

SEQUENCING LONG AMPLICON MICROSATELLITE LOCI USING  
THE OXFORD NANOPORE TECHNOLOGIES  
MINION™ DEVICE

INTERNSHIP PRACTICUM REPORT

Presented to the Graduate Council of the  
Graduate School of Biomedical Science  
University of North Texas  
Health Science Center at Fort Worth  
in Partial Fulfillment of the Requirements

For the Degree of

MASTER OF SCIENCE

By

Courtney L. Hall, B.S.

Fort Worth, Texas

May 2019

## TABLE OF CONTENTS

LIST OF TABLES.....	v
LIST OF FIGURES.....	vi
Chapter	
I. INTRODUCTION.....	1
Specific aims.....	3
II. BACKGROUND.....	4
Short tandem repeats (STRs) .....	4
STR typing.....	6
Hidden variation.....	7
Mainstay DNA sequencing.....	8
Nanopore sequencing.....	12
III. MATERIALS & METHODS.....	19
Sample preparation.....	19
Length-based profiling.....	20
Primer design.....	20
PCR amplification.....	21
Nanopore library preparation.....	23
1D sequencing.....	24
Data analysis.....	25

IV. RESULTS & DISCUSSION.....	28
Designed primers.....	28
Singleplex reactions.....	28
Multiplex development.....	29
Workflow optimization.....	29
Depth of coverage.....	31
Allele designations.....	34
V. LIMITATIONS.....	38
VI. FUTURE RESEARCH.....	42
VII. CONCLUDING REMARKS.....	45
REFERENCES.....	46
Appendix	
A. SAMPLE INFORMATION.....	51
B. PRIMER DESIGN.....	57
C. PCR AMPLIFICATION.....	62
D. NANOPORE SEQUENCING.....	64
E. READ COUNT DATA.....	74

## LIST OF TABLES

<b>Table 1.</b> Forensically-relevant autosomal and Y STR loci targeted in this project.....	3
<b>Table 2.</b> Autosomal and Y primer sets in each multiplex PCR panel.....	23
<b>Table 3.</b> Autosomal STR multiplex read counts.....	32
<b>Table 4.</b> Y-STR multiplex read counts.....	33
<b>Table 5.</b> Sample information and quantification values for the twenty buccal swabs.....	52
<b>Table 6.</b> Sample genotype data.....	53
<b>Table 7.</b> Sample haplotype data.....	55
<b>Table 8.</b> Control genotype/haplotype data.....	56
<b>Table 9.</b> Autosomal loci genomic information.....	58
<b>Table 10.</b> Y loci genomic information.....	59
<b>Table 11.</b> Custom autosomal primer information.....	60
<b>Table 12.</b> Custom Y primer information.....	61
<b>Table 13.</b> PCR components.....	63
<b>Table 14.</b> PCR parameters.....	63
<b>Table 15.</b> Autosomal STR sample read counts.....	75
<b>Table 16.</b> Y-STR sample read counts.....	77
<b>Table 17.</b> Autosomal STR control read counts.....	78
<b>Table 18.</b> Y-STR control read counts.....	79

## LIST OF FIGURES

<b>Figure 1.</b> How nanopore sequencing works.....	12
<b>Figure 2.</b> Nanopore data structure.....	13
<b>Figure 3.</b> Library preparation workflow.....	14
<b>Figure 4.</b> Reads generated during nanopore sequencing.....	15
<b>Figure 5.</b> MinION™ node and flowcell anatomy.....	16
<b>Figure 6.</b> Amplicon quality assessment.....	22
<b>Figure 7.</b> Preliminary data analysis pipeline.....	26
<b>Figure 8.</b> Modified data analysis pipeline.....	27
<b>Figure 9.</b> DynaMag-2™ modifications.....	31
<b>Figure 10.</b> Alignment strategy comparison.....	34
<b>Figure 11.</b> Representative ngmlr & Sniffles results.....	36
<b>Figure 12.</b> Representative STRait Razor 3.0 results.....	37
<b>Figure 13.</b> 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108).....	65

## CHAPTER I

### INTRODUCTION

DNA evidence has long been considered the gold standard for human identification in forensic investigations. DNA typing exploits the high variability of short tandem repeat (STR) sequences to differentiate between individuals at the genetic level [1,2]. Comparison of STR profiles can be used for human identification in a wide range of forensic cases including homicides, sexual assaults, missing persons, and victims of mass disasters. In addition to autosomal short tandem repeats (auSTRs), lineage-specific markers on the Y chromosome (Y-STRs) can facilitate in the identification of male sources and may provide critical information in cases involving sexual assault mixture evidence and unestablished paternity.

