonly used for forensic SNP analysis. However, it is a multi-step process with potential post-PCR contamination risk. The single tube high resolution melting (HRM) temperature real-time PCR method is an alternative, eliminating the post-PCR tube transfer of SNaPshot®. Eight individual DNA samples were genotyped at the six IrisPlex SNP loci using both the IrisPlex published primer set and a set of custom designed HRM primers. The performance of MeltDoctor™ (Life Technologies®) and SensiFast™ (Bioline®) HRM mastermixes was examined on the ViiA™ 7 Real Time PCR platform for 10 ng and 1 ng DNA template amounts. The resultant genotypes were compared with those derived from SNaPshot®. This preliminary study demonstrates HRM potentially offers a fast and flexible alternative to SNaPshot® for small numbers of SNP loci without the associated contamination risk from post-PCR processes.

1. Introduction

SNaPshot® (Life Technologies) is a single base extension (SBE) assay commonly used for forensic SNP genotyping. However, SNaPshot® is a multi-step process with a potential post-PCR contamination risk. High resolution melting (HRM) analysis is a single-tube process where the DNA region containing the SNP of interest is amplified by PCR in the presence of an intercalating fluorescent dye. After PCR, the amplified product is denatured by increasing temperature in small increments which produces a characteristic melting profile. The differences in melting profiles are used to determine the SNP genotypes [1] (Fig. 1). HRM results are interpreted using: (i) Derivative melt curves (Temperature vs. first derivative of fluorescence ($-dF/dT$)) to identify homozygote and heterozygote genotypes for each SNP (ii) Difference plots to visualize sample clusters according to SNP genotypes. In this preliminary study, we investigated the use of HRM analysis as an alternative to SNaPshot® for the IrisPlex six SNP panel [2] using IrisPlex and custom designed HRM primer sets and two different HRM chemistries on varying DNA template amounts.

2. Materials and methods

DNA was extracted from the buccal swabs of eight DNA donors using the QIAamp DNA Mini Kit (Qiagen®). Eight donors were selected to represent all eye color phenotypes (blue, brown and intermediate) in order to detect all possible SNP variants. HRM analysis was performed on a ViiA™ 7 Real Time PCR platform (Life Technologies) using Meltdoctor™ (Life Technologies®) and SensiFast™ (Bioline®) HRM chemistries which incorporate SYTO® 9 and Evagreen® intercalating fluorescent dyes respectively. The IrisPlex primer set [2] and a set of custom designed HRM primers were assessed on 10 ng and 1 ng DNA templates. Genotype variants were identified by the ViiA™ 7 software. The SNaPshot™ assay was performed on the same eight DNA samples using published assay conditions [2] and genotyping was performed using GeneMapper IDX™ software.

3. Results and discussion

MeltDoctor™ was more sensitive and produced more consistent results in comparison with the SensiFast™ HRM chemistry. Genotypes obtained using HRM and SNaPshot™ were concordant for five of the six SNPs for both 10 ng and 1 ng DNA template amounts using both the IrisPlex and custom designed primer sets. Due to the high GC content in the sequence around rs1800407 (A/G), the genotype variants were not clearly identifiable from the melting profile. Both of the primer sets produced three distinguishable variants for all the other SNPs in the IrisPlex panel. The difference plots in Fig. 2 show the three distinguishable variants (CC, CT and TT) for rs12913832 using both of the primer sets and the two DNA template amounts. Thus, the HRM technique can be used for primers designed for existing SBE assays. SNaPshot™ requires 8–10 h whereas HRM can be performed in 2–4 h.

* Corresponding author. Tel.: +61 413250288.
E-mail addresses: bhavik.mehta@canberra.edu.au, bhavik.mehta@hotmail.com (B. Mehta).

1875-1768/$ – see front matter © 2014 Elsevier Ireland Ltd All rights reserved.
http://dx.doi.org/10.1016/j.fsigss.2013.10.191

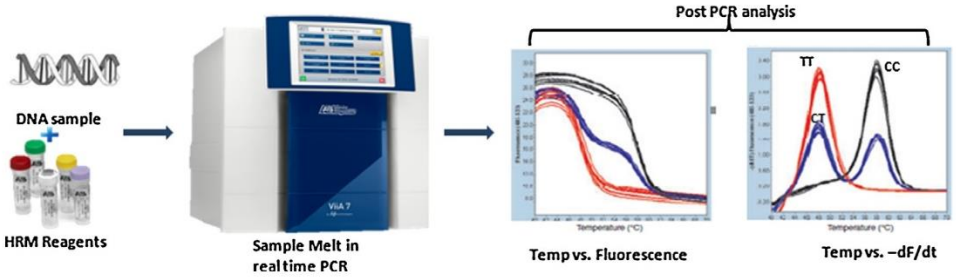


Fig. 1. Principle of HRM (adapted from http://www.appliedbiosystems.com/HRM products).

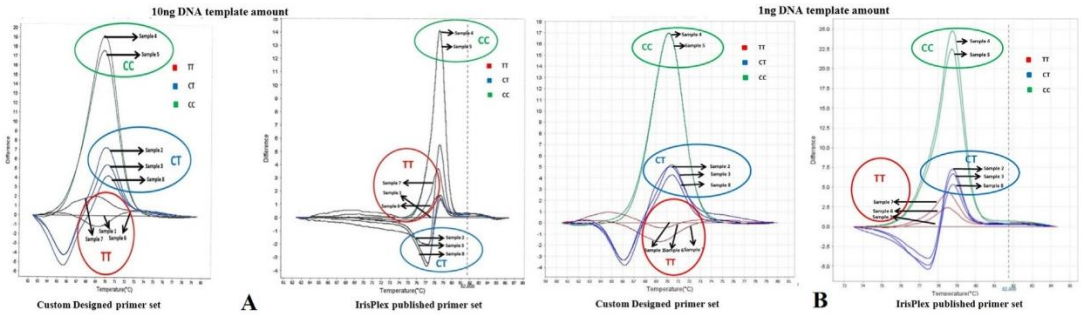


Fig. 2. Difference plots for SNP rs12913832 (C/T) using both the custom designed HRM primers and IrisPlex published primer sets showing three genotype variants (CC, CT and TT) over 10 ng (A) and 1 ng (B) DNA template amounts for eight samples.

4. Conclusion

This preliminary study demonstrates the potential of HRM SNP genotyping for a small number of forensically informative SNP loci, avoiding post-PCR processing. It will be extended to further assess forensic parameters such as sensitivity, reproducibility, mixture studies and the potential for multiplexing the HRM technique. HRM is a versatile technique and potentially offers a faster and viable alternative to SNaPshot™.

Role of funding

This work was supported by funding from the Australian Research Council (LP110100121: *From genotype to phenotype – Molecular photofitting for criminal investigations*) which had no involvement in the conduct of the research or preparation of this article.

Conflict of interest

None.

Acknowledgements

Dr. Samantha Venables (National Center for Forensic Studies, University of Canberra) for technical assistance and Dr. Roland van Oorschot (Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department) for general advice.

References

[1] G.H. Reed, J.O. Kent, C.T. Wittwer, High-resolution DNA melt analysis for simple and efficient molecular diagnostics, Future Med. 8 (2007) 597–608.
[2] S. Walsh, A. Lindenbergh, S.B. Zuinga, et al., Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, Forensic Sci. Int. Genet. 5 (2011) 464–471.

3.2 Mehta B, Daniel R, McNevin D (2017) HRM and SNaPshot as alternative forensic SNP genotyping methods. Forensic Science, Medicine, and Pathology 13 (3):293-301

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 3.2

Declaration by candidate

In the case of Chapter 3.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Proposal of the research, experimental work, data collection and analysis, writing the first draft of the manuscript	70

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student UC Y/N
Dennis McNevin	Revising and editing of the manuscript, critical feedback on the research plan and data analysis methods	15	N
Runa Daniel	Revising and editing of the manuscript, critical feedback on the research plan and data analysis methods	15	N

Candidate's Signature		Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.

2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	National Centre for Forensic Studies, Faculty of STeM, University of Canberra

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	Dennis McN.
5/4/18	RUMA DANIEL	Ruma Daniel

[* Please insert additional rows as needed.]

[Material from: 'Mehta, B., Daniel, R. & McNevin, D. HRM and SNaPshot as alternative forensic SNP genotyping methods. Forensic Sci Med Pathol 13, 293–301 (2017). https://doi.org/10.1007/s12024-017-9874-5, © 2020 Springer Nature Switzerland AG. Part of Springer Nature.]

Forensic Sci Med Pathol
DOI 10.1007/s12024-017-9874-5

ORIGINAL ARTICLE

HRM and SNaPshot as alternative forensic SNP genotyping methods

Bhavik Mehta[1] · **Runa Daniel**[2] · **Dennis McNevin**[1]

Accepted: 17 April 2017

Abstract Single nucleotide polymorphisms (SNPs) have been widely used in forensics for prediction of identity, biogeographical ancestry (BGA) and externally visible characteristics (EVCs). Single base extension (SBE) assays, most notably SNaPshot® (Thermo Fisher Scientific), are commonly used for forensic SNP genotyping as they can be employed on standard instrumentation in forensic laboratories (e.g. capillary electrophoresis). High resolution melt (HRM) analysis is an alternative method and is a simple, fast, single tube assay for low throughput SNP typing. This study compares HRM and SNaPshot®. HRM produced reproducible and concordant genotypes at 500 pg, however, difficulties were encountered when genotyping SNPs with high GC content in flanking regions and differentiating variants of symmetrical SNPs. SNaPshot® was reproducible at 100 pg and is less dependent on SNP choice. HRM has a shorter processing time in comparison to SNaPshot®, avoids post PCR contamination risk and has potential as a screening tool for many forensic applications.

Keywords High resolution melt (HRM) · SNaPshot · Single nucleotide polymorphism (SNP) · Forensic SNP genotyping · Single base extension (SBE)

Electronic supplementary material The online version of this article (doi:10.1007/s12024-017-9874-5) contains supplementary material, which is available to authorized users.

✉ Bhavik Mehta
bhavik.mehta@hotmail.com

[1] National Centre for Forensic Studies, Faculty of Education, Science, Technology & Mathematics, University of Canberra, ACT, Bruce 2617, Australia

[2] Office of the Chief Forensic Scientist, Forensic Services Department, Victoria Police, 31 Forensic Drive, Macleod, VIC 3085, Australia

Introduction

Single nucleotide polymorphisms (SNPs) have gained popularity as forensic markers due to some advantages over short tandem repeats (STRs) such as lower mutation rates, high abundance in the genome and short amplicon lengths [1]. These advantages make SNPs a better choice for degraded DNA samples. SNPs can also provide DNA intelligence in the form of biogeographical ancestry (BGA) and externally visible characteristics (EVCs) in a process commonly referred to as forensic DNA phenotyping (FDP) or molecular photofitting [2].

Forensic SNP genotyping is commonly performed using single base extension (SBE) assays, with the SNaPshot® assay (Thermo Fisher Scientific: TFS) most popular. SNaPshot® is a two-step PCR method consisting of target amplification and SBE product formation. After an initial PCR, fluorescently labelled dideoxynucleotidetriphosphates (ddNTPs) are incorporated at the 3′ end of a primer, which binds to the PCR amplicon immediately upstream of the SNP. Incorporation of the chain-terminating complementary ddNTP provides a means to determine the SNP in the template strand as each of the four ddNTPs (A, G, C and T) are labelled with a different fluorophore. The resulting SBE products are sized by capillary electrophoresis (CE) which generates electropherograms of SNP profiles [3]. There are a number of SNaPshot® forensic SNP assays such as the SNP*for*ID 52-plex [4] (an identity assay), SNP*for*ID 34-plex [5] (a BGA assay) and IrisPlex [6] (an EVC assay).

Recent studies have shown the potential of high resolution melting (HRM) analysis for forensic SNP genotyping [7, 8]. HRM is a post-PCR technique where double stranded PCR products are subject to an increasing temperature gradient using real time PCR instruments such as the ViiA™ 7 and 7500 Fast Real Time PCR systems (TFS) [7, 8]. After PCR, the reaction temperature is lowered to anneal single stranded DNA and then

Published online: 18 May 2017

subsequently raised. Dissociation of a saturating, intercalating fluorescent dye from the double stranded DNA as it melts results in loss of fluorescent signal with fluorescence proportional to the amount of double stranded DNA remaining. A HRM melt profile is then defined as the plot of fluorescence intensity with respect to temperature over a range of incremental temperature steps. A HRM difference profile is produced by plotting fluorescence intensity for all variants relative to an arbitrary reference melt profile. Finally, HRM derivative profiles express the change in fluorescence intensity with respect to the change in temperature ($-dF/dT$) and this allows for accurate determination of melting temperature (T_m) [8, 9].

Two apparent melting temperatures (melt peaks) are associated with the derivative plot for a heterozygote. These correspond with the melt profiles for the two homoduplexes (allele 1 + allele 1 complement and allele 2 + allele 2 complement) and the two heteroduplexes (allele 1 + allele 2 complement and allele 2 + allele 1 complement) formed when the temperature is initially lowered (before the melt profile is generated) [10, 11]. When combined, the melting profiles of the homoduplexes and the heteroduplexes generally form two melt peaks. It should be remembered that the derivative reflects the rate of melting. The heteroduplexes and the homoduplexes melt at different rates (thus giving rise to a different derivatives) and at different temperatures. Heteroduplexes have base pair mismatches at the SNP position while homoduplexes do not. As a result, homoduplexes (and, indeed, symmetrical SNPs) are difficult to distinguish, hence producing only one other melt peak, separate from the heteroduplexes.

This study extends our previous assessment of HRM [7] by including additional forensic parameters and comparing genotype concordance with SNaPshot® using the IrisPlex [6] SNP panel for eye (i.e. iris) color. The parameters reported here are assay sensitivity, reproducibility (intra- and inter- run variability), multiplexing capability, mixture detection and cost.

Materials and methods

Sample collection and quantitation

Cotton-tipped buccal swabs (Interpath, Cat # 8150CIS) were used to collect buccal epithelial cells from eight human donors with ethics approval from the University of Canberra Committee for Ethics in Human Research (Project number 11–119). DNA was extracted using the QIAmp™ DNA Mini Kit (Qiagen®). DNA from two standard cell lines, 007 (TFS) and 9947A (TFS), were also used for this study. All ten samples were quantified by the Quantifiler® Human DNA Quantification Kit (TFS) according to the standard manufacturer's protocol. DNA was diluted in Tris-EDTA (TE: 10 mM Tris, 0.1 mM EDTA) buffer to final concentrations of 10 ng/μL, 1 ng/μL, 0.5 ng/μL and 0.1 ng/μL.

Primers and PCR protocols

Two sets of primers were used for HRM SNP genotyping of the IrisPlex assay for eye color prediction. In addition to the published IrisPlex SNaPshot® PCR primers [12], we also designed our own custom primers for the same six SNPs which are described in Venables et al. [8]. The published IrisPlex primers generate amplicons greater than 100 bp in size [10] whereas the custom primers were designed so that amplicon sizes were in the optimum range of 40–80 bps for HRM Increasing the size of PCR amplicons reduces the melting temperature difference between alternate genotypes and thus decreases differentiation between variants [8, 9]. The IrisPlex SNaPshot® assays were performed as prescribed by Walsh et al. [12]. SNaPshot® SBE products were electrophoresed on a 3500xl Genetic Analyser (TFS) with a 36 cm capillary array and POP-4™ polymer using Dye Set E5, Matrix Standard DS-02 and GS120 LIZ® Size Standard, according to the manufacturer's recommended protocol [11]. HRM reactions were performed according to Venables et al. [8] using a ViiA™ 7 Real Time PCR instrument (TFS).

SNP genotyping and eye color prediction

HRM SNP genotyping was performed using the ViiA™ 7 Research Use Only (RUO) software following the procedure described by Venables et al. [8]. SNaPshot® electropherograms were generated using Genemapper™ ID-X software (TFS, v1.4). The HRM and SNaPshot® genotypes were compared. In both cases, eye color prediction was performed using the multinomial logistic regression (MLR) prediction algorithm excel macro described by Walsh et al. [12]. The predicted phenotypes were compared against the self-declared phenotypes for eight human donors.

Assay sensitivity

Eight human samples (A-H) were typed with DNA template inputs of 10 ng, 1 ng, 0.5 ng and 0.1 ng using both HRM and SNaPshot®.

Reproducibility

Intra- and inter-plate variability was assessed on eight human donors and two standard samples (007 and 9947A) with DNA template inputs of 0.5 ng and 0.1 ng. Our custom designed primers were used for HRM. Inter-plate variability was performed using three replicates of each of the ten samples on three 96-well plates (TFS). Intra-plate variability was assessed on a 384-well plate (TFS) plate in triplicate for all ten samples

at both the above input amounts. For SNaPshot®, inter- and intra-plate variability was performed on 96-well (TFS) plates in triplicate for each of the ten samples.

Mixtures

Two samples (A and 007) were mixed at ratios 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8 with the total template input constant at 1 ng. The HRM mixture experiment was carried out using the custom primers and SNaPshot® was performed using the published assay conditions.

Multiplexing capability

Although the IrisPlex SNaPshot® assay is a six-SNP multiplex, the default HRM analysis involved one PCR reaction (i.e. one tube) for every SNP. However, four samples (D, G, 007 and 9947A) were subject to a HRM multiplex with all six IrisPlex SNPs included. The reaction conditions for this HRM multiplex are documented in Supplementary Table 1.

Half volume reactions

Two samples (D and G) were subject to HRM using half volume reactions (10 μl) to extend the cost-effectiveness of the approach using the custom designed primer set. All concentrations remained the same.

Results

The results of the comparative study are described below.

Genotype and phenotype comparison

The genotypes for all ten samples are presented in Table 1. HRM derivative curves for homozygotes display a single peak whereas two peaks were observed for heterozygotes (an example has been shown in Supplementary Fig. 1). DNA input amounts of 0.5 ng are typically used in forensic laboratories for DNA profiling (personal communication: Australian Federal Police and Victoria Police forensic laboratories). Hence, the genotype comparison between HRM and SNaPshot® was performed at 0.5 ng. The HRM genotypes of four IrisPlex SNPs (rs12913832, rs12203592, rs128936399, rs1393350) were 100% concordant with SNaPshot® for all samples at 0.5 ng DNA input. HRM was unable to type SNPs rs1800407 and rs16891982 as the former had GC-rich flanking regions affecting melt analysis (Supplementary Fig. 2) and the latter was a symmetrical G/C SNP for both the custom (Supplementary Fig. 3) and published primer sets (data not shown). GC-rich amplicons resulted in multiple melting domains and the variants of symmetrical SNPs are often not clearly differentiated using HRM [8]. Hence, these two SNPs were discounted from further HRM analysis. The IrisPlex Excel MLR macro assigns eye color to three categories: blue, brown and intermediate. Phenotype prediction using HRM was not performed due to the unavailability of genotypes for SNPs rs1800407 and rs16891982. Phenotypes predicted by IrisPlex for the SNaPshot® genotypes were concordant with the corresponding self-declared phenotypes (Table 1). Sample H with self-declared hazel eye color and 007 did not return more than 50% probability for any of the three eye color categories when predicted using MLR and hence can be referred to as "inconclusive."

Assay sensitivity

HRM produced genotypes that were concordant with SNaPshot at 10 ng, 1 ng and 0.5 ng. However, at 100 pg DNA input amounts, some genotypes were not concordant with SNaPshot® (Table 2). Fig. 1 shows the HRM derivative curves using our custom primers, Fig. 2 shows the same HRM derivative curves for the primers published by Walsh et al. [12] and Supplementary Fig. 4 shows the fluorescence signals for SNaPshot® electropherograms, all for Sample A. At 10 ng DNA input, SNaPshot® generated over amplified peaks which saturated the electropherograms and hence these were not included in Table 2 and Supplementary Fig. 4. The published IrisPlex SNaPshot® PCR primers [12] resulted in sensitivity equal to that from the custom designed primers.

Reproducibility

Inter- and intra-plate comparison of genotypes indicated that both HRM and SNaPshot® achieved full genotype concordance at 0.5 ng except for SNPs rs1800407 and rs16891982 for which HRM genotypes were not obtained. However, HRM profiles were not reproducible within (Supplementary Fig. 5) or between (Supplementary Fig. 6) plates at 0.1 ng for the four remaining SNPs. SNaPshot® was reproducible within and between plates at 0.1 ng.

Mixtures

HRM was unable to resolve the two sample mixtures employing our custom designed primers. SNP genotypes of mixture ratios up to 1:8 always formed apparent heterozygotes comprised of the alternate alleles of samples A and 007 except for rs12203592 where both samples were homozygous for the same allele (Supplementary Table 2 and Supplementary Fig. 7). SNaPshot® suffers the same disadvantage as HRM in resolving mixtures as we have previously shown [8].

Table 1 SNaPshot® and HRM genotypes for eight human donors (A-H) and two cell line controls (007 and 9947A) used in the study at 0.5 ng DNA input amount

Samples	SNPs						Predicted eye colour	Self-declared eye colour
	rs12913832 (C/T)	rs12203592 (C/T)	rs12896399 (G/T)	rs1393350 (C/T)	rs1800407 (G/A)	rs1689182 (G/C)		
SNaPshot® genotypes								
Human donors								
A	TT	CC	GG	CC	GG	CC	99.4% brown	brown
B	CT	CC	TT	CC	GG	CC	94.1% brown	brown
C	CT	CT	GT	CC	GG	CG	78.7% brown	brown
D	CC	CC	GT	CC	AG	GG	92.7% blue	blue
E	CC	CC	GT	CT	GG	GG	93.7% blue	blue
F	TT	CC	GG	CC	GG	CC	99.4% brown	brown
G	TT	CT	GG	CC	GG	CC	99.4% brown	brown
H	CT	TT	GT	CC	GG	GG	40.5% brown, 28% blue and 31.5% intermediate	hazel
Controls								
007	CT	CC	GT	TT	AG	GG	45.4% blue, 26.5% intermediate, 28% brown	----------
9947A	CC	CT	TT	CC	GG	GG	95.6% blue	----------
HRM genotypes using both custom and SNaPshot published PCR primers								
Human donors								
A	TT	CC	GG	CC	----------	----------	N/A	brown
B	CT	CC	TT	CC	----------	----------	N/A	brown
C	CT	CT	GT	CC	----------	----------	N/A	brown
D	CC	CC	GT	CC	----------	----------	N/A	blue
E	CC	CC	GT	CT	----------	----------	N/A	blue
F	TT	CC	GG	CC	----------	----------	N/A	brown
G	TT	CT	GG	CC	----------	----------	N/A	brown
H	CT	TT	GT	CC	----------	----------	N/A	hazel
Controls								
007	CT	CC	GT	TT	----------	----------	N/A	----------
9947A	CC	CT	TT	CC	----------	----------	N/A	----------

Table 2 Concordance between SNaPshot and HRM (using our custom primer set) at 1 ng, 0.5 ng and 0.1 ng DNA input amount for samples A to H. A DNA input amount of 10 ng saturated the SNaPshot electropherograms (Y –yes and N- no)

SNPs	Concordance between HRM and SNaPshot (%)							
	A	B	C	D	E	F	G	H
1 ng DNA input								
rs12913832	Y	Y	Y	Y	Y	Y	Y	Y
rs12203592	Y	Y	Y	Y	Y	Y	Y	Y
rs12896399	Y	Y	Y	Y	Y	Y	Y	Y
rs1393350	Y	Y	Y	Y	Y	Y	Y	Y
rs1800407	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
rs1689182	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.5 ng DNA input								
rs12913832	Y	Y	Y	Y	Y	Y	Y	Y
rs12203592	Y	Y	Y	Y	Y	Y	Y	Y
rs12896399	Y	Y	Y	Y	Y	Y	Y	Y
rs1393350	Y	Y	Y	Y	Y	Y	Y	Y
rs1800407	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
rs1689182	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
0.1 ng DNA input								
rs12913832	Y	Y	Y	Y	Y	Y	N	N
rs12203592	Y	N	N	Y	Y	N	Y	Y
rs12896399	Y	Y	N	Y	N	Y	Y	N
rs1393350	Y	Y	Y	N	N	Y	Y	Y
rs1800407	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
rs1689182	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Multiplexing capability

The IrisPlex SNaPshot® assay is a multiplex PCR and was able to unambiguously genotype all six SNPs (Table 1) however HRM was only able to genotype singleplex reactions. HRM multiplexes produce multiple melting domains (Fig. 3) and while HRM differential curves indicated the presence of three distinct variants, this was less than the range of variants expected from six SNPs (from 6 to 18 variants) and hence the ViiA™ 7 RUO software could not assign them to distinguishable genotypes. This demonstrates the limited multiplexing capability of HRM.

Half volume reaction

It is recommended that SNaPshot® reactions are performed in a total volume of 10 μl [13] however IrisPlex SNaPshot amplification is performed in 5 μl [6, 12]. Hence, half volume SNaPshot® reactions do not compromise performance. HRM was tested at half volume for two samples: D and G using custom designed primers. Fig. 4 shows that the half volume reaction derivative curves generated secondary melting domains in some instances (rs12203592, rs1393350, rs16891982), which resulted in incorrect genotype assignments using ViiA™ 7 software. While HRM showed some potential to generate profiles with half volume reactions, further optimization is required for consistency and reliability of the genotypes.

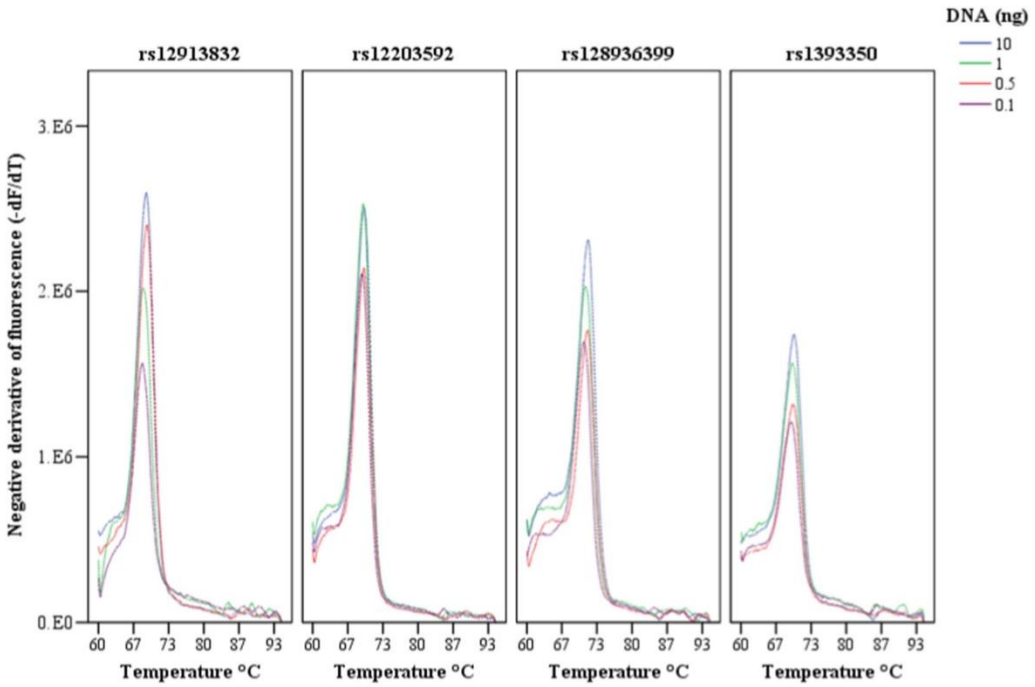


Fig. 1 HRM derivative curves for sample A using custom PCR primers with 10, 1, 0.5 and 0.1 ng DNA template input amounts

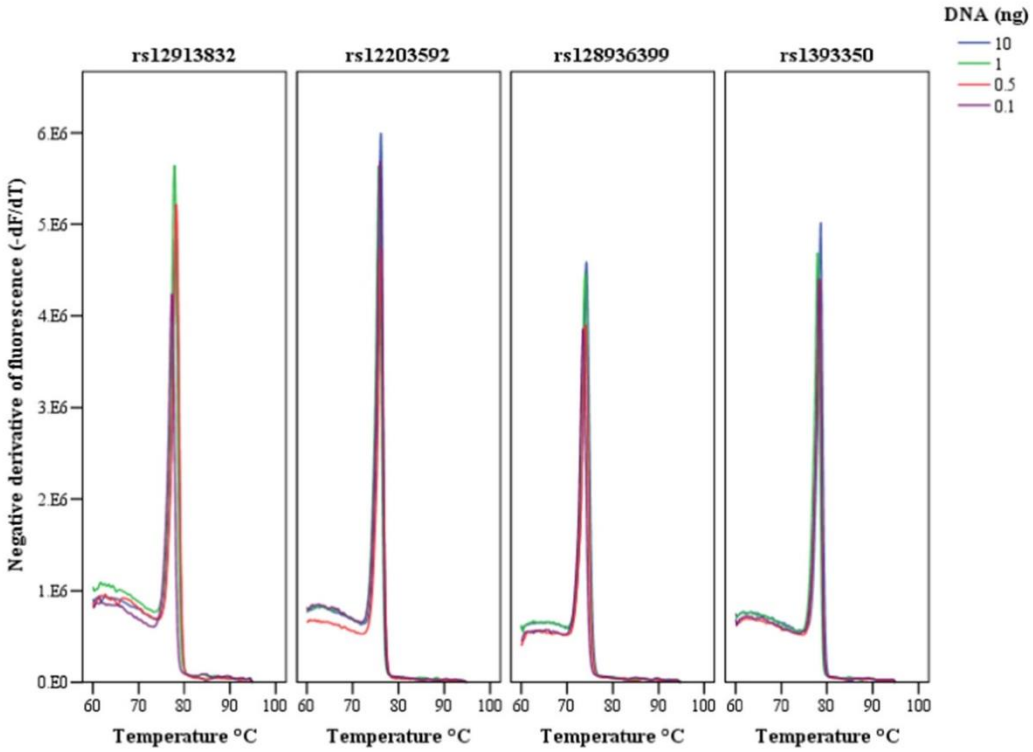


Fig. 2 HRM derivative curves for sample A using IrisPlex PCR primers with 10, 1, 0.5 and 0.1 ng DNA template input amounts

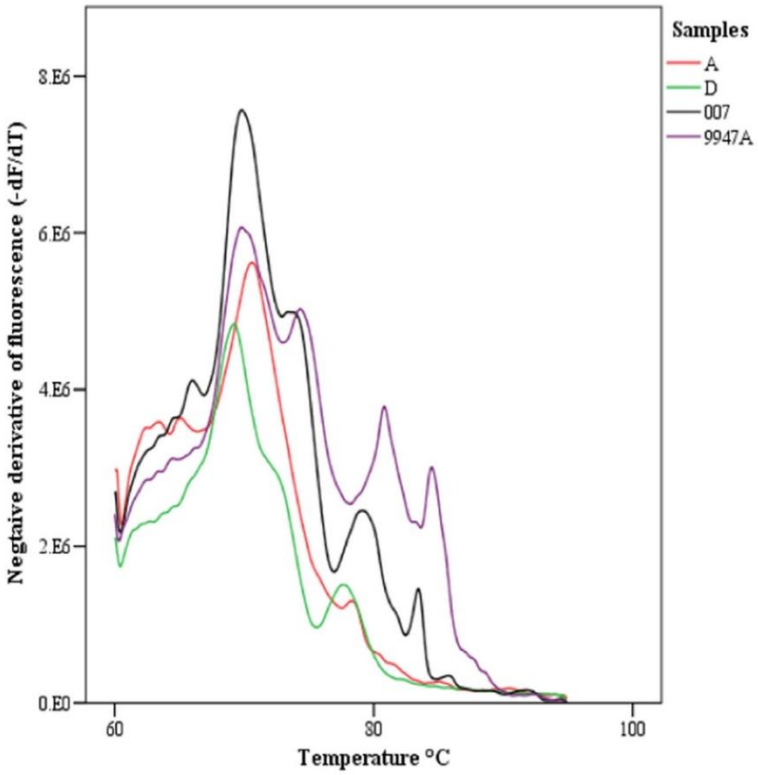


Fig. 3 HRM derivative curve for a multiplex of the six IrisPlex SNPs in a single reaction for four samples showing multiple melting domains using custom designed PCR primers

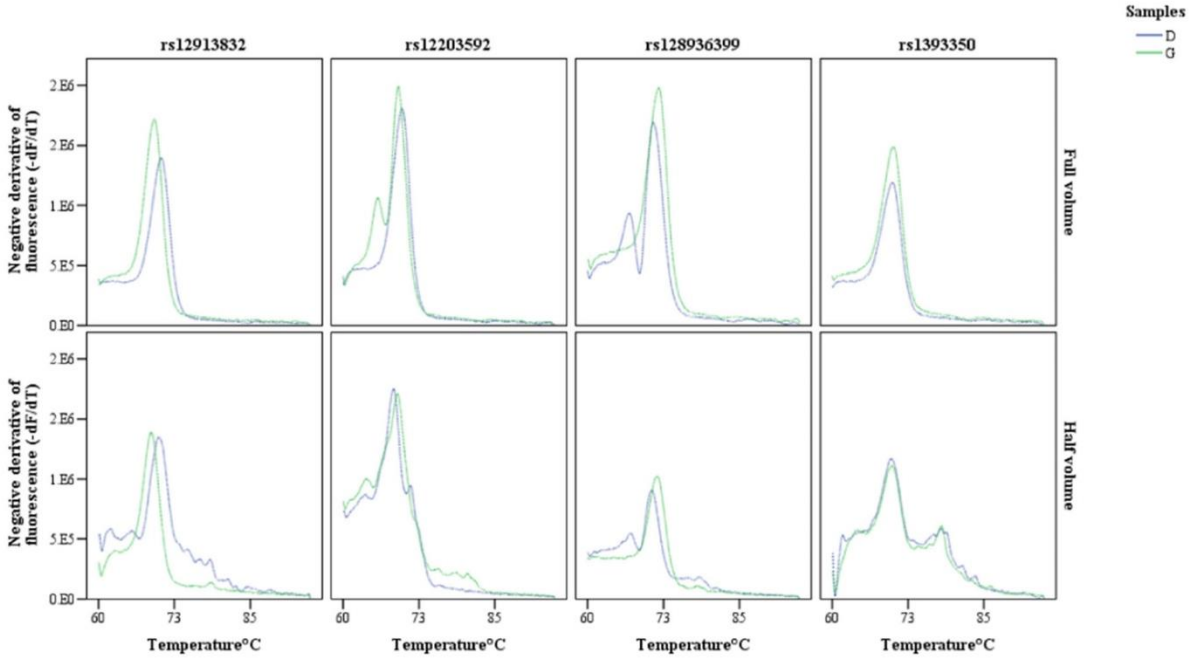


Fig. 4 HRM derivative curves for samples D and G for full- and half-volume reactions using custom designed PCR primers

Cost analysis and assay comparison

A cost comparison between HRM and SNaPshot® was performed based on the available list prices of each on the manufacturer's website (https://www.thermofisher.com/au). The cost per SNP for SNaPshot® was less in comparison to HRM for typing all six IrisPlex SNPs (Table 3). The cost for SNaPshot® can further decrease with increase in the number of SNPs per multiplex over and above the six used in IrisPlex. However, the assay comparison also showed HRM to be faster and less labor intensive than SNaPshot®.

Discussion

High resolution melting (HRM) is a closed, single tube assay requiring less processing time than SNaPshot®. HRM also offers the advantage of avoiding multiple tube transfers and thus avoiding post-PCR contamination [7]. The selection of SNPs with relatively large melting temperature differences is a key requirement for reliable HRM genotyping and hence symmetrical SNPs (G/C or A/T) are not recommended [8, 9]. However, well-established forensic SNP assays such as IrisPlex contain some symmetrical SNPs, which are critical to phenotype prediction. SNP rs16891982 (G/C) is one such example and did not generate reliable and reproducible HRM genotypes. The derivative curves for SNP rs1800407 (Supplementary Fig. 2) demonstrate the effect of GC-rich flanking regions on HRM genotyping and confirm our previous findings [8]. SNP rs1800407 is the second best eye color predictor in the IrisPlex assay and thus cannot be excluded. In contrast to HRM, SNaPshot® was able to genotype this SNP and seems less sensitive to GC-rich amplicons.

The MLR prediction of the phenotypes for eight human donors was completely concordant with their self-declared eye colors using SNaPshot® genotypes (Table 1). The MLR Excel macro designed by Walsh et al. [6, 12] could not be applied to HRM because of its limited ability to genotype SNPs rs1800407 and rs16891982 (for both custom and published SNaPshot® primer sets) however complete genotype concordance with SNaPshot® was observed for the other four SNPs at 0.5 ng DNA input amounts (Table 1). This suggests that HRM can only be applied if SNPs are selected judiciously and meet specific criteria, namely that variants have sufficient difference in melting temperatures (i.e. no symmetrical SNPs) and do not have GC-rich flanking regions [8]. Redesigning the published IrisPlex primers for HRM such that they produced smaller amplicons had little effect on the sensitivity of the HRM assay.

The HRM assay displayed inter- and intra-plate reproducibility and genotype concordance with SNaPshot at 500 pg DNA input amount. The melt profile of samples at 100 pg displayed more melting domains than expected and hence the HRM software interpreted the genotypes with additional variants. Training the ViiA™7 RUO HRM software with known genotypes as controls did not improve the reproducibility. On the other hand, SNaPshot® yielded reproducible genotypes at 100 pg. The IrisPlex SNaPshot® assay has elsewhere been shown to be reproducible at 62 pg DNA input amount [6, 12].

Table 3 Summary of HRM and SNaPshot® SNP genotyping techniques

Condition	HRM	SNaPshot®
Assay type	Real time PCR	Single base extension (SBE)
Assay protocol	Simple, single tube assay	Multiple tube transfers
PCR	Single PCR per SNP	Two PCR steps
Additional equipment and software requirement for forensics	High resolution real time PCR instrument (eg. ViiA™ 7) with HRM software	Compatible with existing equipment in most forensic labs (*eg.* capillary electrophoresis)
Time	2–4 h	12–15 h
Reproducibility (for full profiles)	0.5 ng DNA	0.062 ng DNA amount [10]
Multiplex capability	Limited	Yes (up to 30–40 SNPs per PCR)
Mixed DNA sources	Not suitable	Limited suitability [10] (unbalanced fluorescence signals from fluorophores)
Sensitivity to GC rich flanking sequence	Yes (flanking region rich with GC introduces extra melting domains)	No
SNP selection	SNPs with >0.5 °C melting temperature difference are ideally selected. Symmetrical SNPs are difficult to genotype accurately.	No (except each SNP should meet the PCR assay design and multiplexing criteria).
Genotyping software	Simple	Requires interpretation of electropherograms
Cost/ SNP genotype for genotyping IrisPlex panel	AU$ 0.85	AU$ 0.50 (decreases with increase in number of SNPs per multiplex)

Neither HRM nor SNaPshot was able to resolve mixtures. This is due to the biallelic nature of SNPs that make heterozygote genotypes indistinguishable from the genotypes of two or more contributors with alternate alleles. Further, a combination of homozygote and heterozygote always results in a heterozygote in the mixture (Supplementary Table 2). The limited ability of SNaPshot® to resolve mixtures with biallelic SNPs has also been highlighted in other studies [12, 14]. Tri-allelic SNPs can reduce this limitation and both HRM and SNaPshot are known to be capable of typing tri-allelic SNPs [15, 16].

Forensic SNaPshot® assays such as the SNP*for*ID 52-plex [4], SNP*for*ID 34-plex [17], HIrisPlex [14] and IrisPlex [6, 12] have demonstrated the ability to multiplex in the range 6–34 SNPs in a single PCR reaction. HRM has been shown here and elsewhere to have limited multiplexing capability [8, 18] and this appears to be a limitation of the technique. The melting profiles of the heteroduplexes formed between the various combinations of single stranded DNA give rise to multiple melting domains in any multiplex HRM. These differ from the two homoduplex and two heteroduplex melting profiles for a single heterozygote (giving rise to two derivative melt peaks) or the one possible homoduplex for a single homozygote (giving rise to one melt peak). The multiple, heteroduplex melting domains cannot be resolved into the mixture of genotypes that produced them. Consequently, multiple singleplex HRM assays may consume more DNA compared to a single multiplex SNaPshot® assay and hence HRM is not a viable choice for samples with low DNA yield. HRM is also more costly than multiplexed SNaPshot® assays on a per SNP basis. Half volume reactions could potentially improve the cost effectiveness of HRM, however, our study indicated reduced accuracy with limited multiplex optimization.

SNaPshot® has been the most common approach for forensic SNP typing until recent times; its major advantage being that it can be performed using equipment readily available in most forensic laboratories (e.g. thermal cyclers, CE). However, HRM can be easily incorporated with a minor upgrade of real time PCR instrumentation to high resolution (e.g. 7500 Fast Real Time PCR systems: TFS). This would introduce a fast, reliable and versatile low throughput SNP genotyping alternative for eligible SNPs and could be used as a screening tool for many forensic applications. For example, a 22 forensic STR HRM screening assay exists that enables the discrimination of samples from different individuals and could be utilized for the selection of potential samples to be allowed for routine STR profiling [19]. A HRM assay of just ten non-symmetrical identity SNPs, for example, could be used as a screen to determine which DNA samples can be excluded as being from the same source as an evidentiary sample and which could be included and thus subject to standard STR profiling. This has the potential to save money on the far more costly STR profiling assay. HRM is also known to be useful for bacterial [20], plant [21, 22] and animal [23] species identification; and for identifying human body fluids [24].

Key points

1. High resolution melt (HRM) analysis is a simple, fast, closed tube alternative SNP typing method to SNaPshot.
2. High resolution melt (HRM) analysis produced reproducible SNP genotypes at 500 pg.
3. High resolution melt (HRM) analysis, unlike SNaPshot, was unable to genotype symmetrical SNPs and SNPs with GC-rich flanking regions.
4. High resolution melt (HRM) analysis offers limited multiplexing ability.

Acknowledgments The authors gratefully acknowledge financial support from the Australian Research Council (LP110100121 - From genotype to phenotype: Molecular photofitting for criminal investigations).

Compliance with ethical standards

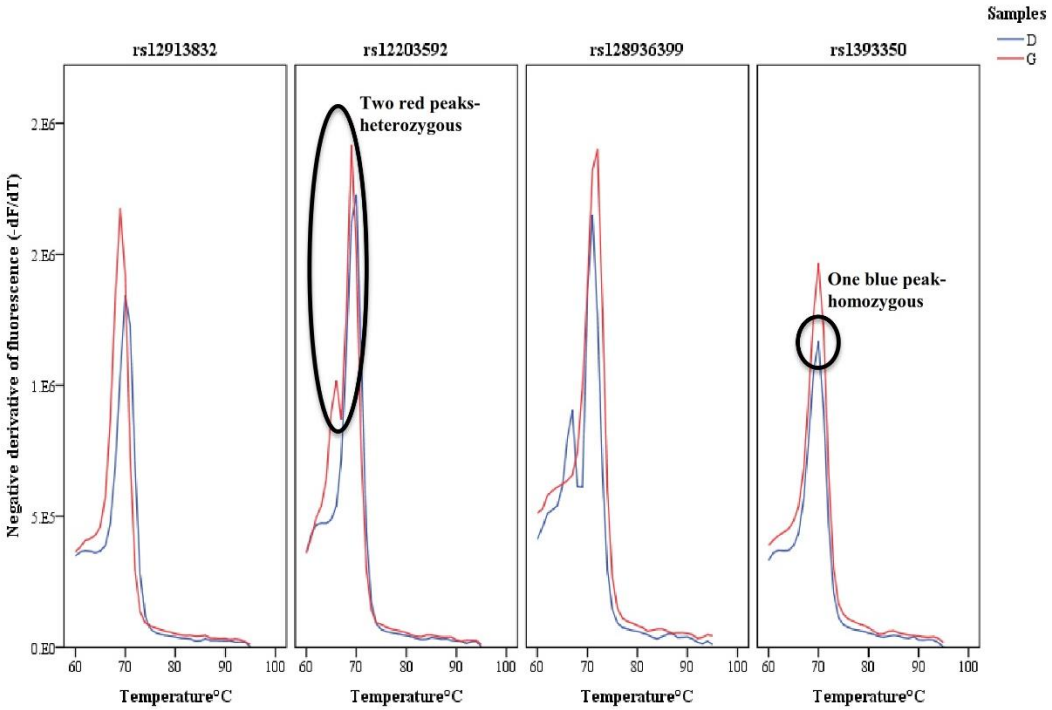
Ethics approval All procedures performed in this study involving human participants were in accordance with the ethical standards of the University of Canberra Human Ethics Committee (Project number 11–19) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Conflict of interest The authors declare that they have no conflict of interest.