Typical STR typing workflow consists of amplification followed by size-based separation and detection via capillary electrophoresis (CE) [3–5]. The power of discrimination achieved by the twenty loci in the expanded core CODIS panel is often adequate for routine forensic casework [6]. However, standard STR typing approaches may be insufficient for the deconvolution of mixed DNA profiles and some complex kinship analyses even when additional loci are interrogated [7]. The abundance of nucleotide variation observed within and around common forensic STR markers [7–15] demonstrates that sequence-level information is highly beneficial in human identity testing. Analyses based solely on repeat length fail to capture sequence-level variation that may occur in many STR loci. Detection of hidden variation at microsatellite regions of forensic interest would significantly expand upon the level of resolution realized via CE. By enabling differentiation between alleles with the same base composition but alternate motif organizations, these data have

revealed additional allelic forms that alter the diversity and distribution within a given population [16]. Ultimately, the ability to obtain full nucleotide sequences could alleviate interpretational difficulties encountered in certain types of forensic casework and also increase the discriminatory power of current suites of loci.

Recent advances in deep-sequencing technologies have made it possible to routinely identify nucleotide variations for forensic DNA typing applications. Although massively parallel sequencing (MPS) platforms have attracted significant interest from the forensic research community, their applicability to STR loci is limited by the low complexity of the target repeat sequences, restrictions in read length, and high cost of implementation [17]. Long-read data are essential for correct alignment and identification of single nucleotide polymorphisms (SNPs) in phase with the repeat motif and flanking regions. Therefore, a novel deep-sequencing approach that is not restricted to instrument-specified read lengths must be employed in order to harness the potential discriminatory power within and around both autosomal and Y microsatellites of forensic interest, increasing the resolution achieved with current STR typing techniques.

Oxford Nanopore Technologies (ONT) offers the ability to obtain the full nucleotide sequences of STR loci on a pocket-sized device known as the MinION™ [18]. DNA sequencing on an array of nanopores bypasses some major limitations, including read length restrictions and cost, of mainstay MPS platforms. Nanopore-based sequencing is scalable, portable, and capable of simultaneously interrogating the entire panel of forensic markers, making it an efficient and cost-effective alternative to MPS technologies. Adoption of this technology in forensic laboratories would preclude complete dependence on length-based genotypes, providing the most comprehensive representation of the genetic variability at STR loci. Therefore, this study aimed to assess the ability to sequence autosomal and Y STR markers of forensic interest (**Table 1**) using the MinION™ device.

**Table 1.** Forensically-relevant autosomal and Y STR loci targeted in this project.

Autosomal		Y	
D1S1656	vWA	DYS19	DYS439
TPOX	D12S391	DYS385a/b	DYS448
D2S441	D13S317	DYS389I	DYS456
D2S1338	Penta E	DYS389II	DYS458
D3S1358	D16S539	DYS390	DYS481
FGA	D18S51	DYS391	DYS533
D5S818	D19S443	DYS392	DYS549
CSF1PO	D21S11	DYS393	DYS570
D7S820	Penta D	DYS435	DYS576
D8S1179	D22S1045	DYS437	DYS643
D10S1248	Amelogenin	DYS438	GATAH4
TH01			

### Specific Aims

The following specific aims were addressed to evaluate the applicability of nanopore sequencing to a panel of forensically-relevant autosomal and Y STR markers:

**Aim 1:** Design primer sets targeting auSTRs and Y-STRs of forensic interest.

**Aim 2:** Test primers and optimize PCR amplification in singleplex and multiplex reactions.

**Aim 3:** Assess the ability to correctly identify forensic STR loci based on length and sequence using available bioinformatics tools.

**Aim 4:** Begin testing and developing a customized pipeline for the analysis of STR data produced via nanopore sequencing to determine concordance between typing results obtained to those generated by the fragment length approaches typically employed in forensic DNA laboratories.