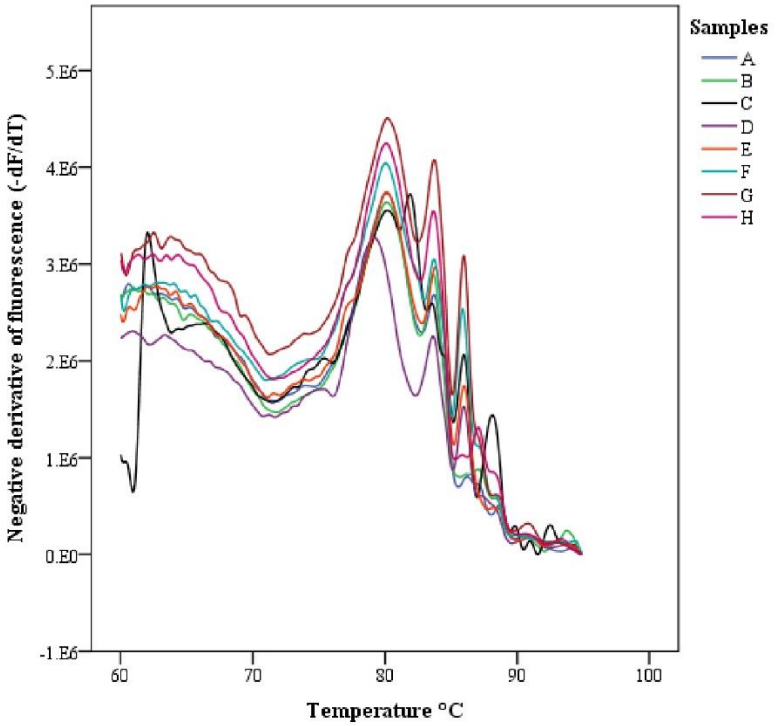
References

1. Budowle B, van Daal A. Forensically relevant SNP classes. BioTechniques. 2008;44(5):603–10.
2. Kayser M, de Knijff P. Improving human forensics through advances in genetics, genomics and molecular biology. Nature Rev Genet. 2011;12(3):179–92.
3. Mehta B, Daniel R, Phillips C, McNevin D. Forensically relevant SNaPshot® assays for human DNA SNP analysis: a review. Int J Legal Med. 2017;131(1):21–37.
4. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, Fondevila M, et al. A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis. 2006;27(9):1713–24.
5. Phillips C, Salas A, Sánchez JJ, Fondevila M, Gómez-Tato A, Álvarez-Dios J, et al. Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Sci Int Genet. 2007;1(3–4):273–80.
6. Walsh S, Liu F, Ballantyne KN, Oven MV, Lao O, Kayser M. IrisPlex: a sensitive DNA tool for accurate prediction of blue and brown eye colour in the absence of ancestry information. Forensic Sci Int Genet. 2011;5:170–80.
7. Mehta B, Daniel R, McNevin D. High resolution melting (HRM) of forensically informative SNPs. Forensic Sci Int Genet Supp Ser. 2013;4(1):e376–7.
8. Venables SJ, Mehta B, Daniel R, Walsh SJ, van Oorschot RAH, McNevin D. Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. Electrophoresis. 2014;35(21–22):3036–43.
9. Reed GH, Kent JO, Wittwer CT. High resolution DNA melting analysis for simple and efficient molecular diagnostics. Pharmacogenomics. 2007;8(6):597–608.
10. Gundry CN, Vandersteen JG, Reed GH, Pryor RJ, Chen J, Wittwer CT. Amplicon melting analysis with labeled primers: a closed-tube method for differentiating homozygotes and heterozygotes. Clin Chem. 2003;49(3):396–406.
11. Liew M, Pryor R, Palais R, Meadows C, Erali M, Lyon E, et al. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clin Chem. 2004;50(7):1156–64.
12. Walsh S, Lindenbergh A, Zuniga SB, Sijenb T. Knijff Pd, Kayser M, et al. developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence. Forensic Sci Int Genet. 2011;5(5):464–71.
13. AppliedBiosystems. ABI PRISM® SNaPshot™ Multiplex Kit. P/N 4323357 Rev. B ed: Thermo Fisher Scientific. 2010;1–42.
14. Walsh S, Chaitanya L, Clarisse L, Wirken L, Draus-Barini J, Kovatsi L, et al. Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Sci Int Genet. 2014;9:150–61.
15. Westen AA, Matai AS, Laros JF, Meiland HC, Jasper M, de Leeuw WJ, et al. Tri-allelic SNP markers enable analysis of mixed and degraded DNA samples. Forensic Sci Int Genet. 2009;3(4):233–41.
16. Swen JJ, Baak-Pablo RF, Guchelaar H-J, van der Straaten T. Alternative methods to a TaqMan assay to detect a tri-allelic single nucleotide polymorphism rs757210 in the HNF1β gene. Clin Chem Lab Med. 2012;50(2):279–84.
17. Fondevila M, Phillips C, Santos C, Freire Aradas A, Vallone PM, Butler JM, et al. Revision of the SNPforID 34-plex forensic ancestry test: assay enhancements, standard reference sample genotypes and extended population studies. Forensic Sci Int Genet. 2013;7(1):63–74.
18. Seipp MT, Durtschi JD, Voelkerding KV, Wittwer CT. Multiplex amplicon genotyping by high-resolution melting. J Biomol Tech. 2009;20(3):160–4.
19. Nicklas JA, Buel E. Improving the efficiency of DNA casework analysis through simple, effective, PCR-based screening methods. NIJ Grant#2005-DA-BX-K003. 2011;1–115.
20. Derzelle S, Laroche S, Le Fleche P, Hauck Y, Thierry S, Vergnaud G, et al. Characterization of genetic diversity of bacillus anthracis in France by using high-resolution melting assays and multilocus variable-number tandem-repeat analysis. J Clin Microbiol. 2011;49(12):4286–92.
21. Bosmali I, Ganopoulos I, Madesis P, Tsaftaris A. Microsatellite and DNA-barcode regions typing combined with high resolution melting (HRM) analysis for food forensic uses: a case study on lentils (*Lens culinaris*). Food Res Int. 2012;46(1):141–7.
22. Ganopoulos I, Bazakos C, Madesis P, Kalaitzis P, Tsaftaris A. Barcode DNA high-resolution melting (bar-HRM) analysis as a novel close-tubed and accurate tool for olive oil forensic use. J Sci Food Agr. 2013;93(9):2281–6.
23. Malewski T, Draber-Mońko A, Pomorski J, Łoś M, Bogdanowicz W. Identification of forensically important blowfly species (Diptera: Calliphoridae) by high-resolution melting PCR analysis. Int J Legal Med. 2010;124(4):277–85.
24. Sijen T. Molecular approaches for forensic cell type identification: on mRNA, miRNA, DNA methylation and microbial markers. Forensic Sci Int Genet. 2015;18:21–32.

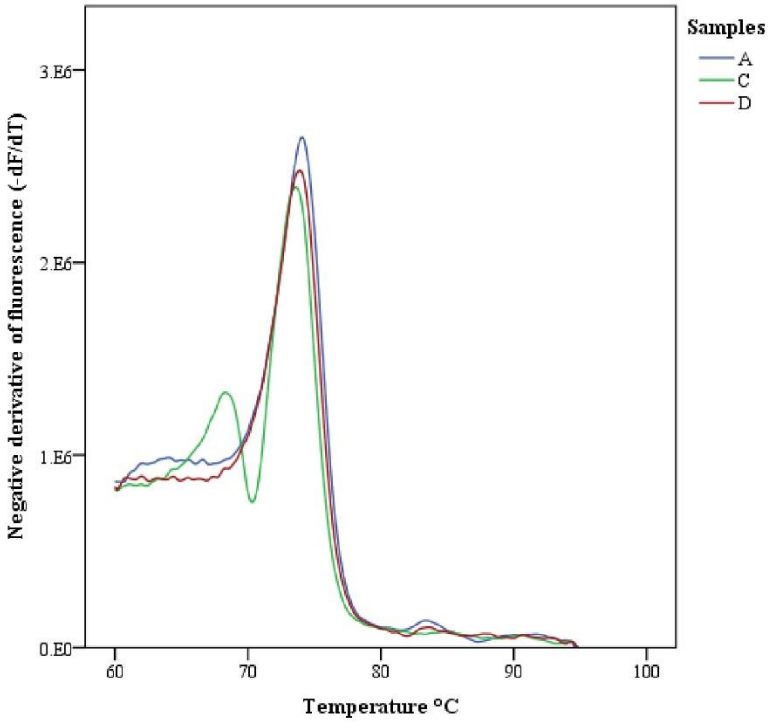
Supplementary Figures



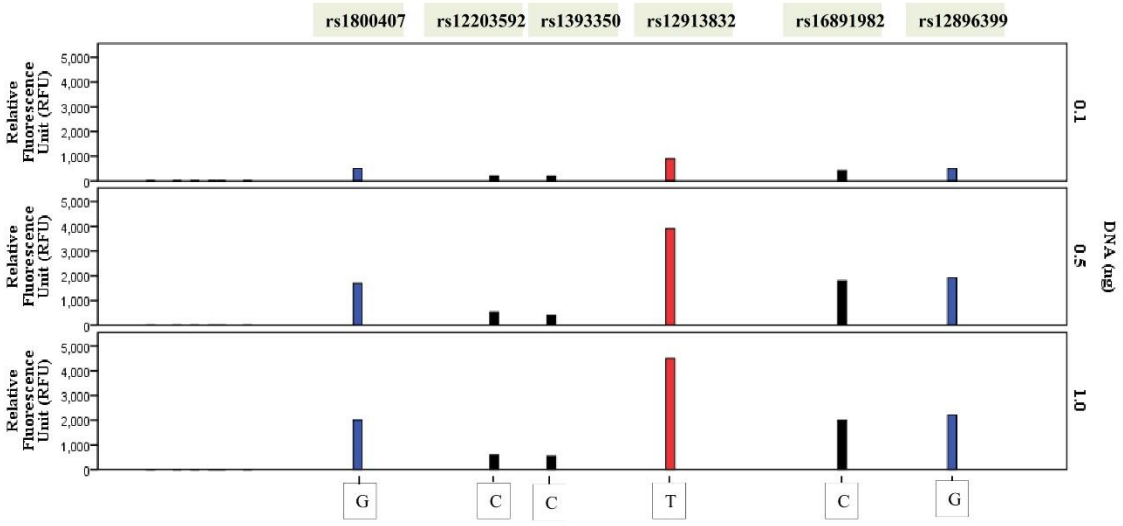
Supplementary Fig. 1: HRM derivative curves for sample D and G at 0.5ng DNA input amount showing homozygous (single peak) and heterozygous (double peak) genotypes using our custom HRM primers. The genotypes of samples D and G are in Table 1



Supplementary Fig. 2: HRM derivative curves for samples A-H for SNP rs1800407 using our custom primers showing secondary peaks due to high GC content in the flanking regions.



Supplementary Fig. 3: HRM derivative curves for samples A, C and D using our custom primer set for the symmetrical (G/C) SNP rs16989182 showing a heterozygous genotype (double peak in green) for sample C. Samples A and D yielded the same homozygous genotype (CC) whereas SNaPshot indicated CC and GG genotypes, respectively.



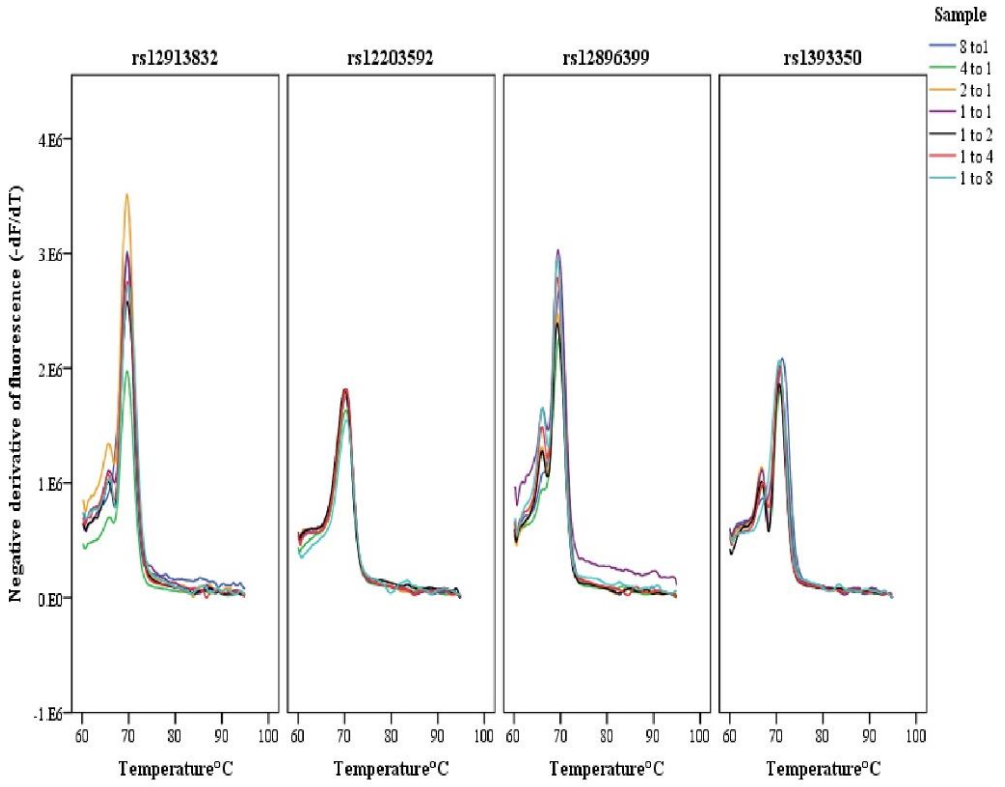
Supplementary Fig. 4: Fluorescence intensity (RFU) for SBE products of SNaPshot assays for sample A at 0.1, 0.5 and 1ng DNA template input amounts (nucleotides are represented as A- green, C-black, G-blue and T- red) A DNA input amount of 10 ng saturated the SNaPshot electropherograms.

SNPs	Genotype concordance between HRM and SNaPshot of samples and replicates																													
Samples	**A**			**B**			**C**			**D**			**E**			**F**			**G**			**H**			**007**			**9947A**		
Replicates	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
rs12913832	Y	Y	N	N	Y	N	N	Y	Y	Y	Y	N	Y	N	Y	N	N	Y	N	N	Y	Y	N	N	N	Y	N	N	Y	Y
rs12203592	Y	Y	N	Y	N	N	Y	N	Y	N	Y	N	N	Y	N	Y	N	N	N	N	Y	N	N	Y	N	N	Y	Y	N	N
rs12896399	Y	Y	N	N	Y	N	N	Y	N	N	Y	N	Y	Y	N	Y	N	Y	N	Y	N	N	N	Y	Y	N	N	N	N	Y
rs1393350	Y	N	Y	Y	N	Y	Y	N	N	Y	N	N	N	Y	N	N	N	Y	N	N	Y	N	N	Y	N	N	Y	N	Y	N
rs1800407	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
rs16891982	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Supplementary Fig. 5: Genotype concordance between HRM and SNaPshot at 0.1ng (intra plate reproducibility) using custom primer set (Y – yes and N – no).

SNPs	Genotype concordance between HRM and SNaPshot of samples and replicates																													
Samples	**A**			**B**			**C**			**D**			**E**			**F**			**G**			**H**			**007**			**9947A**		
Replicates	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	1	2	3
rs12913832	Y	N	N	N	N	Y	N	N	Y	N	Y	N	Y	Y	Y	N	N	Y	N	Y	N	N	Y	N	Y	Y	N	N	N	Y
rs12203592	N	Y	N	Y	N	N	Y	N	N	Y	Y	N	N	N	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	N	N	Y	Y	N
rs12896399	Y	Y	N	N	Y	N	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	N	Y	Y	N	Y	N	Y	N	N	N	Y
rs1393350	N	N	Y	Y	N	Y	Y	N	N	N	N	N	N	N	Y	N	Y	N	N	N	Y	N	N	Y	N	N	Y	N	Y	Y
rs1800407	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A
rs16891982	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A	N/A

Supplementary Fig. 6: Genotype concordance between HRM and SNaPshot at 0.1ng (inter plate reproducibility) using custom primer set (Y - yes and N - no).



Supplementary Fig. 7: HRM derivative curves for mixtures of two samples (A and 007) across all six SNPs for ratios 8:1, 4:1, 2:1, 1:1, 1:2, 1:4 and 1:8 with a total DNA input amount of 1ng using our custom designed primers.

Supplementary Tables

Supplementary Table 1: High Resolution Melting (HRM) assay reaction conditions for the IrisPlex multiplex using custom designed primers.

PCR				**Thermal cycling conditions**		
Component	**[Stock]**	**Vol (µL)**	**[Final]**	**Step**	**Temp (°C)**	**Time**
HRM Master Mix	2 ×	10	1 ×	Activation	95	10 mins
FWD Primer-rs12913832		0.6	0.3 µM	45 × Denaturation	95	15 sec
REV Primer-rs12913832		0.6	0.3 µM	Annealing/Extension	60	60 sec
FWD Primer-rs12203592		0.6	0.3 µM	HRM	95	10 sec
REV Primer-rs12203592		0.6	0.3 µM		60	60 sec
FWD Primer-rs128936399		0.6	0.3 µM		60-95	0.025°C/sec
REV Primer-rs128936399		0.6	0.3 µM		95	15 sec
FWD Primer-rs1393350		0.6	0.3 µM		60	15 sec
REV Primer-rs1393350		0.6	0.3 µM			
FWD Primer-rs1800407		0.6	0.3 µM			
REV Primer-rs1800407		0.6	0.3 µM			
FWD Primer-rs16891982		0.6	0.3 µM			
REV Primer-rs16891982		0.6	0.3 µM			
dH2O		1.8				
Total Master Mix		19				
DNA (1ng)		1				
Total Volume		20				

Supplementary Table 2: Genotypes obtained for a two sample mixture at different ratios using our HRM custom designed primers

Ratios (A:007)	**SNPs**			
	rs12913832 (C/T)	rs12203592 (C/T)	rs12896399 (G/T)	rs1393350 (C/T)
A	**TT**	**CC**	**GG**	**CC**
1:8	CT	CC	GT	CT
1:4	CT	CC	GT	CT
1:2	CT	CC	GT	CT
1:1	CT	CC	GT	CT
2:1	CT	CC	GT	CT
4:1	CT	CC	GT	CT
8:1	CT	CC	GT	CT
007	**CT**	**CC**	**GT**	**TT**

Chapter 4: High-throughput genotyping tools: Illumina MiSeq massively parallel sequencing

4.1 Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37 (21):2832–2840

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 4.1

Declaration by candidate

In the case of Chapter 4.1, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Proposal of the research, experimental work, data collection and analysis and writing the first draft of the manuscript	55

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student UC Y/N
Dennis McNevin	Proposal of the research, revising and editing of the manuscript, critical feedback on the data analysis tools	15	N
Runa Daniel	Proposal of the research, revising and editing of the manuscript, critical feedback on the data analysis tools	15	N
Chris Phillips	Provided research material (reagents), critical feedback on the manuscript completion	5	N
Stephen Doyle	Resources and training in experimental work, critical feedback on the manuscript draft	5	N
Gareth Elvidge	Experimentation and data analysis training, critical feedback on the manuscript draft	5	N

Candidate's Signature		Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	National Centre for Forensic Studies, Faculty of STeM, University of Canberra

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	
06/04/2018	Chris Phillips	

[* Please insert additional rows as needed.]

Candidate's Signature	I.S.M McNEE	Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	National Centre for Forensic Studies, Faculty of STeM, University of Canberra

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	Dennis McN
5/4/18	RUMA DANIEL	Ruma Daniel

[* Please insert additional rows as needed.]

Candidate's Signature		Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author who accepts overall responsibility for the publication;
4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**National Centre for Forensic Studies, Faculty of STeM, University of Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-author	Signature
04/04/2018	Dennis McNevin	
06/04/2018	STEPHEN DOYLE	

[* Please insert additional rows as needed.]

Bhavik Mehta[1]
Runa Daniel[2]
Chris Phillips[3]
Stephen Doyle[4]
Gareth Elvidge[5]
Dennis McNevin[1]

[1]National Centre for Forensic Studies, Faculty of ESTeM, University of Canberra, Canberra, Australia
[2]Office of the Chief Forensic Scientist, Victoria Police Forensic Services Department, Melbourne, Australia
[3]Forensic Genetics Unit, Institute of Legal Medicine, University of Santiago de Compostela, Santiago de Compostela, Spain
[4]Department of Animal, Plant and Soil Sciences, La Trobe University, Melbourne, Australia
[5]Illumina Inc. San Diego, CA, USA

Received April 14, 2016
Revised August 20, 2016
Accepted August 23, 2016

Research Article

Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq

Forensic DNA-based intelligence, or forensic DNA phenotyping, utilises SNPs to infer the biogeographical ancestry and externally visible characteristics of the donor of evidential material. SNaPshot® is a commonly employed forensic SNP genotyping technique, which is limited to multiplexes of 30–40 SNPs in a single reaction and prone to PCR contamination. Massively parallel sequencing has the ability to genotype hundreds of SNPs in multiple samples simultaneously by employing an oligonucleotide sample barcoding strategy. This study of the Illumina MiSeq massively parallel sequencing platform analysed 136 unique SNPs in 48 samples from SNaPshot PCR amplicons generated by five established forensic DNA phenotyping assays comprising the SNP*for*ID 52-plex, SNP*for*ID 34-plex, Eurasiaplex, Pacifiplex and IrisPlex. Approximately 3 GB of sequence data were generated from two MiSeq flow cells and profiles were obtained from just 0.25 ng of DNA. Compared with SNaPshot, an average 98% genotyping concordance was achieved. Our customised approach was successful in attaining SNP profiles from extremely degraded, inhibited, and compromised casework samples. Heterozygote imbalance and sequence coverage in negative controls highlight the need to establish baseline sequence coverage thresholds and refine allele frequency thresholds. This study demonstrates the potential of the MiSeq for forensic SNP analysis.

Keywords:
Forensic DNA genotyping / Illumina MiSeq / Next generation sequencing (NGS) / SNaPshot / SNPs
DOI 10.1002/elps.201600190

Additional supporting information may be found in the online version of this article at the publisher's web-site

1 Introduction

Forensic DNA-based intelligence, commonly known as forensic DNA phenotyping or molecular photofitting [1, 2], utilises genetic markers associated with phenotypes including biogeographical ancestry (BGA) and externally visible characteristics (EVCs) to predict the appearance of the donor of evidential material. Forensic DNA phenotyping is rapidly emerging as a potentially powerful tool in criminal investigations particularly when STR genotyping produces partial or non-informative profiles [1].

The most common approach for forensic SNP genotyping has been single-base extension (SBE) using the SNaPshot® assay (Applied Biosystems), which utilises CE detection [3, 4] and other equipment commonly used in forensic laboratories. Numerous SNP-based forensic intelligence SNaPshot® assays have been developed, including the SNP*for*ID 34-plex [5, 6], Eurasiaplex [7], IrisPlex [8] and HIrisPlex [9]. Some limitations associated with SNaPshot® include an upper multiplexing limit of approximately 30–40 SNPs in a single PCR assay [10] and the need for multiple tube transfers, which increase the risk of contamination [10, 11]. Next-generation sequencing (NGS), alternatively termed massively parallel sequencing (MPS), can simultaneously genotype hundreds of markers in multiple samples using small amounts of DNA. High-throughput MPS platforms, such as the HiSeq (Illumina) and SOLiD (Applied Biosystems) systems, are cost-effective for sequencing whole genomes [12]. Low-to-medium throughput benchtop

Correspondence: Bhavik Mehta, National Centre for Forensic Studies, Level D, Building 7, University of Canberra, Bruce ACT 2617, Australia
E-mail: bhavik.mehta@hotmail.com

Abbreviations: BGA, biogeographical ancestry; **EVCs**, externally visible characteristics; **GQx**, genotyping quality; **HA**, humic acid; **MPS**, massively parallel sequencing; **NGS**, next-generation sequencing; **SBE**, single-base extension; **UV**, ultra violet

sequencers such as the Ion PGM™ (Applied Biosystems) and MiSeq (Illumina) operate at a more appropriate scale for forensic laboratories. Recently, the applicability of the Ion PGM™ for forensic autosomal SNP genotyping has been demonstrated [10, 13]. This study reports on the application of the MiSeq system to genotype autosomal SNPs in a combination of existing customised panels.

The MiSeq employs sequencing by synthesis chemistry. Individual DNA molecules are attached to a glass slide (flow cell) and clonally amplified in clusters via bridge PCR [14, 15]. The MiSeq can generate up to 15 GB (approximately 25 million reads) of sequence data on a single flow cell (version 3) and can be applied to targeted sequencing of forensically informative markers [12]. This has been demonstrated on the forensic specific MiSeq FGx™ (Forensic Genomics System) with a beta version of the ForenSeq™ DNA Signature Prep Kit (Illumina) [16].

In this study, SNPs from five forensic SBE assays were combined and genotyped on the MiSeq. These were the SNP*for*ID 52-plex for identity [17]; the SNP*for*ID 34-plex [6], Eurasiaplex [7] and Pacifiplex [18] for BGA; and IrisPlex [8] as an EVC predictive test. Six forensic validation parameters were examined: sensitivity, reproducibility, genotype concordance, effect of different DNA extraction methods, ability to genotype compromised samples including bone and humic acid (HA) inhibited extracts and ability to genotype ultra violet (UV) degraded extracts.

2 Materials and methods

2.1 Sample preparation

Ethics approval to collect DNA for this study was granted by the University of Canberra Committee for Ethics in Human Research (project number 11–119 and its extension, 15–64). Seven human DNA templates (S1–S7) were extracted from buccal swabs using the DNA-IQ™ System (Promega) following the manufacturer's recommended protocol. Extracts were quantified using Quantifiler™ Human DNA Quantitation Kit (Applied Biosystems) following the manufacturer's recommended protocol together with the two standard reference materials: human male cell line control DNA 007 (Applied Biosystems) and human female cell line control DNA 9947A (Applied Biosystems).

2.2 Preparation of PCR amplicons

PCR amplicons were generated using published primer sequences and reaction protocols for five forensic multiplex PCR assays: SNP*for*ID 52-plex, SNP*for*ID 34-plex, Eurasiaplex, Pacifiplex and IrisPlex. The five multiplex assays together comprise 145 SNP amplicons with nine SNPs (rs1024116, rs1335873, rs12913832, rs16891982, rs1886510, rs204041, rs3827760, rs722098 and rs917118) shared in multiple assays resulting in 136 unique SNP amplicons ranging from 51 to 156 bp.

2.3 SNaPshot® genotyping of PCR amplicons

SNaPshot® genotyping was performed following the published protocols for each assay [6–8, 17, 18] to assess the efficiency of the PCR reactions used to generate the amplicons for sequencing and to obtain genotypes for concordance studies.

2.4 Forensic validation parameters

The study assessed the following six forensic validation parameters.

2.4.1 Sensitivity

A sensitivity study was conducted on three DNA templates: 9947A, 007, and S1 using template input amounts for each multiplex PCR assay of 0.05, 0.1, 0.2, 0.3 and 0.5 ng (total of 0.25, 0.5, 1.0, 1.5, 2.5 ng).

2.4.2 Reproducibility

Four replicates of sample S1 at 0.3 ng for each multiplex PCR assay (total of 1.5 ng) were used to assess reproducibility.

2.4.3 Genotype concordance

Seven human DNA templates (S1–S7), 9947A and 007 were genotyped on the MiSeq at 0.5 ng for each multiplex PCR assay (total of 2.5 ng) and the resulting genotypes compared with those from SNaPshot®.

2.4.4 Effect of different DNA extraction methods

Sample S2 was extracted using three DNA extraction methods: DNA-IQ™ System (Promega), Isolate II (Bioline), both following the manufacturer's recommended protocols, and a standard phenol/chloroform extraction with ethanol precipitation [19]. The three DNA extracts were genotyped at 0.5 ng for each multiplex PCR assay (total of 2.5 ng) in the same MiSeq run and genotype concordance between the samples was assessed.

2.4.5 Effect of UV irradiation

A one-step UV degradation method was adapted for generating artificially degraded samples [20]. Aliquots (of 5 μL) DNA extracts of S2 and S3 (0.5 ng/μL in 0.2 mL) in PCR tubes were exposed to UV light for 30- and 60-min intervals. The UV light was generated from a 10 W source (Sankyo Denki, 254 nm, UV-C) at a distance of 13 cm. PCR amplicons were generated from 1 μL of each irradiated sample for each time interval.

2.4.6 Effect of HA inhibition

HA was used to mimic the inhibition encountered in casework samples from items such as soil. HA (Sigma-Aldrich) at 75 and 100 ng quantities were added to the PCR reactions of samples S2 and S3 (both at 0.5 ng/μL).

2.4.7 Compromised casework samples

DNA extracts from five aged bone samples (S8, S9, S10, S11 and S12) and one aged blood sample (S13) were obtained from three forensic laboratories. Due to limited sample availability, they were only submitted for four multiplex PCR assays: 34-plex, Eurasiaplex, Pacifiplex and IrisPlex (total 93 SNPs). The aged bone samples (S10 and S11) were recovered from Papua New Guinea (suspected to be from World War II). The aged blood sample S13 had been stored at room temperature for 45 years.

2.5 MiSeq MPS library preparation

PCR products (2 μL) from the five multiplex PCR assays for each template were pooled together (10 μL total). A 5 μL aliquot of the pool was used for library preparation. The PCR negative controls from each assay were also pooled. The pooled templates were subjected to the TruSeq® ChIP ligation (Illumina) library preparation step following the manufacturer's protocols. Normalisation of the barcoded libraries was based on quantitation using Qubit (Applied Biosystems) following the manufacturers' recommended protocol. The normalised libraries were pooled into a final concentration of 10 nM.

2.6 MiSeq sequencing template preparation

The 10 nM barcoded library pool was diluted to 4 nM and denatured using 0.2 N NaOH following the manufacturer's recommended protocol [21]. The denatured library was further diluted to 1 pM for loading onto the MiSeq sequencing cartridge. The sequencing control comprised the phiX control library (Illumina). A volume of 600 μL of the 1 pM barcoded denatured library with 5% (v/v) 12.5 pM phiX control was sequenced using the MiSeq v3 600 cycles sequencing kit [21–23]. Paired-end sequencing was performed using a 2 × 101 bp cycle setting. Two flow cells were used to sequence 24 samples per chip (Supporting Information Table S1).

2.7 MiSeq data analysis

Image processing, base calling and base quality scoring were performed with MiSeq Control Software v. 2.5 (Illumina) using default parameters. The MiSeq Reporter software had a default upper limit coverage reporting maximum of 5000 reads per amplicon. The human reference genome hg19 (GRCh37) was used for alignment [24] and sequence output was generated in binary alignment/map format. The binary alignment/map files were used to generate variant calling and genome variant calling files for each sample. The variant calling files were analysed by VariantStudio (v2.1) variant analysis software (Illumina) to generate Excel output files. The Excel and genome variant calling files provided the input for our custom macros to obtain coverage data for each nucleotide (Supporting Information File S1).

2.8 Allele calls

MiSeq Reporter variant caller default allele call thresholds (80% allele frequency or greater for homozygotes and between 20 and 80% for heterozygotes) and parameters including genotyping quality (GQx) scores were employed. GQx is a phred-scale confidence score for genotype designation [25, 26]. No baseline coverage thresholds were applied. SNPs with no genotype calls and genotypes with GQx < 99 were categorised collectively as 'missing' and discounted from further analysis.

2.9 Statistical analysis

Non-parametric statistical tests using the IBM SPSS package (v. 21) were applied to the data due to skewed (non-normal) amplicon coverage distribution. A Mann–Whitney *U*-test was used to assess the null hypothesis of no significant difference in amplicon length (bp) and GC content (%) between amplicons with the highest and lowest 10% combined coverage for three templates (9947A, 007 and S1). Kendall's tau and Spearman's rho rank correlation coefficients were determined to identify any correlation between amplicon length and GC content over the entire coverage distribution for each template (9947A, 007 and S1). A Kruskal–Wallis test was employed to assess the null hypothesis of no significant difference in combined coverage distribution across all three templates in each assay. A Kruskal–Wallis test was also used to test the null hypothesis of no difference between the combined coverage across all four replicates (of S1 at 1.5 ng) in each assay.

3 Results

The 48 samples, sequenced on two MiSeq flow cells, generated 29.5 million reads in total. Allele frequency variation is compared with depth of coverage in Supporting Information Fig. S1 for samples 9947A, 007 and S1.

3.1 Sensitivity

MiSeq genotype concordance between template amounts (0.25, 0.5, 1.0, 1.5 and 2.5 ng) for 9947A, 007 and S1 was

97.6, 99.3 and 97.0%, respectively. The genotype concordance between SNaPshot® and MiSeq was 96.0–99.3% across all template amounts and samples. Genotypes from the same three samples at 0.5, 1.0 and 1.5 ng template amounts were compared with previously obtained Ion PGM™ genotypes reported by Daniel et al. [10] with concordance between 97 and 100%. The percentage of missing data for 9947A, 007 and S1 ranged from 0.7 to 9.5% across all template amounts (Supporting Information Table S2). On average, MiSeq genotypes were 97.7 and 98.5% concordant with SNaPshot® and Ion PGM™, respectively (Supporting Information Table S3).

3.2 Reproducibility

A Kruskal–Wallis test confirmed that there was a significant difference in coverage distribution across all four replicates ($p = 0.000$). Reproducibility was 97.2–99.3% (with 4/145 and 1/145 SNPs not present in one sample and three samples, respectively: Supporting Information Table S4). However, excluding missing SNPs, genotypes between all four replicates of S1 (at 1.5 ng) were 100% concordant.

3.3 Genotype concordance

Missing SNaPshot® genotypes are shown in Supporting Information Table S5. Excluding these missing genotypes, there were between zero and five discordant SNPs for S1, S2, S3, S4, S5, S6, S7, 9947A and 007, respectively (Supporting Information Table S5). The overall concordance between MiSeq and SNaPshot® genotypes ranged from 96.5 to 100.0% for all samples (Supporting Information Table S6).

3.4 Effect of different DNA extraction methods

The three different extractions of sample S2 (DNA IQ, Isolate II and phenol/chloroform) had two, three and two missing SNPs, respectively (Supporting Information Table S7). Excluding the missing SNPs, 100% genotype concordance was observed between all three extracts.

3.5 Effect of UV irradiation

Exposure of sample S2 to UV for 30 and 60 min resulted in 54.5 and 29.0% reportable SNPs, respectively (Fig. 1). Sample S3 yielded 60.0 and 30.0% reportable SNPs at 30 and 60 min UV exposure, respectively (Fig. 1 and Supporting Information Table S7). Genotype concordance between UV exposed samples and corresponding original samples are indicated in Supporting Information Table S7. Non-concordant genotypes ranged from 5.0 to 20.0% and were solely due to loss of alleles in the UV exposed samples (an example is shown in Supporting Information Fig. S2). SNaPshot® genotyping of sample S2 exposed to 60 min of UV was unsuccessful using the 52- and 34-plex assays, whereas the MiSeq produced partial profiles under these extreme UV degradation conditions.

3.6 Effect of HA inhibition

Sample S2 spiked with 75 and 100 ng of HA returned 65.5 and 60.0% reportable SNPs, respectively (Fig. 1). Sample S3 similarly returned 69.5 and 59.0% SNPs, respectively (Fig. 1). Excluding missing SNPs, the genotype concordance between the HA-inhibited samples and the original samples ranged between 99.0 and 100.0% (Supporting Information Table S7). Figure 2 shows multiplex PCR assays containing BSA in their PCR reaction mixes (52- and 34-plex) generated better coverage compared to those assays without BSA. However, no SNaPhsot® SNP profiles were generated with sample S2 spiked with 100 ng HA using the SBE 52- and 34-plex assays (data not shown).

3.7 Compromised casework samples

Samples S8, S10 and S11 were below the detection limit for Quantifiler, but produced 4, 10 and 12% reportable SNPs (out of a total of 93), respectively (Table 1). Samples S9, S12 and S13, with DNA concentrations either undetected or less than 0.01 ng/μL, gave 89, 92 and 86% reportable SNPs, respectively (Table 1).

3.8 Amplicon coverage bias

Inconsistent and skewed coverage between different amplicons was observed for all templates. In samples 9947A, 007 and S1, 66% of amplicons with the highest 10% of combined coverage across all template amounts were common to all three. Similarly, all the amplicons with the lowest 10% of coverage were common to all three. Table 2 shows the SNPs with the lowest and highest 10% of combined coverage across the three samples at all template amounts. The data indicate that coverage is amplicon-dependent with rarely sequenced amplicons in common across templates and template amounts, and highly sequenced amplicons also in common. The effect of amplicon length, GC content and PCR assay on coverage bias was subsequently examined.

3.8.1 Effect of amplicon length on coverage

Sequence coverage as a function of amplicon length for samples 9947A, 007 and S1 at five template input amounts is shown in Supporting Information Fig. S3. Amplicon lengths for the highest and lowest 10% of amplicons by coverage ranged from 86 to 118 bp and 51 to 156 bp, respectively, with medians of 93 and 83 bp, respectively (Table 2). A Mann–Whitney *U*-test rejected the hypothesis that there was no difference between the amplicon lengths of the SNPs with the highest and lowest 10% of combined coverage ($p = 0.040$). In

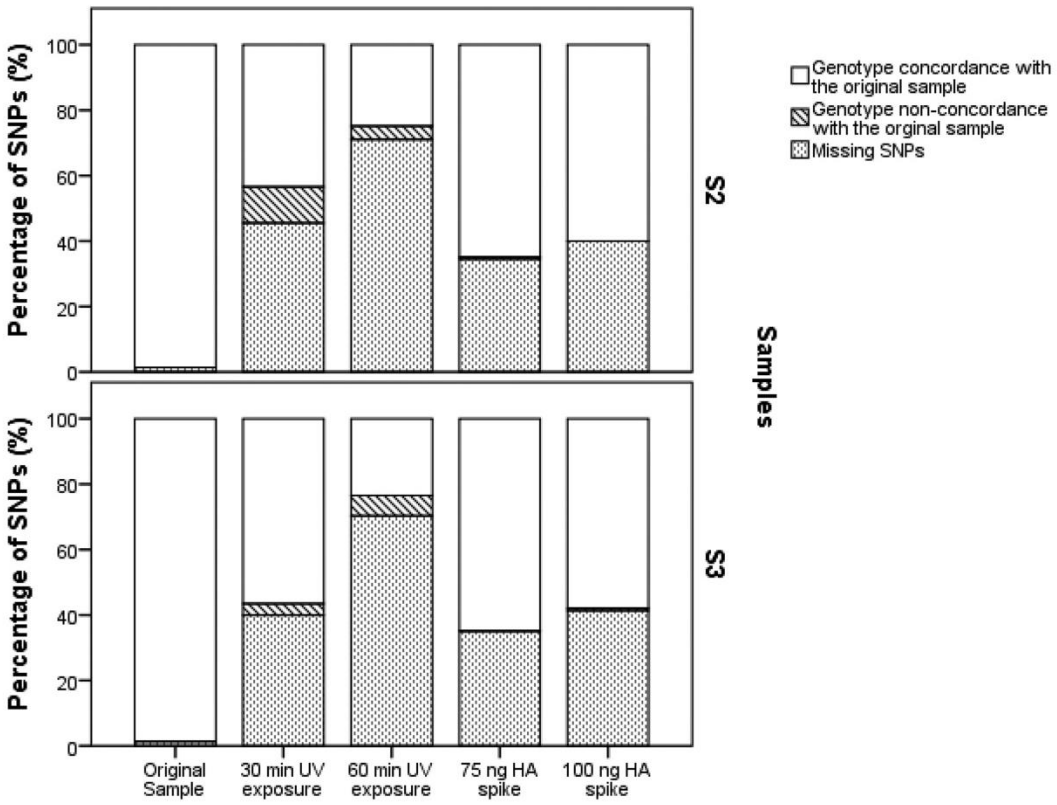


Figure 1. The effects of UV irradiation (30 and 60 min) and humic acid (HA at 75 and 100 ng) on genotype concordance (as a percentage of a total of 136 unique SNPs) for samples S2 and S3.

addition, Kendall's tau and Spearman's rho tests performed on the entire coverage distribution showed a weak correlation between amplicon length and coverage ($r^2 = 0.251$ with $p = 0.000$ and $r^2 = 0.354$ with $p = 0.000$, respectively).

3.8.2 Effect of GC content on coverage

Sequence coverage as a function of amplicon GC content for samples 9947A, 007 and S1 at five template input amounts is shown in Supporting Information Fig. S4. The average GC content across the amplicons was 44%. The GC content of the highest and lowest 10% of amplicons by coverage ranged from 37 to 51% and 31 to 47%, respectively, with medians of 45 and 43% (Table 2). A Mann–Whitney *U*-test supported the hypothesis of no difference in GC content between the amplicons with the highest and lowest 10% of combined coverage ($p = 0.436$). In addition, Kendall's tau and Spearman's rho tests performed on the entire coverage distribution indicated that there was no significant correlation between GC content and coverage ($r^2 = -0.013$ with $p = 0.410$ and $r^2 = -0.017$ with $p = 0.437$, respectively). Thus, there is no evidence to support the hypothesis that coverage bias is associated with GC content.

3.8.3 Effect of PCR assay on coverage

A skewed distribution of coverage was observed within each multiplex PCR assay (Supporting Information Fig. S5) as well as inconsistency in representation of assays between samples. For example, the 52-plex was under-represented in 007 at 1.0 ng and Eurasiaplex was under-represented in 9947A at 1.5 ng. Seven of the nine amplicons with the highest 10% coverage originated from Eurasiaplex with a combined coverage range from 16 442× to 25 000×, indicating Eurasiaplex amplicons were generally over-represented (Table 2). A Kruskal–Wallis test rejected the hypothesis of no difference in combined coverage (across all five template amounts for samples 9947A, 007 and S1) between amplicons from different multiplex assays ($p = 0.000$) indicating significant differences in SNP coverage from different assays in all samples.

ELECTROPHORESIS

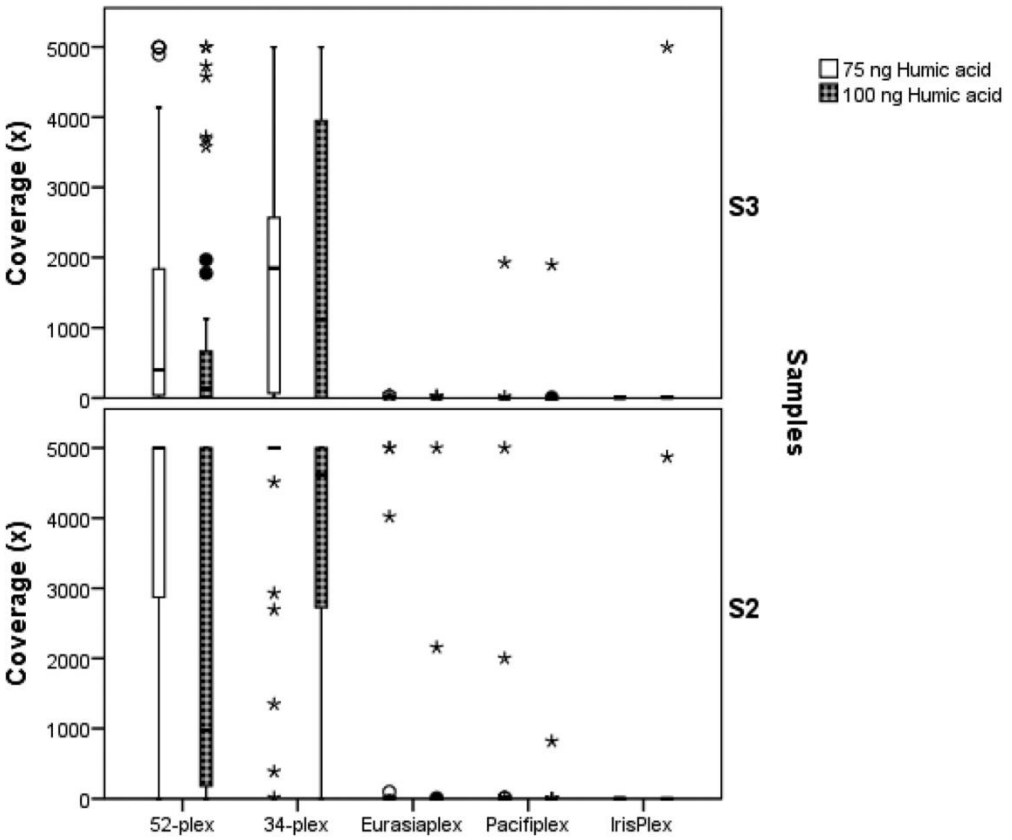


Figure 2. Coverage distributions in each of the five multiplex PCR assays spiked with humic acid (HA) at 75 and 100 ng for samples S2 and S3. The 52-plex and 34-plex assays both contain BSA in their PCR reaction mix.

3.9 Negative control

Eleven SNPs were observed in the negative control with nine of these from the 52-plex assay (Supporting Information Table S8). The coverage ranged from 69× to 5000×. Except for five SNPs (rs1355366, rs1463729, rs1028528, rs734482 and rs2227203), all other genotypes corresponded to one or more possible templates used in the run (9947A, 007, S1 and S2). Daniel et al. [10] also observed coverage of SNPs in a negative control with the same three samples (9947A, 007 and S1) sequenced on the Ion PGM™.

3.10 Cost estimate

Genotyping costs were estimated to be US $1.4 per SNP per sample based on library preparation and sequencing reagent costs only (Supporting Information Table S9). One Miseq v3 flow cell has the capacity to genotype approximately 10 000 SNPs per sample at 100× coverage when genotyping 24 samples in a run. Thus, per SNP costs could be further reduced by adding more markers and/or more samples per Miseq run.

4 Discussion

This study demonstrates the potential of the MiSeq as a medium throughput MPS platform for forensic analysis using modular, customised SNP panels that are already established as sensitive forensic assays. The 3 GB of sequence data obtained from two runs allowed us to obtain SNP genotypes for identity, BGA and EVC inferences from 48 samples.

MiSeq sequencing using a pool of five non-commercial multiplex PCR assays produced uneven sequence coverage that was also observed for the same amplicons by the Ion PGM™ system [10]. The overall sequence data profile has the following characteristics: (a) uneven coverage of amplicons between multiplex PCR assays, (b) uneven coverage within each assay and (c) non-normal (skewed) distribution of coverage (Supporting Information Fig. S5). High and low

Table 1. Reportable SNPs (from a total of 93) for the compromised samples

Sample	Substrate	Quantity (ng/μL)	Percentage of reportable SNPs (%)
S8	Bone	Undetected	4
S9	Bone	0.006	89
S10	Bone	Undetected	11
S11	Bone	Undetected	13
S12	Bone	0.004	92
S13	Blood	Undetected	86

coverage is consistent between amplicons, suggesting that coverage bias is not random but amplicon-dependent. Additionally, coverage was weakly associated with amplicon length (Supporting Information Fig. S3), possibly due to sequence length bias in the magnetic bead cleanup steps of library preparation, favouring longer amplicons.

GC content is often implicated as a source of coverage bias in MPS and associated library preparation [27]. Several studies have identified GC bias in MiSeq sequencing [27–29]; however, we did not encounter such bias (Supporting Information Fig. S4). This may be because the amplicons sequenced in this study were from optimised SNaPshot® assays where optimal GC content of both primer and amplicon sequence had been important considerations during primer design. Also, there were no extremes of GC content in our amplicons (28–65%). This result matches a lack of detectable GC bias in previous Ion PGM™ sequencing of the same templates [10].

The most likely reason for the observed amplicon coverage bias is the amplification imbalance between and within each of the five multiplex assays. The bias may have arisen from differing amplicon representation between multiplexes prior to library preparation, with Eurasiaplex clearly showing over-representation (Supporting Information Fig. S5). SNaPshot® PCR assay protocols were applied without modification for amplicon generation and were not optimised for MPS. Furthermore, PCR products from the five assays were pooled in equal volumes, whereas equimolar pooling may have reduced the imbalance between PCR assays. Any bias within assays may be addressed by further fine-tuning of primer concentrations. Nevertheless, the equal volume pooling strategy used resulted in high genotype concordance with SNaPshot® for both the MiSeq and Ion Torrent [10], suggesting that this approach can be utilised effectively without investing resources in amplicon quantitation or multiplex optimisation to achieve balanced amplicon production.

Baseline coverage thresholds were not applied as discounting genotypes with GQx $<$ 99 resulted in filtering out most of the genotypes with less than 20× coverage. Genotype non-concordance did not appear to be related to low coverage (Supporting Information Table S5) and contaminating alleles in the negative control were similarly not related to low coverage (Supporting Information Table S8: only one of 11 contaminating alleles with less than 250× coverage). The genotypes of six of 11 SNPs observed in the negative controls corresponded to those of one or more templates that could be attributed to contamination between samples. However, the unmatched genotypes of the other five SNPs must either have been due to external DNA or PCR/sequencing errors. In this

Table 2. SNPs with lowest and highest 10% combined coverage across all template amounts for 9947A, 007 and S1

SNP	GC content (%)	Amplicon length (bp)	Multiplex	9947A (×)	007 (×)	S1 (×)
Lowest 10% coverage						
rs3785181	47	156	34-plex	82	126	314
rs2069945	43	83	Pacifiplex	47	137	876
rs1357617	44	90	52 Auto 1	69	111	970
rs12434466	33	51	Pacifiplex	168	155	998
rs239031	47	70	34-plex	222	384	1124
rs876724	40	83	52 Auto 1	107	108	1401
rs2046361	28	79	52 Auto 1	212	92	3025
rs2274636	46	81	Pacifiplex	253	446	1652
rs826472	31	85	52 Auto 1	269	420	4483
Median	43	83				
Highest 10% coverage						
rs9809818	45	89	Pacifiplex	16 416	18 431	25 000
rs39897	50	78	Eurasiaplex	16 442	21 083	25 000
rs1544656	46	90	Eurasiaplex	17 418	24 999	25 000
rs1519654	51	86	Eurasiaplex	20 004	25 000	25 000
rs10008492	47	94	Eurasiaplex	20 008	24 999	25 000
rs2196051	34	115	Eurasiaplex	20 013	25 000	25 000
rs734482	38	118	Eurasiaplex	20 022	25 000	25 000
rs17625895	41	99	Eurasiaplex	20 023	24 999	25 000
rs354439	37	93	52 Auto 2	20 264	18 432	25 000
Median	45	93				

study, the single negative control was a pool of the negative controls from individual multiplex PCR assays and sequencing those individual negative controls would have been more informative. Thus, while this approach is not sufficient to evaluate baseline coverage thresholds, it is an informative preliminary study of negative controls on this platform for any customised forensic assay designs.

This customised MiSeq approach was sensitive enough to provide reliable genotypes with a total template amount as low as 0.25 ng (0.05 ng for each multiplex assay) using default allele frequency thresholds, yielding an average overall concordance of 98% with SNaPshot and Ion Torrent data (for 9947A, 007 and S1). The sensitivity study was performed on the same three templates with the same five PCR assays employed by a similar Ion PGM study [10] as a means of comparison and is indicative only. Daniel et al. [10] similarly found >98% genotype concordance down to 0.1 ng template amount per assay (0.5 ng total). Greater resolution would require the use of replicates for each dilution.

The reproducibility study indicated a high genotype concordance between the four replicates (100% concordant, excluding missing SNPs). This occurred in spite of a significant difference in coverage distribution between the replicates, the probable result of highly uneven amplicon coverage. Genotyping of rs1592672 consistently failed in all replicates and the primers for this SNP may require redesigning if this customised assay was to be routinely employed. One of the replicates failed to produce genotypes for a further three SNPs (rs1357617, rs188650 and rs938283) in a total of 136 unique SNPs.

Genotype concordance between MiSeq and SNaPshot for 9947A, 007 and S1 to S7 ranged from 97.8 to 100%. Applying stringent allele frequency thresholds (such as 95% for homozygotes and 40–60% for heterozygotes) may increase the already high concordance by eliminating some of the ambiguous allele frequencies shown in Fig. 1 (in the range 10–30% and 70–90%). However, this may reduce the number of usable reads and genotypes [10].

Some of the MiSeq SNP genotypes that were non-concordant with SNaPshot® were concordant with Sanger genotypes (from data in Daniel et al. [10]). Two SNPs (rs1029407 and rs717302) were non-concordant across all three platforms (MiSeq, SNaPshot and Sanger) likely due to homopolymeric stretches in flanking regions prompting misalignments. MPS is known to perform poorly in homopolymeric regions [27]. SNP rs1029407 has been mistyped by the Ion PGM™ [12] as well as by the MiSeq in this study (Supporting Information Table S5), whereas the GAIIx (Illumina) MPS platform has produced a correct AA genotype for 9947A in another study [24]. The MiSeq omitted a single base in the flanking homopolymer region, which the alignment algorithm then mis-aligned (Supporting Information Fig. S6), whereas the GAIIx alignment software could align the sequences properly and call the correct genotype [24]. This provides further evidence that SNP mistyping in markers sited in homopolymeric regions can occur from misalignment as well as from mis-incorporation of nucleotides.

MiSeq genotyping was not affected by the methods used to extract DNA and was able to produce partial SNP profiles from samples exposed to 60 min of UV radiation and 100 ng of HA, whereas SNaPshot® SNP genotyping and standard STR profiling failed to detect any alleles in these samples. In addition, this approach successfully typed degraded casework (compromised) samples, producing genotypes for up to 92% of SNPs for aged blood and bones, when real-time PCR quantitation using Quantifiler failed to detect DNA in most cases (< 0.01 ng/μL in all cases). This demonstrates the robustness and applicability of MiSeq using customised SNP panels for highly degraded and inhibited sample analysis typical of disaster victim identification and exhumed remains.

This customised approach offers modularity and flexibility to add and subtract SNP panels providing better ancestry resolution (to sub-population level) and EVC inclusion in contrast to the commercially available ForenSeq™ DNA Signature Kit (Illumina) consisting of only 56 ancestry informative SNPs, which limits ancestry resolution to three or four continental populations only. MPS is a potential DNA-based intelligence tool that can type a large battery of forensically informative markers with consequent reduction in inter-run variability, cost, time and effort.

4.1 Conclusions

MiSeq MPS employing customised, modular SNP panels has been demonstrated here to be able to genotype more than 100 identity, BGA and EVC markers simultaneously in multiple samples. This offers the potential to maximise the use of scarce evidentiary material in comparison to the standard SNaPshot® genotyping. In addition, our customised method provides the option of adding optimised marker sets to increase the resolution and accuracy of ancestry and phenotype prediction in a single run. Future work should be conducted to evaluate the baseline coverage thresholds that may increase genotyping accuracy.

The authors gratefully acknowledge technical support and consumables from Illumina, Inc. and financial support from the Australian Research Council (LP110100121 – From genotype to phenotype: Molecular photofitting for criminal investigations). S. R. D. was supported by an Illumina MiSeq Grant. The compromised forensic samples were provided by Kirsty Wright (School of Natural Sciences, Griffith University), Jodie Ward (Forensic & Analytical Science Services, NSW) and David Bruce (Forensic & Analytical Science Services, NSW). We also acknowledge the Unrecovered War Causalities-Army (UWC-A) unit and Jeremy Austin (Australian Centre for Ancient DNA, Adelaide) for their support with aged bone sample recovery.

The authors have declared no conflict of interest.

5 References

[1] Butler, K., Peck, M., Hart, J., Schanfield, M., Podini, D., *Forensic Sci. Int. Genet. Suppl. Ser.* 2011, *3*, e498–e499.

[2] Kayser, M., de Knijff, P, *Nat. Rev.* 2011, *12*, 179–192.

[3] Sobrino, B., Brion, M., Carracedo, A., *Forensic Sci. Int.* 2005, *154*, 181–194.

[4] Pati, N., Schowinsky, V., Kokanovic, O., Magnuson, V., Ghosh, S., *J. Biochem. Biophys. Methods* 2004, *60*, 1–12.

[5] Phillips, C., Salas, A., Sánchez, J. J., Fondevila, M., Gómez-Tato, A., Álvarez-Dios, J., Calaza, M., de Cal, M. C., Ballard, D., Lareu, M. V., Carracedo, Á., *Forensic Sci. Int. Genet.* 2007, *1*, 273–280.

[6] Fondevila, M., Phillips, C., Santos, C., Freire Aradas, A., Vallone, P. M., Butler, J. M., Lareu, M. V., Carracedo, Á., *Forensic Sci. Int.: Genet.* 2013, *7*, 63–74.

[7] Phillips, C., Aradas, A. F., Kriegel, A. K., Fondevila, M., Bulbul, O., Santos, C., Rech, F. S., Carceles, M. D. P., Carracedo, Á., Schneider, P. M., Lareu, M. V., *Forensic Sci. Int. Genet.* 2013, *7*, 359–366.

[8] Walsh, S., Lindenbergh, A., Zuniga, S. B., Sijenb, T., Knijff, P. d., Kayser, M., Ballantyne, K. N., *Forensic Sci. Int. Genet.* 2011, *5*, 464–471.

[9] Walsh, S., Liu, F., Wollstein, A., Kovatsi, L., Ralf, A., Kosiniak-Kamysz, A., Branicki, W., Kayser, M., *Forensic Sci. Int. Genet.* 2013, *7*, 98–115.

[10] Daniel, R., Santos, C., Phillips, C., Fondevila, M., van Oorschot, R. A. H., Carracedo, Á., Lareu, M. V., McNevin, D., *Forensic Sci. Int. Genet.* 2015, *14*, 50–60.

[11] Mehta, B., Daniel, R., McNevin, D., *Forensic Sci. Int. Genet. Suppl. Ser.* 2013, *4*, e376–e377.

[12] Berglund, E. C., Kiialainen, A., Syvänen, A.-C., *Investig. Genet.* 2011, *2*, 23–23.

[13] Børsting, C., Fordyce, S. L., Olofsson, J., Mogensen, H. S., Morling, N., *Forensic Sci. Int. Genet.* 2014, *12*, 144–154.

[14] Grada, A., Weinbrecht, K., *J. Invest. Dermatol.* 2013, *133*, e11–e14.

[15] Illumina, Illumina Sequencing Technology, Technology Spotlight: Illumina Sequencing, San Diego, CA, 2010.

[16] Churchill, J. D., Schmedes, S. E., King, J. L., Budowle, B., *Forensic Sci. Int. Genet.* 2016, *20*, 20–29.

[17] Sanchez, J. J., Phillips, C., Børsting, C., Balogh, K., Bogus, M., Fondevila, M., Harrison, C. D., Musgrave-Brown, E., Salas, A., Syndercombe-Court, D., Schneider, P. M., Carracedo, A., Morling, N., *Electrophoresis* 2006, *27*, 1713–1724.

[18] Santos, C., Phillips, C., Fondevila, M., Daniel, R., van Oorschot, R. A., Burchard, E. G., Schanfield, M. S., Souto, L., Uacyisrael, J., Via, M., *Forensic Sci. Internat. Genet.* 2016, *20*, 71–80.

[19] Köchl, S., Niederstätter, H., Parson, W., *Forensic DNA Typing Protocols*, Springer, NY, USA, 2005, pp. 13–29.

[20] Pang, B. C. M., Cheung, B. K. K., *Anal. Biochem.* 2007, *360*, 163–165.

[21] Illumina, MiSeq System User Guide, Rev O.

[22] Illumina, Preparing sequencing libraries for Loading on a MiSeq, Rev B.

[23] Illumina, TruSeq CHIP sample preperation guide, Rev A.

[24] GenomeReferenceConsortium, Human Genome Overview: GRCh37, National Institutes of Health, MD, USA, 2013.

[25] Li, H., Durbin, R., *Bioinformatics* 2010, *26*, 589–595.

[26] McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., *Genome Res.* 2010, *20*, 1297–1303.

[27] Ross, M. G., Russ, C., Costello, M., Hollinger, A., Lennon, N. J., Hegarty, R., Nusbaum, C., Jaffe, D. B., *Genome Biol.* 2013, *14*, R51–R70.

[28] Quail, M. A., Smith, M., Coupland, P., Otto, T. D., Harris, S. R., Connor, T. R., Bertoni, A., Swerdlow, H. P., Gu, Y., *BMC Genom.* 2012, *13*, 341–353.

[29] Chen, S., Li, S., Xie, W., Li, X., Zhang, C., Jiang, H., Zheng, J., Pan, X., Zheng, H., Liu, J. S., *PloS One.* 2014, *9*, 9–16.

Supplementary figures

Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37 (21):2832–2840

Figure S1

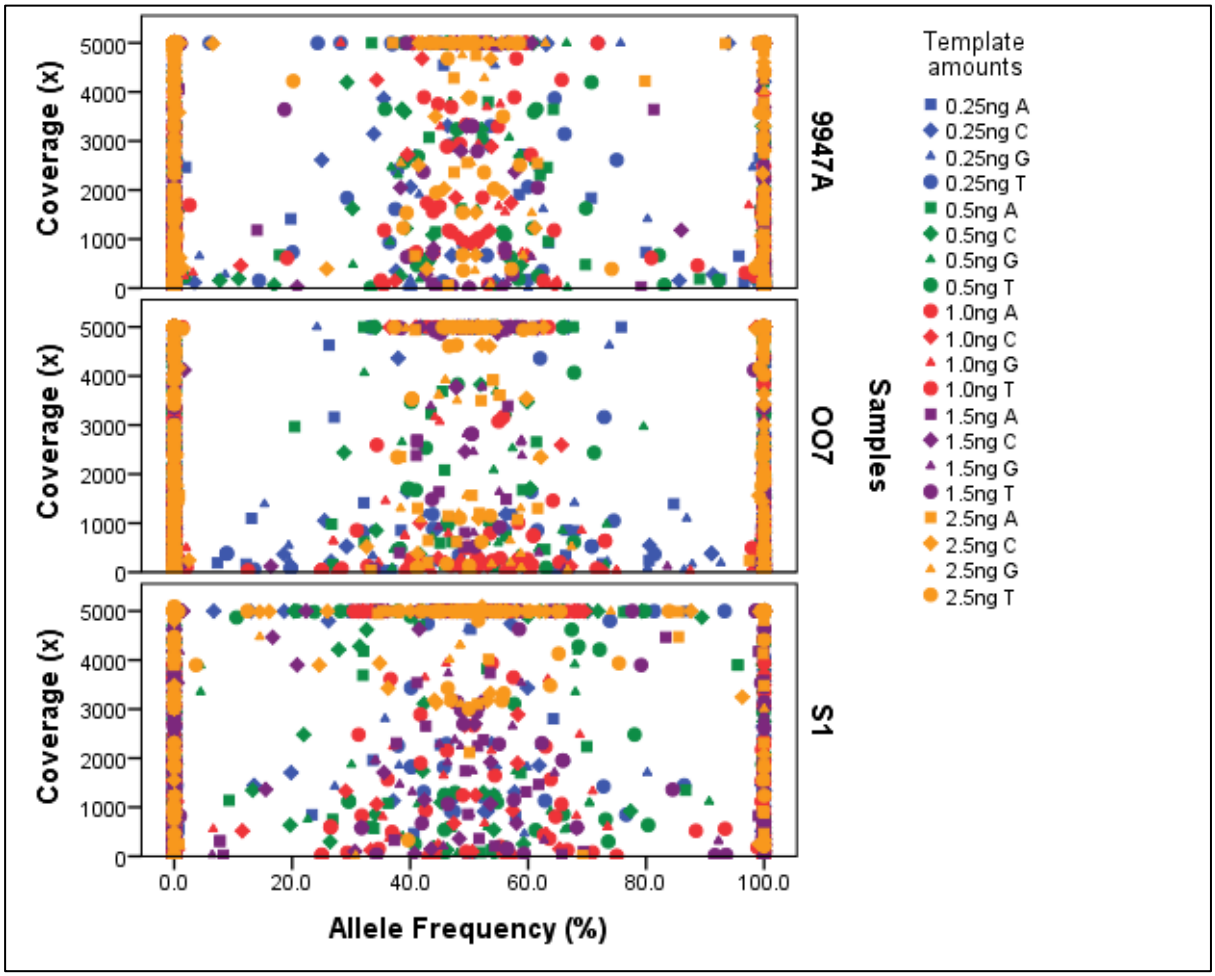


Figure S2

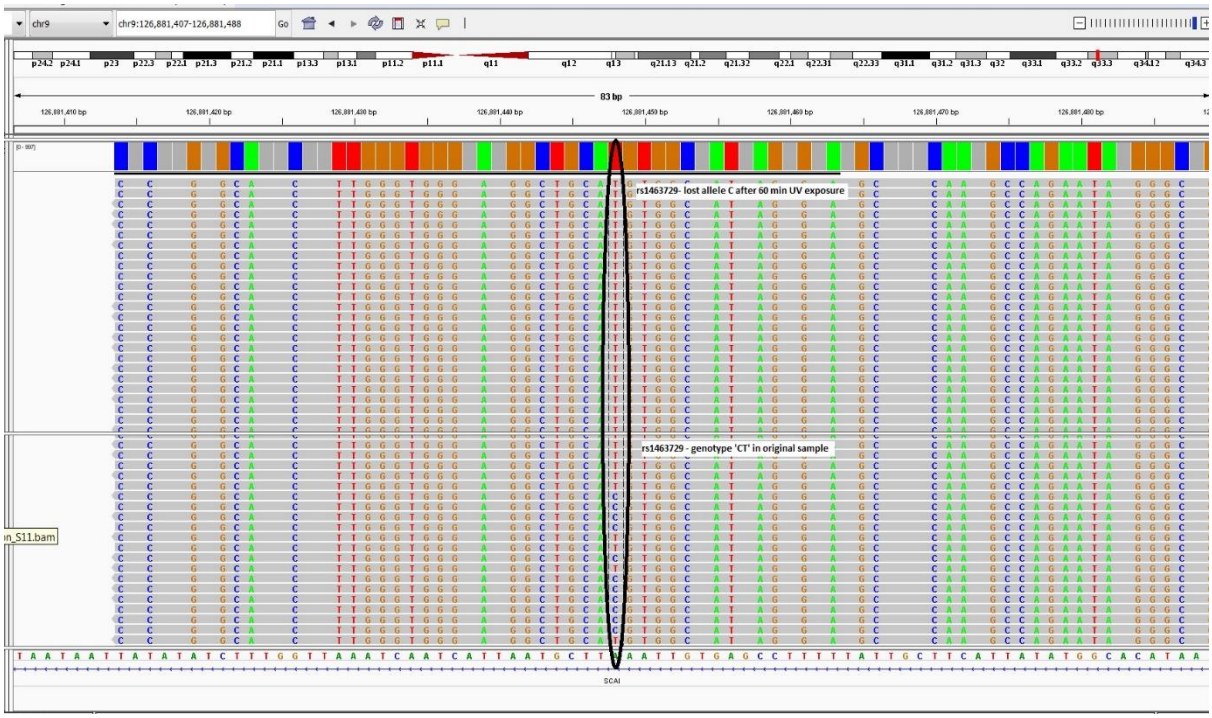


Figure S3

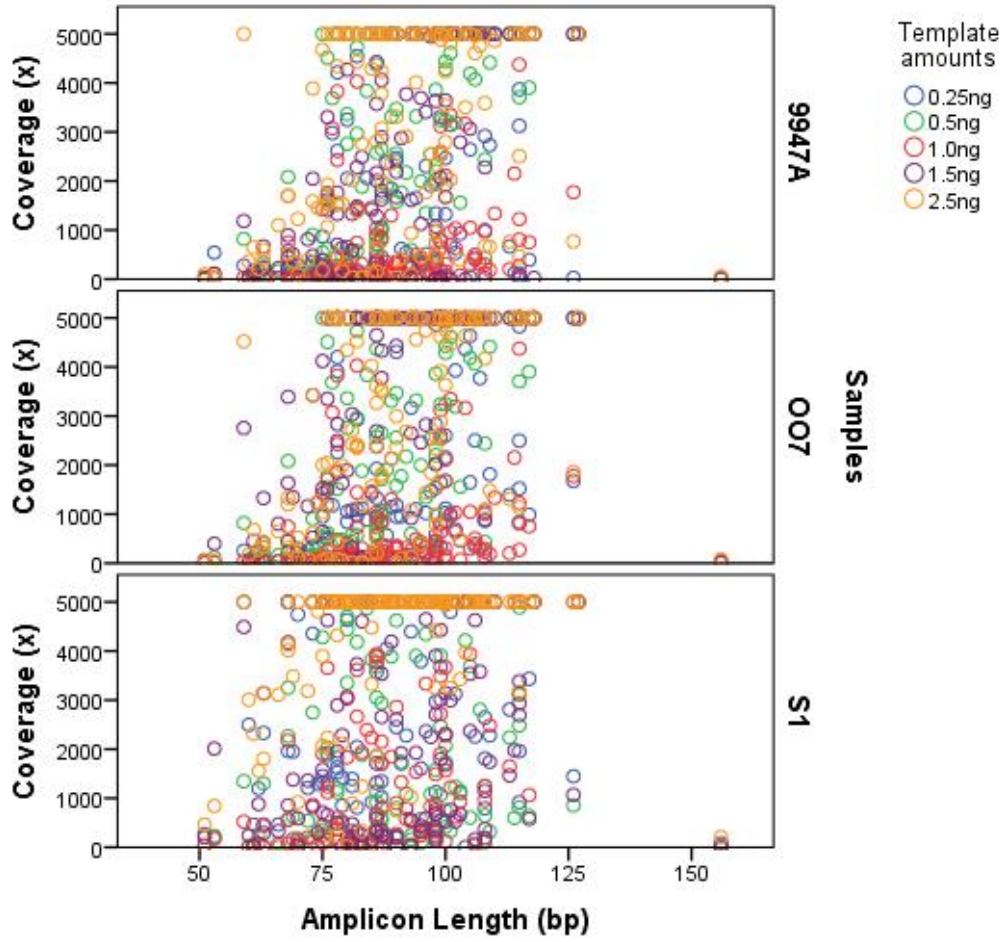


Figure S4

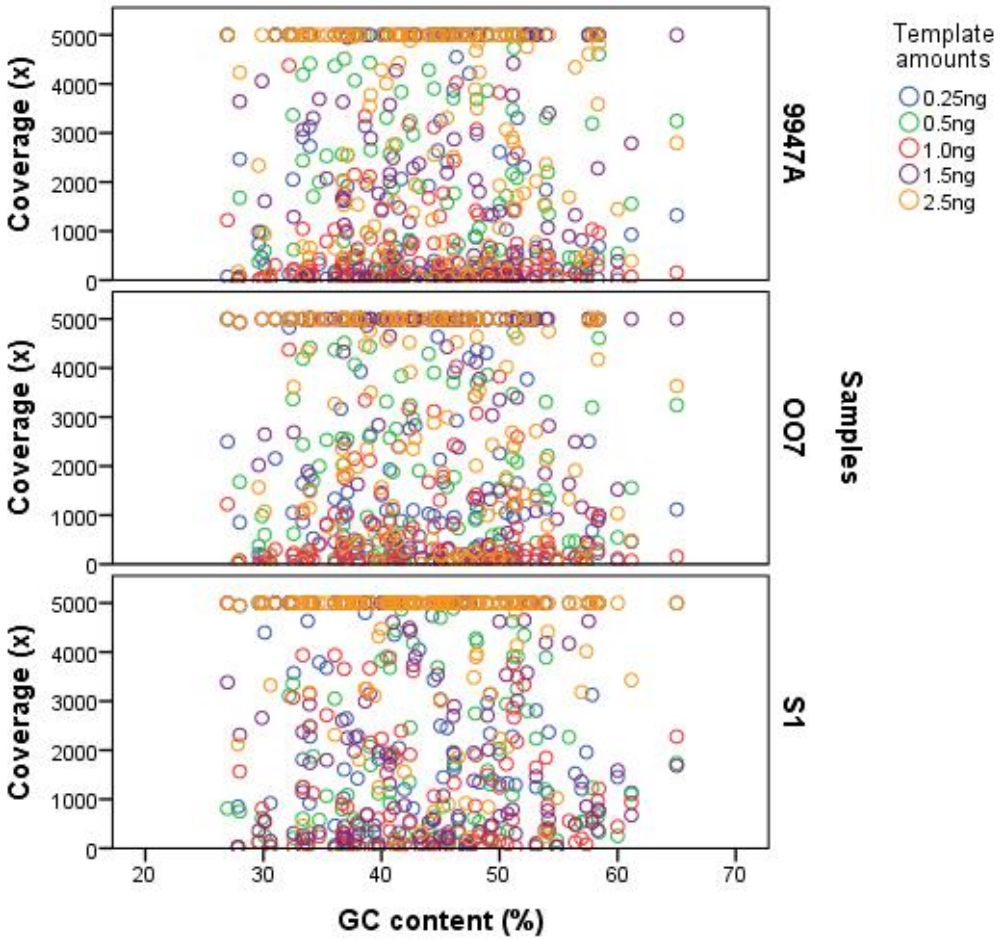
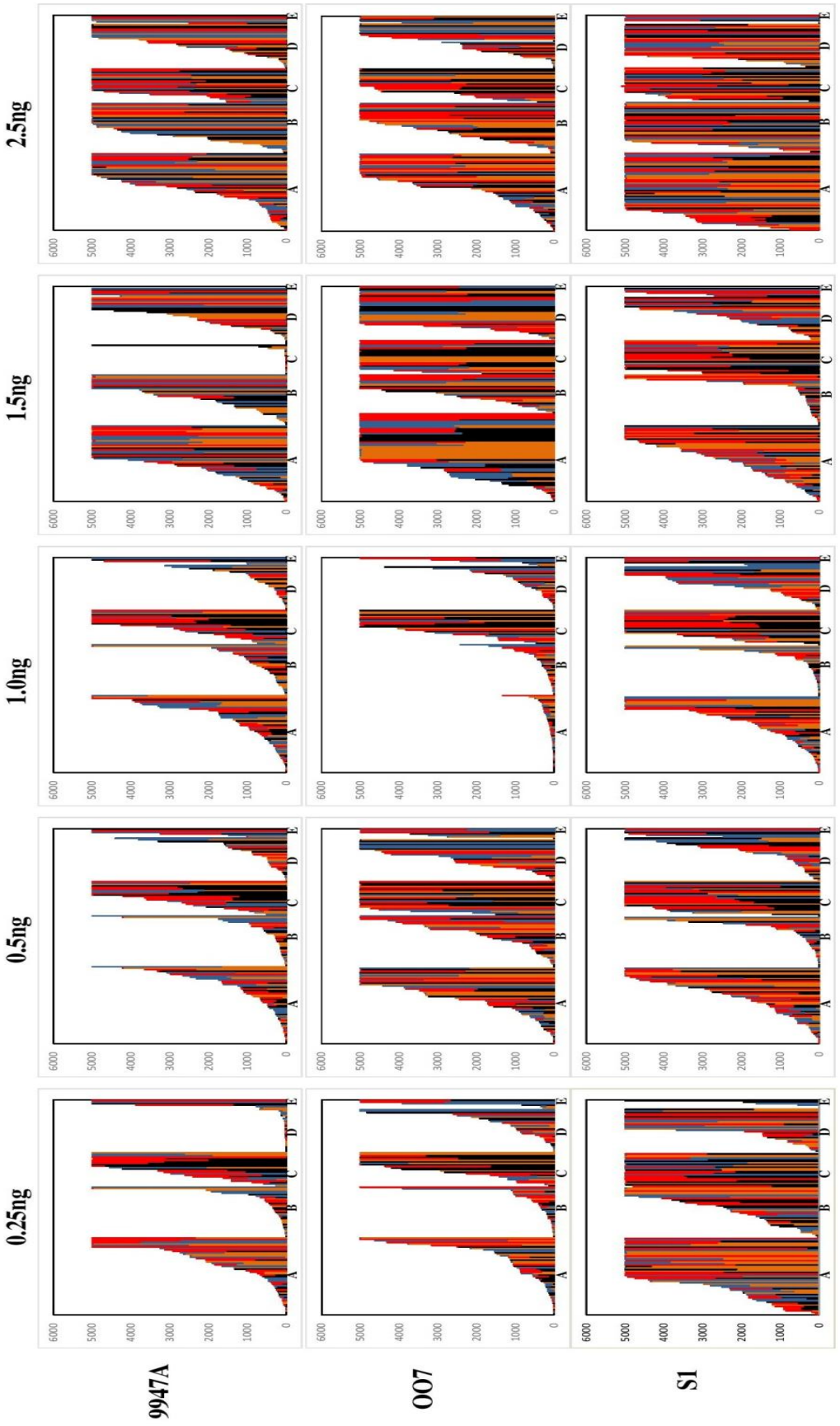

Coverage (x)
5000
4000
3000
2000
1000
0
9947A
OO7
S1
Samples
Template amounts
0.25ng
0.5ng
1.0ng
1.5ng
2.5ng
20
30
40
50
60
70
GC content (%)

Figure S5


0.25ng
0.5ng
1.0ng
1.5ng
2.5ng
9947A
007
S1
A
B
C
D
E
6000
5000
4000
3000
2000
1000
0

4.2 Mehta B, Venables S, Roffey P (2018) Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. International Journal of Legal Medicine 132 (1):125–132.

FORM E: DECLARATION OF CO-AUTHORED PUBLICATION CHAPTER

For use in theses which include publications. This declaration must be completed for each co-authored publication and to be placed at the start of the thesis chapter in which the publication appears.

Declaration for Thesis Chapter 4.2

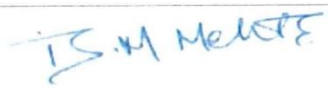
Declaration by candidate

In the case of Chapter 4.2, the nature and extent of my contribution to the work was the following:

Nature of contribution	Extent of contribution (%)
Proposal of the research, experimental work, data collection and analysis and writing the first draft of the manuscript	75

The following co-authors contributed to the work.

Name	Nature of contribution	Extent of contribution (%)	Contributor is also a student at UC Y/N
Samantha Venables	Revising and editing of the manuscript, critical feedback on the data analysis	10	N
Paul Roffey	Revising and editing of the manuscript, critical feedback on the research design and data analysis	15	N

Candidate's Signature		Date 04/04/2018

Declaration by co-authors

The undersigned hereby certify that:

1. The above declaration correctly reflects the nature and extent of the candidate's contribution to this work, and the nature of the contribution of each of the co-authors.
2. They meet the criteria for authorship in that they have participated in the conception, execution, or interpretation, of at least that part of the publication in their field of expertise;
3. They take public responsibility for their part of the publication, except for the responsible author

who accepts overall responsibility for the publication;

4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	**Specialist Operations- Forensics, Australian Federal Police, Canberra**

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	**Name of the co-authors**	**Signature**
4/4/18	Paul Roffey	P.Roffey

[* Please insert additional rows as needed.]

who accepts overall responsibility for the publication;

4. There are no other authors of the publication according to these criteria;
5. Potential conflicts of interest have been disclosed to (a) granting bodies, (b) the editor or publisher of journals or other publications, and (c) the head of the responsible academic unit; and
6. The original data are stored at the following location(s) and will be held for at least five years from the date indicated below:

Location(s)	Specialist Operations- Forensics, Australian Federal Police, Canberra

[Please note that the location(s) must be institutional in nature, and should be indicated here as a department, centre or institute, with specific campus identification where relevant.]

Date	Name of the co-authors	Signature
4/4/18	Paul Roffey	
6/4/2018	SAMANTHA VENABLES	

[* Please insert additional rows as needed.]

[Material from: 'Mehta, B., Venables, S. & Roffey, P. Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. Int J Legal Med 132, 125–132 (2018). https://doi.org/10.1007/s00414-017-1591-9, ©Springer-Verlag Berlin Heidelberg 2017]

Int J Legal Med
DOI 10.1007/s00414-017-1591-9

METHODS PAPER

Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping

Bhavik Mehta [1,2] · **Samantha Venables** [1,2] · **Paul Roffey** [1,2]

Received: 6 December 2016 / Accepted: 6 April 2017

Abstract Massively parallel sequencing (MPS) is fast approaching operational use in forensic science, with the capability to analyse hundreds of DNA identity and DNA intelligence markers in multiple samples simultaneously. The ForenSeq™ DNA Signature Kit on MiSeq FGx™ (Illumina) workflow can provide profiles for autosomal short tandem repeats (STRs), X chromosome and Y chromosome STRs, identity single nucleotide polymorphisms (SNPs), biogeographical ancestry SNPs and phenotype (eye and hair colour) SNPs from a sample. The library preparation procedure involves a series of steps including target amplification, library purification and library normalisation. This study highlights the comparison between the manufacturer recommended magnetic bead normalisation and quantitative polymerase chain reaction (qPCR) methods. Furthermore, two qPCR chemistries, KAPA® (KAPA Biosystems) and NEBNext® (New England Bio Inc.), have also been compared. The qPCR outperformed the bead normalisation method, while the NEBNext® kit obtained higher genotype concordance than KAPA®. The study also established an MPS workflow that can be utilised in any operational forensic laboratory.

Electronic supplementary material The online version of this article (doi:10.1007/s00414-017-1591-9) contains supplementary material, which is available to authorized users.

✉ Bhavik Mehta
bhavik.mehta@hotmail.com

[1] National Centre for Forensic Studies, Faculty of Education, Science, Technology and Mathematics (ESTeM), University of Canberra, Canberra, Bruce ACT 2617, Australia

[2] Specialist Operations – Forensics, Australian Federal Police, Tambreet Street, Majura ACT 2609, Australia

Published online: 18 April 2017

Keywords Forensic DNA profiling · Next generation sequencing (NGS) · Massively parallel sequencing (MPS) · Illumina MiSeq FGx · Library normalisation · Quantitative polymerase chain reaction (qPCR)

Introduction

Forensic DNA identification using short tandem repeats (STRs) has been continuously performed for over two decades using capillary electrophoresis (CE) based detection systems. These profiling techniques changed the course of judicial systems in the modern era, making DNA one of the important evidence types for solving crimes. However, cases are known where these STR profiles do not provide information useful to the investigation, either because no match can be found in the reference database; a partial profile is generated or required to narrow down the pool of suspects [1]. In these circumstances, intelligence information gathered from DNA, commonly referred as forensic DNA phenotyping (FDP) or molecular photofitting, can provide useful leads [2]. Single nucleotide polymorphisms (SNPs) are commonly used for FDP in the form of predictions for biogeographical ancestry (BGA) and externally visible characteristics (EVCs) [3, 4]. The ability of massively parallel sequencing (MPS), also known as next generation sequencing (NGS), to simultaneously generate DNA identity and intelligence profiles at the same time for a sample has been of significant interest to forensic scientists [5–7].

At present, two MPS bench top platforms are suited to forensic application namely, the MiSeq FGx™ (Illumina) and the Ion™ PGM or Ion™ S 5 (Thermo Fisher Scientific). Both of these technologies are capable of genotyping hundreds of markers in multiple samples using its barcode labelling strategy [5, 8, 9]. Commercially released forensic panels are available for these MPS platforms, including the

ForenSeq™ DNA Signature (Illumina) and Precision ID Identity, Precision ID Ancestry, Precision ID mtDNA Whole Genome, Precision ID mtDNA Control Region, Precision ID GlobaFiler™ NGS STR panels (Thermo Fisher Scientific). Some of these commercial panels have already undergone forensic evaluation studies [6, 8, 10, 11]. It is also possible to take a custom approach and genotype existing forensic SNaPshot® ancestry and phenotype panels in any combination with these MPS technologies [5, 9].

The MiSeq FGx™ (Illumina) system along with the ForenSeq™ DNA Signature panel (Illumina) is commercial MPS product available for operational forensic purposes. This multiplex kit consists of more than 200 markers including 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 identity SNPs, 56 ancestry SNPs, 24 phenotypic SNPs and Amelogenin. The ForenSeq™ workflow involves template and library preparation, sequencing and the analysis of sequence data using Illumina's Universal Analysis Software (UAS) [6, 11]. The sequencing template preparation step in the Illumina recommended protocol includes the normalisation of libraries using magnetic beads before loading the sequencing cartridge. The magnetic bead based normalisation method uses a fixed volume of beads which captures equal amounts of DNA from each sample library and enables the creation of an equimolar library pool [12]. However, research has suggested various other additional library normalisation techniques that can be used including NanoDrop® (Thermo Fisher Scientific) [12], Qubit® (Thermo Fisher Scientific) [13] and qPCR methods [14].

This preliminary technical study compares the manufacturer recommended magnetic bead normalisation and qPCR based normalisation methods to assess which approach produces the highest level of consistency and reproducibility required for utilisation in forensic casework. This article also recommends an MPS workflow that any forensic laboratory can utilise for processing the samples using the ForenSeq™ DNA Signature panel on the MiSeq FGx™ (Illumina).

Materials and methods

Sample preparation

The Standard Reference Material (SRM) 2800 M (Promega) DNA stock (10 ng/μL) was diluted to 0.2 ng/μL with amplification grade water. Each sequencing run consisted of 32 samples made up of 31 replicates of 2800 M at 1 ng DNA input amount and one PCR negative control (amplification grade water).

PCR amplicons and library preparation

PCR amplicons were generated using the ForenSeq™ DNA Signature kit (231 markers) for all samples. The PCR amplicons were subjected to an enrichment step (second PCR step in the protocol) to barcode the sample with unique oligo-nucleotides following the Illumina protocol. The generated libraries were purified using sample purification beads supplied within the ForenSeq™ kit following the manufacturer's protocol. The enrichment and purification pipetting steps were carried out on the EpMotion® 5075 (Eppendorf) liquid handling platform. EpMotion scripts were developed in-house with the help of Eppendorf personnel. The thermal cycling steps were performed on 9700 Thermal Cycler (Thermo Fisher Scientific).

Magnetic bead-based library normalisation

The bead normalisation was carried out using the library normalising magnetic beads (LNB1) provided in the ForenSeq™ DNA Signature kit (Illumina) following the manufacturer's protocol. Each normalised library was pooled in equal volume (5 μL) in a 1.5 mL DNA low bind tube (Eppendorf). The bead normalisation step was performed on the EpMotion® 5075 (Eppendorf) liquid handling platform. A volume of 11 μL of the library pool was used for loading on the FGx™ sequencing cartridge. The cartridge was run in the MiSeq FGx™ sequencer using the manufacturer's recommended conditions.

qPCR library normalisation

Two qPCR library normalisation kits were tested, these being the KAPA® library quantification kit for Illumina® platforms (KK4873, KAPA Biosystems) and the NEBNext® library quant kit for Illumina® (E7630, New England BioLabs®). Samples of the purified ForenSeq™ libraries were diluted 1: 10,000 with amplification grade water. Diluted libraries were subjected to qPCR on a 7500 real-time PCR instrument (Thermo Fisher Scientific) following the conditions outlined in Table 1. KAPA® qPCR data analysis was performed using their supplied Microsoft Excel macro sheet, while the Nebiocalculator web tool (http://nebiocalculator.neb.com/#!/) was used for the analysis of NEBNext® qPCR data. The libraries were then diluted to 10 nM concentration with amplification grade water based on the qPCR quantitation values. Each qPCR normalised library was pooled in equal volume (5 μL) in a 1.5 mL DNA low bind tube (Eppendorf). For the KAPA® qPCR library pool, MiSeq sequencing was tested using three different concentrations, a 7 μL volume of the library pool, a 7 μL volume of a 1:2 dilution of the library pool and then a 7 μL volume of a 1:5 dilution of the library pool. Based on the sequencing results obtained from the KAPA® pool, the NEBNext® pool was only run using 7 μL volume of a 1:2 dilution of the library pool.

Table 1 Master Mix and PCR conditions for KAPA® and NEBNext® qPCR methods

Kit	PCR set up		PCR Conditions			
	Component	Vol. (μL)	Step	Temp (°C)	Time	Cycles
KAPA® Library Quantification kit Illumina® platforms	2X KAPA SYBR® FAST qPCR Master Mix + 10X Primer Premix	6	Initial denaturation	95	5 min	1
	DNA (Standards and samples)	4	Denaturation	95	30 s	35
	Total	10	Extension/data acquisition	60	45 s	
NEBNext® Library Quant kit for Illumina®	NEBNext Library Quant Master Mix (with primers)	8	Initial denaturation	95	1 min	1
	DNA (Standards and samples)	2	Denaturation	95	15 s	35
	Total	10	Extension/data acquisition	63	45 s	

Data analysis

ForenSeq™ Universal Analysis software (UAS) (Illumina) was used to analyse and genotype sequencing data using the default analysis settings. The quality metric parameters including cluster density, cluster passing filter, phasing and pre-phasing were used to assess the run data quality using the default UAS settings prior to genotyping. The default baseline and stochastic thresholds are locus specific and set as percentage of total number of reads per locus (Supplementary File 1). The minimum number of reads to make an allele call is fixed at >30 reads in the UAS software. Genotype concordance was assessed by comparing the 2800 M profile (provided by Illumina) to the genotypes obtained for each of the 2800 M replicates from each of the normalisation method runs. Statistical tests were performed using IBM SPSS (v.21).

Results

The workflow for ForenSeq™ DNA Signature kit library preparation using QIAgility™ (Qiagen) and EpMotion™ 5075 (Eppendorf) liquid handling robots is shown in Fig. 1.

Quality metrics

The ForenSeq™ UAS generates quality metrics consisting of four important sequencing parameters [15]:

- Cluster density (K/mm^2) which is a measure of the number of clusters per square millimetre for the run
- Cluster passing filter (%) which is a measure of the total number of clusters that generate one fluorescent signal and passes the default manufacturer internal filter
- Phasing rate (%) which is a measure of number of molecules within a cluster that fall behind the current cycle
- Pre-phasing rate (%) which is a measure of molecules within a cluster that run ahead of the current cycle

The manufacturer set default range for each of the quality metrics was used to determine whether to accept the run data and proceed with genotyping [15]. Table 2 shows the quality metric parameters for all five MiSeq FGx™ runs performed in this study. The sequencing run with 7 μL of the KAPA® qPCR library pool failed to meet the required cluster passing filter range as a result of the run being over-clustered. Subsequently, a 7 μL volume of a 1:2 dilution of the KAPA® library pool was sequenced and this run passed all of the quality metrics. The sequencing run with 7 μL of a 1:5 dilution of the KAPA® qPCR library pool obtained the lowest cluster density out of all the KAPA® runs. Based on these results, the NEBNext® sequencing run was performed using a 7 μL volume of a 1:2 dilution of the NEBNext® qPCR library pool and compared with the corresponding KAPA® qPCR run for genotype concordance.

Overall, both KAPA® (1:2) and NEBNext® (1:2) runs obtained higher coverage than the bead normalisation procedure (Supplementary Table 1). One-way Kolmogorov-Smirnov test showed that the coverage distribution followed a normal pattern across all replicates within each runs (except KAPA® qPCR run possibly due to over clustering) (Supplementary Table 1). One-way ANOVA test showed that there is significant difference ($p = 0.001$) in coverage distribution between all runs. *t*-test performed between KAPA® (1:2) and NEBNext® (1:2) showed no variability in coverage distribution for two methods across all replicates ($p = 0.298$). The overall coverage for NEBNext® run was better than the other runs, excluding the KAPA® qPCR 7 μL and KAPA® 1:5 dilution runs which, did not pass all the quality metrics parameters (Supplementary Table 1).

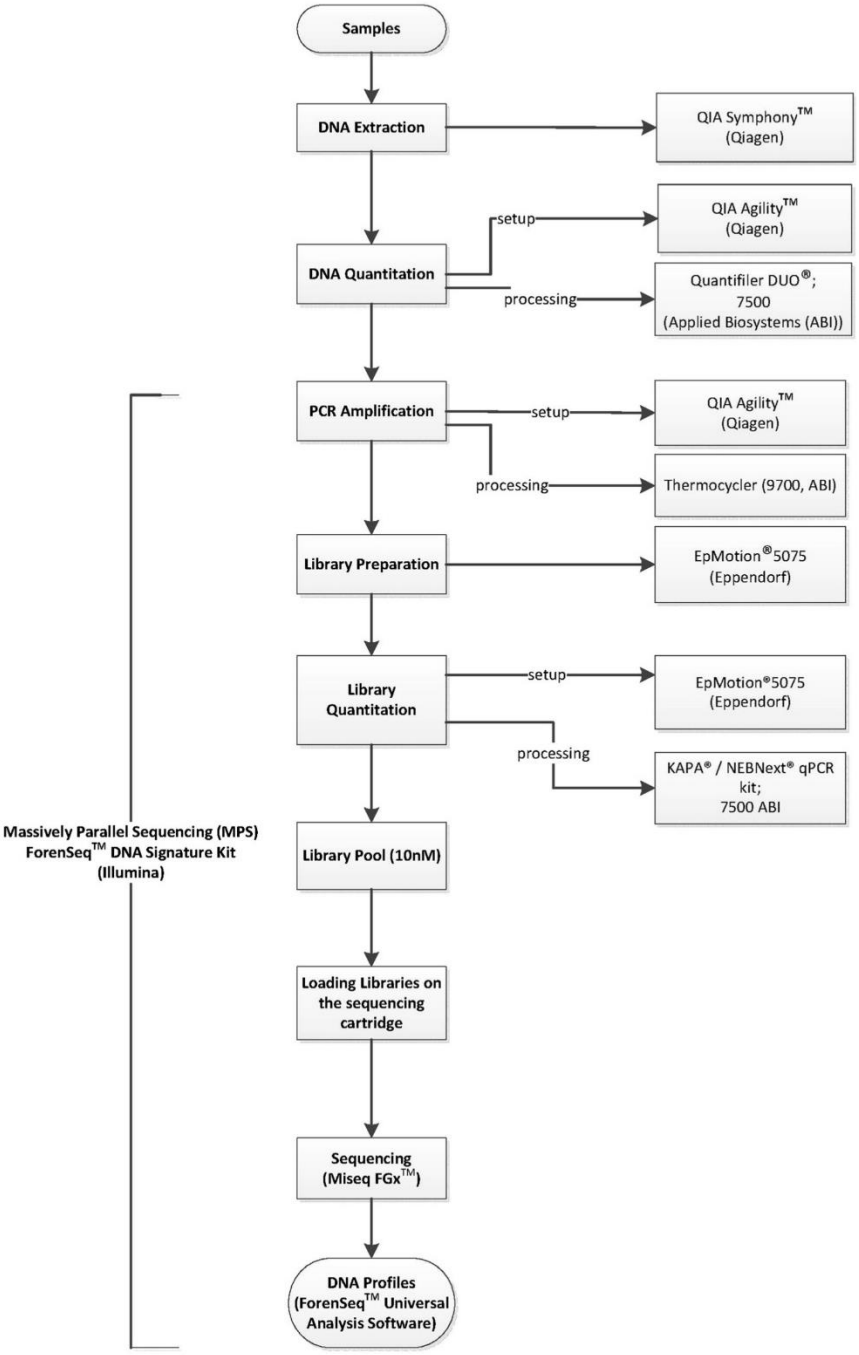


Fig. 1 MPS workflow used for this study

Genotype concordance

The genotype concordance was determined by comparing the Illumina provided 2800 M profile and the profiles obtained using all three normalisation procedures across all ForenSeq™ loci for 31 replicates of 2800 M. Table 3 shows the average genotype concordance of the 31 replicates of 2800 M according to

Table 2 Quality metrics parameters for all the runs

Parameters	Bead[a]	KAPA qPCR[b]	KAPA qPCR (1:2)[c]	KAPA qPCR (1:5)[d]	NEBNext qPCR (1:2)[c]
Cluster density (default range: 400-1600 K/mm^2)	806	1237	1057	380	1073
CPF (default range: 80–100%)	92.96	78.01	90.82	95.03	90.35
Phasing	0.216	0.199	0.222	0.221	0.188
Pre-phasing	0.093	0.102	0.116	0.083	0.113

[a] This is the manufacturer's recommended bead normalisation procedure. Eleven microliters of the library pool was sequenced

[b] Seven microliters of 10 nM library pool was sequenced

[c] The 10 nM library pool was diluted 1:2 and 7 μL of this dilution was sequenced

[d] The 10 nM library pool was diluted 1:5 and 7 μL of this dilution was sequenced

marker type for each of the normalisation runs performed. Sample 2800 M_18 did not produce any result in the bead normalisation run and was excluded from the analysis. The KAPA® qPCR and KAPA® 1:5 dilution runs did not meet all the quality metrics analysis criteria, hence, excluded from further analysis. Both the KAPA® (1:2) and NEBNext® (1:2) qPCR kits offered a higher level of genotype concordance than the Illumina recommended bead normalisation procedure. Out of the four runs analysed, the NEBNext® qPCR kit showed the highest genotype concordance across all marker types included in the ForenSeq™ DNA Signature kit. The overall genotype concordance for the 231 ForenSeq™ markers for the NEBNext® run was 93.8%, with greater than 95% concordance obtained for all marker types except the identity informative SNPs (iiSNPs).

The genotype non-concordance observed between tested and manufacturer supplied profiles were due to the following three reasons:

- Stutter—an allele in the stutter position was above the stutter threshold and was called by the software using the default setting- giving a tri-allelic call at that locus
- Allele drop out (ADO)—an allele in a known heterozygous genotype was not detected above the threshold using the default settings—resulting in an incorrect homozygous genotype being called.
- Locus drop out (LDO)—no data detected above the threshold using the default setting —an "NN" genotype

Table 4 describes the average non concordance for different normalisation runs across all replicates. NEBNext® (1:2) was the best performer among all the runs with average null stutter,

Table 3 Average genotype concordance between standard profile supplied in ForenSeq™ DNA Signature protocol (Illumina) and profiles obtained using different library normalisation procedures for all replicates of 2800 M

Marker type (*n*)	Concordance (%)[a]			
	Bead[b]	KAPA qPCR[c]	KAPA qPCR[d]	NEBNext qPCR[d]
ForenSeq Set (231)	89.4	91.1	92.6	93.8
Amelogenin (1)	54.8	90.3	90.3	96.8
Autosomal STRs (27)	90.7	95.8	96.2	98.0
Y-STRs (24)	96.2	100.0	100.0	100.0
X-STRs (7)	96.3	100.0	100.0	100.0
iiSNPs (94)	79.9	84.5	86.8	87.9
aiSNPs (54)	87.0	91.4	94.0	99.2
piSNPs (24)	95.0	98.9	98.9	99.6

[a] Concordance is a measure of the percentage of loci that are above the reporting threshold and demonstrate the expected genotype

[b] This is the manufacturer's recommended bead normalisation procedure. Eleven microliters of the library pool was sequenced. The 2800_18 replicate produced no sequence data, so the total number of replicates for this run is 30

[c] Seven microliters of 10 nM library pool was sequenced

[d] The 10 nM library pool was diluted 1:2 and 7 μL of this dilution was sequenced

Table 4 Average genotype non-concordance between standard profile supplied by the manufacturer and profiles obtained using different library normalisation procedures across all replicates of 2800 M

Method	Concordance (%)	Non-concordance (%)		
		Stutter	ADO	LDO
Bead[a] ($n = 30$)	89.4	0.5	2.6	7.6
KAPA qPCR[b] ($n = 31$)	91.1	0.1	2.1	6.7
KAPA qPCR[c] ($n = 31$)	92.6	0.2	1.7	5.5
NEBNext qPCR[c] ($n = 31$)	93.8	0	1.6	4.6

ADO allele drop out, *LDO* locus drop out

[a] The manufacturer's recommended protocol (bead) had 11 μL of the library pool sequenced. One of the 31 2800 M replicates produced no sequence data, so the number of replicates for this run is 30

[b] Seven microliters of 10 nM library pool was sequenced

[c] The 10 nM library pool was diluted 1:2 and 7 μL of this dilution was sequenced

~1.6% ADO and ~4.6% LDO non-concordance respectively, across all replicates.

Poor performing markers

The overall genotype concordance for each of the library normalisation procedures was adversely affected by a small number of markers with low genotype concordance. Genetic markers with a concordance of 60% or less from the bead normalised library pool were classified as having low genotype concordance (Supplementary Table 2). The improvement in genotype concordance of these markers when the KAPA® and NEBNext® qPCR based library normalisation methods were used is also indicated in Supplementary Table 2. One ancestry-informative SNP (aiSNP; rs1572018) and four identity-informative SNPs (iiSNP; rs1031825, rs1294331, rs1736442, rs1528460) had less than 5% genotype concordance regardless of the library normalisation method used.

Discussion

Massively parallel sequencing (MPS), an upcoming capability in forensic science, can not only perform DNA identification but also provide DNA intelligence information. Compared to current STR profiling, MPS workflow enables the analysis of multiple samples simultaneously; however, it is complicated by the additional steps required for the formation of the libraries. The automation of the library preparation procedure can help to reduce the risk of background contamination and also provide options for sample tracking. Figure 1 highlights the workflow that can be utilised by any operational forensic laboratory in processing samples for MPS.

The library preparation process can be broadly classified as target amplification (amplification of targeted DNA markers), generation of purified libraries (attachment of the unique DNA barcodes and magnetic bead based purification of libraries), library normalisation (quantitation of libraries for equimolar pooling) and library pooling (equimolar pooling). The pooled library is loaded on to the sequencing cartridge for massively parallel targeted sequencing [6, 10, 16]. Within the library preparation, library normalisation is a vital step which ensures that each of the sample libraries are present in almost equal quantities in the pool with the aim of providing equal representation of total coverage per sample.

MiSeq FGx™ along with ForenSeq™ DNA Signature Kit and UAS software is offered as a validated tool by the Illumina to the forensic operational laboratories, and the quality metrics is recommended as one of the evaluating metrics to confirm the sequencing run has been completed without errors caused by system artefacts. The default range of the quality metrics parameters is not amendable in the UAS software. The magnetic bead normalisation procedure recommended in the ForenSeq™ DNA Signature workflow does not provide a quantitative measure of the libraries and hence the qPCR option was explored. The results indicate that the qPCR normalisation method improved the total coverage per sample (Supplementary Table 1) and produced higher genotype concordance in comparison with the bead normalisation method (Table 3). The improved total coverage per sample may translate to a higher number of reads per allele, which in turn provides greater confidence in genotype calling. A higher pooled library amount of 11 μL was used here in the bead normalised run compared to the manufacturer recommended 7 μL, and still the overall coverage obtained was less than the qPCR normalisation procedures (Supplementary Table 1). The KAPA qPCR run at 7 μL was over-clustered reducing the number of clusters with pure signals and hence decreasing the cluster passing filter percentage. However, the quality metrics indicated that the optimal results were obtained by sequencing a 7 μL volume of a 1:2 dilution of the qPCR normalised library pool. The NEBNext® qPCR chemistry slightly outperformed the KAPA® chemistry; however, both the qPCR chemistries tested outperformed the Illumina recommended bead normalisation procedure. The overall coverage of NEBNext® (1:2) qPCR run was 6.1% higher than the KAPA® (1:2) run. The coefficient of variance (CV) suggests both qPCR procedures provide similar uniformity of the coverage across all the replicates. However, the overall genotype concordance of NEBNext® was higher than the KAPA® approach (Supplementary Table 1).

The overall average genotype concordance of ~89% for magnetic bead normalisation suggests not enough reproducibility and consistency required by the forensic laboratory.

This further may be disadvantageous for forensic geneticists/biologist as can potentially lead to mis-interpretation. In comparison, NEBNext® qPCR approach showed overall genotype concordance of ~94%, and excluding IISNPs, it improved to ~99% (Table 3). Locus drop out (LDO) was by far the biggest contributor to our non-concordance loci for each of the methods tested (Table 4). The genotype concordance of five poor performing markers was less than 5% in all runs across the 31 replicates of 2800 M (Supplementary Table 2). The genotyping accuracy of these poor performing markers appeared independent to the library normalisation procedure, suggesting that there is a problem with the amplification of these markers using the ForenSeq™ DNA Signature Kit.

Library normalisation using qPCR may also help to provide the consistency and reproducibility from run to run that is expected from forensic analysis procedures. In MPS, samples are analysed simultaneously as part of a pool which means each sample is affected by the other samples in that pool. As such, a sample with high library yield can consume more sequencing real estate than a sample with low library yield. For example, sample A with an input amount of 0.5 ng template and sample B with an input amount of 1 ng template may generate different quantities of libraries. Library normalisation using qPCR allows the addition of equal quantities of sample A and B libraries into the sequencing pool and hence may account for the differences in input template quantity.

The time difference between bead normalisation and qPCR method was approximately 2 h. The magnetic bead normalisation method was more time efficient as it does not involve a series of library dilution steps and eliminates the need for quantifying individual libraries [17]. Illumina provides magnetic bead normalisation reagents along with the ForenSeq™ DNA Signature Kit components; however, qPCR reagents would be required to purchase separately. The cost per reaction of KAPA® and NEBNext® was AU$2.15 (half volume reaction AU$ 1.08) and AU$1.8 (half volume reaction AU$ 0.9; costs obtained from the Australian distributors GeneWorks and Genesearch), respectively. This is not significantly higher considering the overall cost of ForenSeq™ library preparation and sequencing per reaction, ~AU$ 140 (includes only library preparation and sequencing reagents costs; calculated based on the list price obtained from Illumina website).

This preliminary comparative study performed on SRM controls using the manufacturer recommended 1 ng DNA input amount, indicates the better performance of qPCR over the bead based normalisation procedure. Where laboratories are analysing samples with DNA input of less than 1 ng with the bead normalisation procedure, it may be necessary to further optimise the volume of beads based on DNA input. This study would be a guide for those laboratories choosing to optimise magnetic bead and/or qPCR normalisation procedures for forensic MPS typing.

Conclusions

Developing an appropriate MPS workflow will be a key factor in utilising this capability for routine forensic use. Library normalisation using qPCR provides a greater level of consistency and reproducibility in comparison to the manufacturer recommended bead normalisation method tested here. This is important if MPS is to be used for forensic application. Using this preliminary study, future normalisation studies can be conducted using reference, casework and degraded samples with varying DNA input amount. More research should be conducted to optimise the magnetic bead normalisation procedure for sample types yielding low level DNA input amounts. In addition, studies on different liquid handling robots must be also performed to identify platforms suitable for use in the MPS workflow within forensic DNA laboratories.

Acknowledgements The authors gratefully acknowledge funding support from the Specialist Operations- Forensics, Australian Federal Police. We would also like to acknowledge Dr. Eric Wenger and Slazana Ristveska from Specialist Operations—Forensics, Australian Federal Police for their consultation support.

Compliance with ethical standards

Conflict of interest The authors declared that they have no conflict of interest.

References

1. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. Nat Rev Genet 12(3):179–192
2. Butler K, Peck M, Hart J, Schanfield M, Podini D (2011) Molecular "eyewitness": forensic prediction of phenotype and ancestry. Forensic Sci Int Genet Suppl Ser 3:e498–e499
3. Budowle B, van Daal A (2008) Forensically relevant SNP classes. BioTechniques 44(5):603–610
4. Kayser M (2015) Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Sci Int Genet 18:33–48
5. Daniel R, Santos C, Phillips C, Fondevila M, van Oorschot RAH, Carracedo Á, Lareu MV, McNevin D (2015) A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Sci Int Genet 14(0):50–60
6. Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina® Beta version ForenSeq™ DNA Signature prep kit for use in genetic profiling. Forensic Sc Int Genet 20:20–29
7. Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. Forensic Sci Int Genet 18:78–89
8. Silvia AL, Shugarts N, Smith J (2016) A preliminary assessment of the ForenSeq™ FGx System: next generation sequencing of an STR and SNP multiplex. Int J Leg Med 131(1):1–14

9. Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, McNevin D (2016) Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis 37(21): 2832–2840
10. Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N (2014) Evaluation of the ion Torrent™ HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing. Forensic Sci Int Genet 12(0):144–154
11. Jäger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, Way L, Walichiewicz P, Silva D, Pham N, Caves G (2017) Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. Forensic Sci Int Genet 28:52–70
12. Harris JK, Sahl JW, Castoe TA, Wagner BD, Pollock DD, Spear JR (2010) Comparison of normalization methods for construction of large, multiplex amplicon pools for next-generation sequencing. Appl Environ Microbiol 76(12):3863–3868
13. Trujillano D, Weiss ME, Köster J, Papachristos EB, Werber M, Kandaswamy KK, Marais A, Eichler S, Creed J, Baysal E (2015) Validation of a semiconductor next-generation sequencing assay for the clinical genetic screening of CFTR. Molecular genetics & genomic medicine 3(5):396–403
14. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13(1):341
15. Illumina (2016) ForenSeqTM Universal Analysis Software guide (v15053876 01)
16. Fordyce SL, Mogensen HS, Børsting C, Lagacé RE, Chang C-W, Rajagopalan N, Morling N (2015) Second-generation sequencing of forensic STRs using the Ion Torrent™ HID STR 10-plex and the Ion PGM™. Forensic Sci Int Genet 14(0):132–140
17. Hayden MJ, Wa SLKK (2013) Method of Producing a Normalised Nucleic Acid Library Using Solid State Capture Material. Google Patents, U.S. Patent Application No. 14/430, 786

Chapter 5: Conclusions and future directions

5.1. Conclusions

DNA evidence has been used in forensic science since 1980s [1], following the development of the restriction fragment length polymorphisms method, which was soon replaced by variable number tandem repeats (VNTRs) used in DNA fingerprinting technique by Sir Alec Jeffrey[2] [3]. DNA was first applied in a landmark double-murder case in Leicester in 1986, which led to the identification of the perpetrator, Collin Pitchfork [4]. Although a successful forensic methodology, DNA fingerprinting was not sufficiently sensitive in cases where small amounts of DNA were obtained and the method was phased out with the introduction of short tandem repeat (STR)-based DNA-typing technology that using polymerase chain reaction (PCR) [5].

STR profiling has prevailed since the 1990s and played a vital role in shaping the modern forensics era, in which DNA is one of the most significant and valuable evidence types. An STR-based human identification assay was first reported in 1991—a three locus fluorescent multiplex PCR assay [6]. The current STR-typing technology involves the use of two to six nucleotide repeat units.—STR loci with tetra-nucleotide repeats being the most common. DNA input requirements are generally between 0.5ng to 1ng for STR-typing [7].

Several multiplex STR assays are now commercially available, including the GlobalFiler PCR Amplification kit (Thermo Fisher ScientificTFS), which enables typing of 24 STRs in a single multiplex [8]. With the availability of multiplex STR assays and advancements in STR profiling technologies, the sensitivity and reliability of STR profiles have significantly increased.

In forensic identification analysis, STR profiles generated from evidentiary DNA samples are required to match with a reference profile or criminal database profile. This requirement can become a limitation in cases where a match is not obtained or a partial STR profile is generated, which does not assist the investigation [9, 10]. In these circumstances, investigators must rely on other evidence, such as eyewitness statements. However, eyewitness statements are notoriously unreliable [10]. For example, in a 1984 case, Kirk Bloodsworth was mistakenly convicted based on unreliable eyewitness testimony and sentenced to death for the murder of a nine year old girl [11]. Forensic DNA phenotyping (FDP) can overcome the limitations of eyewitness testimony and assist investigators by providing leads to further progress investigation. FDP can act as a ‘molecular eyewitness’ and provide phenotyping inferences in the form of externally visible characteristics (EVCs, such as eye, hair colour) and biogeographical ancestry (BGA) of the donor of biological evidence [9]. FDP generally

employs single nucleotide polymorphisms (SNPs) as genetic markers. There are four broad categories of SNPs used in forensics: i) identity informative SNPs (IISNPs) can be used to complement STRs in identification; ii) lineage informative SNPs (LINSPs) are used for paternity and kinship analysis testing as well as BGA inferences; iii) ancestry informative SNPs (AISNPs) are used for BGA inferences; and iv) phenotype informative SNPs (PISNPs) are used for EVCs inferences [12]. STR profiling is mainly performed by separating PCR products on capillary electrophoresis (CE) detection systems, whereas SNP-typing can be performed by a variety of technologies [5, 13, 14]. This thesis highlighted the different genotyping tools that can be utilised for FDP SNPs, depending on the throughput requirements and application types for different laboratories. This thesis examined three main techniques: high resolution melt (HRM) analysis (low-throughput method) in Chapter 3; single base extension (SBE), typified by the SNaPshot assay (medium-throughput method) in Chapter 3; and massively parallel sequencing (MPS) (high-throughput method) in Chapter 4.

HRM analysis detects variants in real-time post-PCR [15]. The method has been utilised in clinical diagnostics for many applications, including microbial strain differentiation [16], pathogen detection [15, 17] and somatic cancer mutation detection [18]. In this thesis, the method was assessed for forensic SNP-typing and demonstrated HRM's utility as a low-throughput technique for 1–10 SNPs [19, 20]. HRM can be an effective SNP profiling technique, provided that the SNPs have a melting temperature difference of greater than 0.5^{o}C, and do not have high guanine-cytosine (GC) content in the flanking regions [21]. HRM being a single, closed-tube assay, is less prone to contamination, which can benefit forensic analysis. In addition, the majority of forensic laboratories possess a real-time PCR instrument and an upgrade of the software would provide these laboratories with access to HRM assays [20]. HRM is simple and quick, useful not only for SNP-typing but also as a screening tool for STR-typing and species identification [22, 23].

Although many HRM technologies are available, this study compared SensiFast™ (Bioline) and MeltDoctor™ (Thermo Fisher Scientific: TFS). SensiFast produced additional melting domains and MeltDoctor generated consistent results [19]. MeltDoctor was shown to be sensitive and reproducible, requiring just 500 pg of DNA input [20]. HRM genotyping calls were made by the software. The advantage of the ViiA™ 7 RUO software (TFS) was the ease of use. The control samples test data were used to help software with accurate distinguishing of variants, especially for samples at 100pg input amount; however, HRM remained non-

reproducible at 100pg [20]. The major disadvantages of this tool are limited multiplexing ability and inability to genotype symmetrical SNPs and SNPs with high GC flanking regions [20].

Symmetrical SNP (G/C or A/T) typing using HRM is unreliable as the technique cannot clearly differentiate variants because the alternate homozygotes of these type of SNPs share similar melting temperatures [21]. GC-rich flanking regions incorporate additional melting domains and impacts on accurate genotype calling [20]. To use HRM effectively in forensics for FDP SNP-typing, appropriate SNP selection is required. If non-symmetrical SNPs with temperature differences greater than 0.5^{o}C without GC-rich flanking regions are selected, the method has the potential to genotype a large population with sufficient accuracy and reproducibility [21]. However, HRM is problematic, particularly for SNP panels with little redundancy—such as IrisPlex eye colour prediction system—in which the symmetrical SNP (rs16891982) and SNP with GC-rich regions (rs1800407) are critical to EVC inference [19, 21]. The poor performing SNPs could be replaced with others in close proximity (i.e., in linkage), although this would require new primers and potential redesign of the multiplex PCR.

In other studies, HRM assays have shown multiplexing capability up to four SNPs in a single reaction [24]. Multiplexing targets enable cost efficiency as the cost per SNP genotype for HRM assay is approximately AU$ 0.85 [20]. However, the work included in this thesis demonstrated that HRM was not successful in multiplexing six IrisPlex SNPs [20]. The preliminary data from studies presented in this thesis suggest that performing half-volume reactions is another way to reduce cost for HRM assays. However, it requires optimisation, which depends on the HRM chemistry used [20]. Different forensic laboratories use different DNA extraction procedures and HRM could be applied in such instances [20]. HRM could be used for typing small FDP panels like IrisPlex, provided the SNPs present in the panel meet the criteria of not being symmetrical, no GC-rich flanking regions and temperature differences greater than 0.5^{o}C between SNPs.

HRM is a potentially useful low-throughput SNP-typing tool, generally, for typing less than 10 SNPs. However, applicability should be considered for symmetrical SNPs, SNPs with GC rich flanking sequences, mixed source and low level DNA samples [20, 21]. The single base extension (SBE)-based SNaPshot method (TFS) can be employed to type 10–40 SNPs in single reactions, using the capillary electrophoresis (CE) detection systems currently used in forensic laboratories [25, 26]. The tool is a medium-throughput option for forensic SNP-typing.

SNaPshot has been the most common method for SNP-typing in forensics since early 2000 [25]. The workflow involves two PCR steps: the first generates the PCR products containing target regions using target-specific primers; and the second PCR detects SNP variants using SBE and dideoxy nucleotide triphosphates (ddNTPs) fluorophores. The method is more prone to contamination as it involves several post-PCR steps. SNaPshot is sensitive and reproducible, requiring as little as 62pg of input DNA [27]. The study included in this thesis demonstrated HRM's reproducibility with 100pg of input DNA [20]. SNaPshot offers multiplexing capability, with examples including the SNPforID 34-plex assays [28], 29-plex Pacifiplex [29] and 24-plex HIrisPlex assays [30]. The major advantage of the SNaPshot method is the ability to utilise the same equipment used for STR analysis—such as 3130xl or 3500xl (TFS)—in forensic laboratories. However, the workflow is time consuming compared to HRM [20].

SNaPshot has been applied to both forensic identification and FDP [25]. SNP*for*ID 52-plex assay is an example of SNaPshot-based IISNP assay [31]. This assay has been validated for routine use in forensic identification casework [32, 33]. The number and types of SNaPshot assays available enable users to adopt a hierarchical approach to the analysis of samples. Users can select LISNPs (such as 28-Y [34] and 36-mt [35] LISNP assays) to infer paternal and maternal lineages. If a sample is European, then 37-Y LISNP assay can infer specific European Y-chromosome haplogroup [36]. Further, a range of mitochondrial parental assays could be selected to differentiate continental lineages, such as the 22-mt LISNP assay differentiating nine European clades [25]. Many SNaPshot assays have been developed for BGA and EVCs SNPs typing. The SNP*for*ID 34-plex AISNP assay is a validated tool for differentiating Asian, African and European populations [37, 38]. The tool was applied in providing leads to investigators regarding a suspect of North African origin in 11-M Madrid bombing case [39]. The assay was also employed in Operation Minstead, informing investigators that the suspect was most likely of admixed African origin from the Caribbean or mainland America [40]. Similarly, the tool helped in a murder investigation in Madrid to confirm that the suspect was of Moroccan origin [41]. Other BGA SNaPshot assays, such as Eurasiaplex, differentiating European and Asian populations [42] and Pacifiplex, differentiating oceanic populations[29], have been developed. HIrisPlex [30] and IrisPlex [43] EVCs assays are SNaPshot-based PISNP tools.

The multiplexing ability of SNaPshot makes it cost effective for forensic laboratories. The approximate cost per SNP is AU$ 0.50 for typing an IrisPlex assay and there is a further cost

reduction per SNP for larger multiplex assays, such as SNP*for*ID 34-plex [20]. However, neither SNaPshot nor HRM are able to resolve mixtures. This is mainly due to the nature of bi-allelic SNPs, in which heterozygote genotypes may be indistinguishable from the genotypes of two or more contributors with alternate alleles. Further, a combination of homozygotes and heterozygotes results in a heterozygote genotype [20]. Although SNaPshot could potentially differentiate between a major and minor contributor to a mixture—based on electropherograms fluorescence intensity or relative fluorescence unit peak heights—the differing intensities of the fluorophores and the use of two PCR rounds make this difficult in practice.

The potential of FDP as a 'biological eyewitness' has received significant interest in the forensic community [10, 44]. However, this requires a large battery of FDP SNPs for making detailed phenotype inferences. The SNaPshot SNP-typing tool can be used to type 40 SNPs, but has limitations when simultaneously typing hundreds of FDP markers. Targeted amplicon sequencing with MPS and DNA barcoding strategies can be used to simultaneously type hundreds of markers in multiple samples [45, 46]. Designing new MPS-based panels may require significant investment in design and optimisation, and consideration is required for the enrichment of multiple targets. FDP SNPs are continuously being discovered and these considerations for developing larger multiplexes are becoming more immediate, with heavier reliance on manufacturers to develop commercial solutions for FDP AISNP or PISNP assays.

However, in the absence of desired commercial assays, existing SNaPshot FDP assays can be utilised with MPS. A custom approach utilising the available SNaPshot panels would be beneficial to the forensic community. This thesis demonstrated an MPS custom approach, using existing SNaPshot panels without requiring investments in primer design or assay optimisation. The custom strategy also showed the flexibility of adding and subtracting BGA and EVC markers, depending on the requirements of the individual laboratory. Two MPS benchtop technologies are available in the market: the Ion PGM and the Ion GeneStudio S5 (TFS); and MiSeq/MiSeq FGx™ (Illumina). MiSeq (Illumina) was used in the studies presented in this thesis.

SNP*for*ID 52-plex, SNP*for*ID 34-plex, Eurasiaplex, Pacifiplex and IrisPlex SNaPshot panels, with a collective total of 136 SNPs, were in the custom approach evaluation in this thesis, using the Illumina MiSeq platform [46]. The MPS custom approach was shown to be a modular and flexible option. MPS involves three broad steps: i) library preparation, in which samples are amplified for the SNPs of interest and barcoded with unique oligo-sequences, followed by

library normalisation and pooling; ii) pooled libraries (samples) are sequenced on the sequencer; and iii) data are analysed using bioinformatics analysis software [47]. Using the custom approach on the MiSeq, the observed sensitivity was 250 pg of DNA input amount—0.05 ng for each multiplex assay—using forensic standards. This approach was also applicable to degraded samples [46]. In addition, the MiSeq-based MPS method generated partial profiles for samples exposed to 60min UV radiation and 100 ng of humic acid inhibition. This method demonstrated robustness and successfully typed compromised samples, whereas Quantifiler™ (TFS) real-time PCR assay failed to detect DNA in most cases [46].

The MPS custom approach using five non-commercial PCR assays produced uneven sequence coverage, which was a common pattern on both MiSeq (Illumina) and Ion Torrent (TFS) platforms [45, 46]. The overall sequence data of the custom approach had the following characteristics: i) uneven coverage of amplicons between multiple PCR assays; and ii) uneven coverage within each assay. The consistent high and low coverage between amplicons suggests that coverage bias may be amplicon dependant. The weak association of coverage with amplicon length observed was possibly due to sequence length bias during magnetic bead clean-up steps, favouring longer amplicons. GC content is often implicated as a source of coverage bias in MPS; however, this was not observed in this study on MiSeq nor in a similar study on Ion PGM system (TFS) [45, 46]. Despite the uneven coverage observed, the MiSeq custom approach obtained an average of 98 percent genotype concordance with SNaPshot and Ion Torrent technologies. In addition, this approach on MiSeq generated 100 percent reproducible genotypes tested between four replicates of a human sample [46].

The above MPS custom approach provides options to forensic laboratories to use the existing SNaPshot panels without primer or assay optimisation. Forensic laboratories can use existing assay protocols and pool amplicons from multiple assays together for sequencing. The work in this thesis also indicated that identity SNPs can be included in the same sequencing analysis. Therefore, both forensic identity and phenotyping information can be generated from the same run. This option is useful for the analysis of degraded samples, in which STR profiling may not be informative and reduces consumption of scant evidential material from multiple analyses.

The study presented in this thesis utilised ligation-based MPS library preparation chemistries, in which adapters and DNA barcodes were ligated with the amplicons of interest. However, PCR-based commercial panels are also available, such as ForenSeq DNA Signature Prep Kit (Illumina). MPS commercial assays provide forensic laboratories a ready-made, optimised

solution for identification and/or phenotype applications. Uptake of commercial assays by forensic laboratories saves time and cost spent on optimisation and also allows for easier standardisation between laboratories. In addition, forensic laboratories get support and technical resources from commercial providers which may make implementation easier. Many commercial assays come with bioinformatics or data analysis pipelines which further benefits forensic users in getting an end to end solution. Many commercial companies also provide the option of a professional technical validation service.

Modular and flexible, MPS is capable of sequencing STRs and SNPs together, creating a complete solution for forensic DNA analyses for identification and phenotyping. To perform forensic STR sequencing on MPS platforms, obtaining sufficient DNA fragment read lengths are essential. Ion Torrent's (TFS) newly improved chemistry can sequence 600 bp reads [48]. Illumina have 2 x 300 bp chemistries for paired-end sequencing; however, their forensic genomics platform, MiSeq FGx, uses the semi-paired-end sequencing chemistry with 351bp in the forward direction and 32 bp in the reverse direction. [49]. ForenSeq DNA Signature Prep Kit comprises of both STRs and SNPs developed to run on the Illumina MiSeq FGx platform. The ForenSeq kit comprises of more than 200 markers, including 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 identity SNPs, 56 AISNPs, 24 PISNPs and an amelogenin sex marker [49]. This is an all-in-one assay, developed for forensic users for the simultaneous analysis of identification and phenotyping.

The ForenSeq kit uses a magnetic bead-based library normalisation method. The method uses a fixed volume of beads, which captures equal amounts of DNA for each library and hence an equimolar pool can be created of all libraries [50]. The study presented in this thesis compared the magnetic bead-based normalisation procedure with a qPCR normalisation method. Library normalisation is a key step in the MPS process to ensure the equimolar concentration of all libraries in the pool, enabling uniform coverage distribution. The performance of two qPCR library normalisation kits KAPA® (Roche) and NEBNext® (New England Bio Inc.) were compared. Overall, KAPA and NEBNext runs obtained higher sequencing coverage compared to the bead normalisation procedure. Universal analysis software was used for data analysis with default parameters. Excluding IISNPs, NEBNext obtained ~99 percent genotype concordance for the SRM 2800M (Promega) samples, higher than the ~89 percent concordance obtained for the magnetic bead-based normalisation procedure. The study also identified five poor performing markers in the ForenSeq kit, which appeared to be independent of the library

normalisation study, suggesting possible amplification issues [50]. Library normalisation using qPCR may also help to achieve consistency and reproducibility from run to run, which is a requirement of forensic analysis. In MPS, samples are analysed simultaneously as part of a pool, which means each sample is affected by the other samples in that pool. As such, a sample with high library yield would dominate coverage compared to a sample with low library yield. For example, sample A, with an input amount of 0.5ng template (trace DNA sample) and sample B, with 1ng input (reference DNA sample), may generate different quantities of library yield. Library normalisation using qPCR allows accurate estimation to enable equal quantities of sample A and B libraries into the sequencing pool. The average cost of qPCR library normalisation was AU$ 1.90 (half-volume reaction AU$1), which is a small fraction of the ~AU$ 140 per sample cost of ForenSeq library and sequencing reagents [50].

MPS library preparation involves a series of pipetting steps for amplicon preparation, purifications and ligation of adapters and barcodes. Given the multiple pipetting steps required and the post-PCR products used, it will be impractical for forensic laboratories to use the technique routinely without automation to avoid errors and contamination. Further work performed at Australian Federal Police laboratory established an automation pathway that can be utilised by forensic laboratories. This MPS automation path used QIA Symphony™ (Qiagen) for DNA extraction. DNA quantitation setup was performed on QIA Agility™ (Qiagen). The first step of MPS workflow was setting up PCR amplification of samples with the chosen marker multiplex. For QIA Agility (Qiagen), a simple robot was used for setting up the plate. The library preparation was done in the post-PCR lab on an EpMotion™ 5075 (Eppendorf) liquid handling system. The EpMotion system is easy to use and adaptable in creating custom protocols. Library normalisation set up and library pooling was also performed on the EpMotion robot. This automation workflow streamlines the use of MPS assays in forensic laboratories.

While there are many advantages of MPS, there are also disadvantages. The MPS custom approach included in this thesis cost US$1.4 per SNP genotype for typing 136 SNPs for 24 samples together. The analysis of more markers would further reduce the cost [46]. However, this indicates that greater numbers of markers and pooling of multiple samples is required for cost efficiency. This can become challenging for forensic laboratories that need to run smaller sample batches. Ion Torrent platforms offer scalability using their Ion chips (2M–130M reads), enabling users to choose different chips for low-throughput to high-throughput sample

requirements [51]. The Illumina MPS technology users do not have this choice and must utilise a 15GB flow cell with every run [52]. MPS running costs on any platform are expensive compared to the current STR technologies and widespread uptake of MPS will probably depend on these costs being reduced and the value of forensic phenotyping to be realised.

MPS runs generate a significant amount of data in a variety of file types. The raw sequencing data files—.DAT (TFS) or.TIFF/.bcl (Illumina)—are large in size (~ 10–300 GB per run). These raw files are converted to FASTQ (indicative file sizes are ~2–10 GB) and then aligned Binary Alignment(BAM) files (size ~3–6 GB) followed by smaller variant calling files(vcf) (size in a few kilobytes (KB)) and excel variant reports(size in the range of a few KB) [53, 54]. For routine use of MPS, an expanded storage capacity or alternate storage solutions are required as most forensic laboratories cannot accommodate these quantities of data in existing storage facilities. MPS manufacturers have availed cloud solutions to store data, such as Thermo Fisher Cloud storage [55] and Base Space Sequence Hub [56]. The forensic community may need to reach a consensus about which types of data files need to be retained. Future bioinformatics software developments may not require raw files for re-analysis. The re-analysis may start with FASTQ formats, which is possible in bioinformatics software, such as GATK [57].

Consideration is also required for the bioinformatics pipelines used for MPS data analysis. The bioinformatics pipelines involve seven broad steps: i) trim adapter sequences; ii) sort amplicon sequence by barcodes; iii) trim barcodes; iv) filter out and/or trim low quality reads; v) align to a reference genome; vi) identify variants with respect to the reference genome—STRs or SNPs; and vii) determine genotypes. Bioinformatics pipelines developed by MPS manufacturers can be a 'black box' for forensic users as the manufacturers may not disclose all aspects of the data processing. These bioinformatics pipelines can result in errors such as misalignments. For example, SNP rs1029407 was mistyped by the Illumina MiSeq Reporter software as well as the Ion Torrent Suite software, whereas the Illumina GAIIx software generated the correct AA-genotype for control 9947A (TFS) sample [46]. The MiSeq Reporter and Torrent Suite software omitted a single base in the homopolymer region, which the alignment algorithm then misaligned [46]. This demonstrates that transparency in each step of sequence data analysis is important, especially to identify potential sources of error. Within each step, the algorithm uses parameters such as quality scores during filtering base calling and alignment; strand bias, baseline thresholds; genotyping quality(GQ) scores for variant calling [54]. Amending these parameters, or requiring quality scores or filters may impact on the results obtained. Therefore,

validation and optimisation of bioinformatics pipelines is a key requirement for accurate and consistent genotyping.

The study presented in this thesis suggests HRM is more suitable as a singleplex system [20], in which laboratories only require a real-time PCR instrument with melt curve analysis software. The higher dye colour systems QuantStudio™ 5 and 6 would offer increased targets multiplexing capabilities compared to the QuantStudio 3 system [58]. Approximate costs are less than AU$ 65K, depending on the type of real-time PCR instrument selected [59]. Given that forensic laboratories use real-time PCR for DNA quantitation, these laboratories may only require HRM software upgrades, which may be less than AU$ 5K per licence [60].

In contrast, SNaPshot SNP-typing requires PCR thermal cyclers—one each in pre- and post-PCR laboratories—and the CE instrument. The cost of CE genetic analysers such as the 3500Xl (TFS) are approximately AU$250K [61]. These genetic analysers are routinely used in STR forensic identification in forensic laboratories, hence laboratories benefit from not acquiring additional costs for equipment to implement SNaPshot for SNP-typing. The multiplexing capability of SNaPshot also provides cost-saving options for medium-throughput assays. The work in this thesis provided cost estimates of AU$ 0.50/SNP (reagents only) for typing IrisPlex panel, which equates to less than AU$ 0.10 for typing SNP*for*ID 34-plex assay [20]. The cost of Ion GeneStudio S5 (with Ion Chef) from TFS and MiSeq FGx from Illumina MPS platforms is ~AU$ 140K [52, 62]. The study presented in this thesis shows the cost estimate is US$ 1.40/SNP for MiSeq sequencing based on typing 24 samples for 136 SNPs [46]. If the number of markers increased to 400 SNPs, the cost would reduce to ~US$ 0.40/SNP. Further cost reductions can be achieved by simultaneously analysing greater numbers of samples. This can be achieved using up to 384 barcodes [63] or commercially produced custom barcodes for greater than 384 samples [64].

In summary, HRM is recommended only when using a small panel, which does not have any symmetrical SNPs or SNPs with high GC flanking regions. HRM can be also utilised as a screening tool because it is simple, fast and easy to use. If a laboratory is considering the implementation of HIrisPlex [65] or SNP*for*ID 34-plex [38] type FDP panels individually, SNaPshot should be an economical and practical option. However, if a laboratory is considering running hundreds of FDP markers for multiple samples, such as employing the global AIMs panel [66], commercial panels like Precision ID Ancestry panel (TFS) [67] or combining ancestry and EVCs SNaPshot panels [45, 46], MPS is the practical option. Laboratories with

the above three technology options will benefit from the options and flexibilities these afford for casework application.

This forensic era ‘genotype to phenotype: molecular photofitting for criminal investigations’ has been possible with the discovery of a large of number of SNP markers associated with various phenotypes. SNP profiling is important for forensic DNA analysis and FDP. HRM and SNaPshot provide realistic options for specific forensic applications; however, MPS is gaining more support from the forensic community as a viable technology option for future forensic DNA analysis, including FDP. In the short-term, MPS is a useful option to supplement established STR profiling systems, while long-term implementation is considered. In summary, this thesis serves as a guide for the forensic community to choose appropriate DNA-typing methods, based on forensic applications and throughput required by forensic laboratories.

5.2.Future directions

This thesis highlighted the performance and considerations of low-, medium- and high-throughput genotyping platforms, and examined the enhanced capacity that MPS provides to numerous forensic applications, as well as the technical considerations associated with validation and implementation of this technology. Further studies should be conducted to assess various liquid handling platforms that can be used to automate the MPS workflow. Studies should also focus on additional informative BGA and EVC SNPs to FDP analysis to progress efforts towards complete, accurate ‘molecular photofit’ from DNA. Further evaluation studies should be conducted on commercial bioinformatics solutions, similar to a recent study that compared HID SNP Genotyper (TFS) and CLC Genomics WorkBench (Qiagen) for Precision ID Identity panel (TFS) [68]. Additional evaluation studies of prediction tools are required, such as a recent multinomial logistic regression, Bayesian based SNIPPER [69] and STRUCTURE [70] comparison to assist in achieving accurate BGA and EVC inferences [71]. MPS implementation for routine forensic use not only requires technical evaluations—such as the establishment of standard technical instructions and laboratory validation based on SWGDAM guidelines—but also requires the establishment of legal frameworks for using FDP lawfully [72]. MPS education and training of forensic biologists and the development of proficiency testing and quality management systems for MPS implementation are required in addition to MPS secure data storage solutions.

Third generation sequencers, including single molecule real-time sequencers like PacBio and Oxford Nanopore MinION, are also potential future candidates for forensic DNA-typing. PacBio offers read lengths of more than 1000 bp, but currently has high error rates with limited throughput capabilities [73]. A recent preliminary study on the applicability of MinIon sequencing for SNP*for*ID 52-plex typing, demonstrated that the technology and associated software is not yet ready for forensic SNP-typing [74]. With the rapid advancement in technology development, these third generation sequencers may soon overcome technology limitations, although forensic DNA analysis requires extensive evaluation and validation prior to application in casework. Being out of scope for this thesis of assessing FDP markers typing methods, the sensitivity-specificity studies of FDP markers is significant particularly, conducting assessments on high statistical weights and false positives (vs. true positives) to determine their true actionable utility.

5.3.References

1. Arnaud, C.H., Thirty years of DNA forensics: How DNA has revolutionized criminal investigations, in Chemical & Engineering News. 2017. p. 16-20.

2. Jeffreys, A.L., V. Wilson, and S.L. Thein, Hypervariable 'minisatellite' regions in human DNA. Nature, 1985a. **314**(67-72): p. 67.

3. Jeffreys, A.L., V. Wilson, and S.L. Thein, Individual specific 'fingerprints' of human DNA. Nature, 1985b. **316**: p. 75-79.

4. Aronson, J.D., DNA fingerprinting on trial: the dramatic early history of a new forensic technique. Endeavour, 2005. **29**(3): p. 126-131.

5. Butler, J., Forensic DNA Typing: Biology, Technology and genetics of STR markers. 2005, London: Elsevier Academic Press, London.

6. Edwards A, et al., DNA typing and genetic mapping with trimeric and tetrameric tandem repeats The American Journal of Human Genetics 1991. **49**(4): p. 746-756.

7. Panneerchelvam, S. and M.N. Norazmi, Forensic DNA Profiling and Database. The Malaysian Journal of Medical Sciences : MJMS, 2003. **10**(2): p. 20-26.

8. AppliedBiosystems, GlobalFiler® PCR Amplification Kit User Guide. Life Technologies Corporation, Carlsbad, CA, 2012.

9. Kayser, M., Forensic DNA phenotyping: predicting human appearance from crime scene material for investigative purposes. Forensic Science International: Genetics, 2015. **18**: p. 33-48.

10. Kayser, M. and P. de Knijff, Improving human forensics through advances in genetics, genomics and molecular biology. Nature Reviews Genetics, 2011. **12**(3): p. 179-192.

11. State of Maryland v. Kirk N. Bloodsworth, in Baltimore County. 1984. p. 84-CR-3138

12. Budowle, B. and A. vanDaal, Forensically relevant SNP classes. Biotechniques, 2008. **44**(5): p. 603-610.

13. Butler, J.M., M.D. Coble, and P.M. Vallone, STRs vs. SNPs: thoughts on the future of forensic DNA testing. Forensic Science, Medicine, and Pathology, 2007. **3**(3): p. 200-205.

14. Sobrino, B., M. Brion, and A. Carracedo, SNPs in forensic genetics: a review on SNP typing methodologies. Forensic Science International, 2005. **154**: p. 181-194.

15. Reed, G.H., J.O. Kent, and C.T. Wittwer, High resolution DNA melting analysis for simple and efficient molecular diagnostics. Pahrmacogenomics, 2007. **8**(6): p. 597-608.

16. Tong, S.Y.C. and P.M. Giffard, Microbiological Applications of High-Resolution Melting Analysis. Journal of Clinical Microbiology, 2012. **50**(11): p. 3418-3421.

17. Talmi-Frank, D., et al., Detection and Identification of Old World Leishmania by High Resolution Melt Analysis. PLOS Neglected Tropical Diseases, 2010. **4**(1): p. e581.

18. Gonzalez-Bosquet, J., et al., Detection of Somatic Mutations by High-Resolution DNA Melting (HRM) Analysis in Multiple Cancers. PLoS ONE, 2011. **6**(1): p. e14522.

19. Mehta, B., R. Daniel, and D. McNevin, High resolution melting (HRM) of forensically informative SNPs. Forensic Science International: Genetics Supplement Series, 2013. **4**(1): p. e376-e377.

20. Mehta, B., R. Daniel, and D. McNevin, HRM and SNaPshot as alternative forensic SNP genotyping methods. Vol. 13. 2017.

21. Venables, S.J., et al., Assessment of high resolution melting analysis as a potential SNP genotyping technique in forensic casework. ELECTROPHORESIS, 2014. **35**(21-22): p. 3036-3043.

22. Nicklas, J.A. and E. Buel, Improving the Efficiency of DNA Casework Analysis through Simple, Effective, PCR-based Screening Methods. 2011.

23. Malewski, T., et al., Identification of forensically important blowfly species (Diptera: Calliphoridae) by high-resolution melting PCR analysis. International journal of legal medicine, 2010. **124**(4): p. 277-285.

24. Pereyra, S., et al., Rapid multiplex high resolution melting method to analyze inflammatory related SNPs in preterm birth. BMC Research Notes, 2012. **5**: p. 69-69.

25. Mehta, B., et al., Forensically relevant SNaPshot® assays for human DNA SNP analysis: A Review. International Journal of Legal Medicine, 2016.

26. AppliedBiosystems, ABI PRISM® SNaPshot™ Multiplex Kit. Thermo Fisher Scientific. p. 1-42.

27. Walsh, S., et al., Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. Forensic Science International: Genetics, 2011. **5**(5): p. 464-471.

28. Phillips, C., et al., Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. Forensic Science International: Genetics, 2007. **1**(3–4): p. 273-280.

29. Santos, C., et al., Pacifiplex: an ancestry-informative SNP panel centred on Australia and the Pacific region. Forensic Science International: Genetics, 2016. **20**: p. 71-80.

30. Walsh, S., et al., The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. Forensic Science International: Genetics, 2013. **7**(1): p. 98-115.

31. Sanchez, J.J., et al., A multiplex assay with 52 single nucleotide polymorphisms for human identification. Electrophoresis, 2006. **27**(9): p. 1713-1724.

32. Borsting, C., et al., Performance of the SNPforID 52 SNP-plex assay in paternity testing. Forensic Science International: Genetics, 2008. **2**(2008): p. 292-300.

33. Børsting, C., H.S. Mogensen, and N. Morling, Forensic genetic SNP typing of low-template DNA and highly degraded DNA from crime case samples. Forensic Science International: Genetics, 2013. **7**(3): p. 345-352.

34. van Oven, M., A. Ralf, and M. Kayser, An efficient multiplex genotyping approach for detecting the major worldwide human Y-chromosome haplogroups. International journal of legal medicine, 2011. **125**(6): p. 879-885.

35. van Oven, M., M. Vermeulen, and M. Kayser, Multiplex genotyping system for efficient inference of matrilineal genetic ancestry with continental resolution. Investigative genetics, 2011. **2**(1): p. 1-14.

36. Onofri, V., et al., Development of multiplex PCRs for evolutionary and forensic applications of 37 human Y chromosome SNPs. Forensic science international, 2006. **157**(1): p. 23-35.

37. Phillips, C., et al., Inferring ancestral origin using a single multiplex assay pf ancestry informative marker SNPs. Forensic Science International: Genetics, 2007. **1**: p. 273-280.

38. Fondevila, M., et al., Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. Forensic Science International: Genetics, 2013. **7**(1): p. 63-74.

39. Phillips, C., et al., Ancestry analysis in the 11-M Madrid bomb attack investigation. PLoS One, 2009. **4**(8): p. e6583.

40. McShane, J., The Night Stalker - The True Story of Delroy Grant, Britain's Most Shocking Serial Sex Attacker. Kindle Edition ed. 2011: John Blake.

41. PATRICIA ORTEGA DOLZ and F.J. BARROSO, Madrid teen's suspected murderer arrested in France 18 years after crime. ELPAIS.

42. Phillips, C., et al., Eurasiaplex: A forensic SNP assay for differentiating European and South Asian ancestries. Forensic Science International: Genetics, 2013. **7**(3): p. 359-366.

43. 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 Science International: Genetics, 2011. **5**: p. 170-180.

44. Kayser, M. and P.M. Schneider, DNA-based prediction of human externally visible characteristics in forensics: Motivations, scientific challenges, and ethical considerations. Forensic Science International: Genetics, 2009. **3**(3): p. 154-161.

45. Daniel, R., et al., A SNaPshot of next generation sequencing for forensic SNP analysis. Forensic Science International: Genetics, 2015. **14**(0): p. 50-60.

46. Mehta, B., et al., Massively parallel sequencing of customised forensically informative SNP panels on the MiSeq. Electrophoresis, 2016. **37**(21): p. 2832-2840.

47. Quail, M.A., et al., A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC genomics, 2012. **13**(1): p. 341.

48. Dinauer, D., et al., P086: IMPROVED ION TORRENT SEQUENCING CHEMISTRY ENABLES RAPID TURN AROUND AND 600BP READS. Human Immunology, 2014. **75**: p. 112.

49. Jäger, A.C., et al., Developmental validation of the MiSeq FGx Forensic Genomics System for Targeted Next Generation Sequencing in Forensic DNA Casework and Database Laboratories. Forensic Science International: Genetics, 2017. **28**: p. 52-70.

50. Mehta, B., S. Venables, and P. Roffey, Comparison between magnetic bead and qPCR library normalisation methods for forensic MPS genotyping. Vol. 132. 2017.

51. ThermoFisherScientific, Introducing the Ion GeneStudio S5 series for next generation sequencing. 2018.

52. Illumina. Illumina Sequencing platforms. 2018.

53. IonTorrent, Storing Ion Torrent ™ System data and results. 2014, Thermo Fisher Scientific.

54. Illumina, MiSeq® Reporter Software Guide. 2017. p. 25-26.

55. ThermoFisherScientific. Thermo Fisher Cloud Storage. 2018.

56. Illumina. BaseSpace Sequence Hub. 2018.

57. McKenna, A., et al., The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 2010. **20**(9): p. 1297-1303.

58. ThermoFisherScientific, Applied Biosystems QuantStudio real-time PCR and digital PCR systems. 2016.

59. ThermoFisherScientific. QuantStudio qPCR Family of Instruments. 2018.

60. ThermoFisherScientific. High Resolution Melt Software v3.1. 2018.

61. ThermoFisherScientific, 3500xL Genetic Analyzer for Fragment Analysis. 2018.

62. ThermoFisherScientific, Ion GeneStudio™ S5 System. 2018.

63. ThermoFisherScientific. IonCode™ Barcode Adapters 1-384 Kit. 2018.

64. IonTorrent. Design custom barcodes. 2018.

65. Walsh, S., et al., Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage. Forensic Science International: Genetics, 2014. **9**(0): p. 150-161.

66. Phillips, C., et al., Building a forensic ancestry panel from the ground up: The EUROFORGEN Global AIM-SNP set. Forensic Science International: Genetics, 2014. **11**: p. 13-25.

67. Al-Asfi, M., et al., Assessment of the Precision ID Ancestry panel. International journal of legal medicine, 2018.

68. Meiklejohn, K.A. and J.M. Robertson, Evaluation of the Precision ID Identity Panel for the Ion Torrent PGM sequencer. Forensic Science International: Genetics. **31**: p. 48-56.

69. Phillips, C., Forensic genetic analysis of bio-geographical ancestry. Forensic Science International: Genetics, 2015.

70. Porras-Hurtado, L., et al., An overview of STRUCTURE: applications, parameter settings, and supporting software. Frontiers in Genetics, 2013. **4**(98).

71. Y Y Cheung, E., M. Gahan, and D. McNevin, Prediction of biogeographical ancestry from genotype: a comparison of classifiers. Vol. 131. 2016.

72. Koops, B.-J. and M. Schellekens, Forensic DNA phenotypong: regulatory issues. The Columbia Science and Technology Law Review, 2008. **9**: p. 158-202.

73. Berglund, E.C., A. Kiialainen, and A.-C. Syvänen, Next-generation sequencing technologies and applications for human genetic history and forensics. Investigative Genetics, 2011. **2**: p. 23-23.

74. Cornelis, S., et al., Forensic SNP Genotyping using Nanopore MinION Sequencing. Scientific Reports, 2017. **7**: p. 41759.