## CHAPTER II

### BACKGROUND

#### **Short tandem repeats (STRs)**

Short tandem repeat (STR) markers are the primary genetic tool utilized in forensic DNA examinations. As the name suggests, STRs are composed of short, repetitive DNA sequences that vary in both length and pattern of the core repeat unit. The number of contiguous repeat units present at a given microsatellite locus varies significantly among individuals, and thus make them useful for human identification purposes [1,3,4,19]. The highly polymorphic nature of STRs is largely attributable to polymerase slippage during the extension phase of DNA replication [16,20–22]. This widely accepted mutation mechanism occurs as DNA polymerase pauses and briefly dissociates from the DNA molecule. The terminal end of the nascent strand reanneals to a neighboring repeat unit in either direction on the template strand. Synthesis then resumes to produce a nascent strand that is expanded or contracted by one or more repeat units. Repetitive microsatellite sequences are inherently prone to replication slippage, resulting in mutation rates several orders of magnitude higher than that of unique sequences in the human genome [20]. Forensic DNA analyses harness the considerable degree of genetic instability exhibited by STR loci for identity testing.

Microsatellite markers located on both autosomal and Y chromosomes are commonly employed in human identity testing. Given that roughly half of an individual's genetic material is inherited from each parent, suites of forensically-relevant autosomal markers (auSTRs) represent

a random shuffling of alleles. The unique allelic composition observed across multiple loci can be used to attribute DNA found at a crime scene to a known source with a high degree of confidence [19]. As highlighted by Hares [6], the expanded suite of twenty core CODIS loci provides adequate discriminatory power for routine casework and reduces the likelihood of adventitious associations in National DNA Indexing System (NDIS) database searches. In contrast to the Mendelian inheritance pattern observed in auSTRs, Y-STRs are passed down from father to son in a linear manner. Consequently, all descendants of a particular paternal lineage share a haplotype in the absence of mutational events and often cannot be differentiated using current typing techniques [23]. The discriminatory power of Y-STR profiles depend on not only the number of loci interrogated, but also the size of the relevant population databases from which the statistical frequency estimations are derived [24]. Although the haplotypes generated from Y-STRs are not as effective as the genotypes obtained from auSTRs for identification, these markers can provide essential male-specific information in sexual assault mixtures as well as corroborate auSTR testing results in father-son and sibling assessments. With expansion of the haplotype definition, many commercially available Y-STR loci employed in forensic kits can be used to differentiate between unrelated males. Some manufacturers have also included rapidly mutating Y-STRs (RM Y-STRs) to further expand upon the haplotype definition. These loci, which have mutation rates 10 to 100-fold higher than that of standard Y-STR markers, improve the resolution between unrelated males and may enable discrimination between individuals of the same paternal lineage [24]. Generally, the resolution achieved using current Y-STR testing kits and the Y-Chromosome Haplotype Reference Database [25] is sufficient for most forensic applications.

For both auSTRs and Y-STRs, the power of discrimination can be increased by expanding the number of loci queried. However, interrogation of even a reasonably large number of loci may be insufficient in certain casework scenarios because the scope of the resultant profiles is limited

by the sizing-based data generated [13]. A more comprehensive approach to harnessing the information within STR loci could further increase the power of discrimination and improve basic database searches in forensic investigations.

### **STR typing**

In order to obtain a sufficient amount of genetic material for subsequent analyses, many forensic DNA laboratories utilize a three-step enzymatic process known as the polymerase chain reaction (PCR) [5,26]. The first step of PCR consists of heating the reaction mixture to disrupt the hydrogen bonds between base pairs in the double-stranded template DNA. Following denaturation, the sequence of interest in the now single-stranded template is targeted for amplification via primer annealing. Reducing the temperature of the reaction mixture in this step enables complementary binding of both a forward and reverse oligonucleotide primer to regions flanking the target DNA sequence. In this manner, the primer set is able to define the precise location of the sequence to be amplified, and thus well-designed primers are critical for successful amplification [5]. Multiple locations within the genome can be simultaneously amplified by combining the appropriate primer sets in a multiplex reaction. A final increase in temperature stimulates the activity of a thermostable DNA polymerase which extends the bound primers using deoxynucleotide triphosphates (dNTPs) provided in the reaction mixture. The thermocycling is typically repeated 25 to 35 times. Each cycle effectively doubles the amount of targeted DNA, hence the total amount of DNA present in a sample increases exponentially [5].

Capillary electrophoresis (CE) is the standard technique used for the separation and detection of PCR amplification products in forensic casework [5,27,28]. Upon application of a voltage, the negatively charged DNA molecules present in the sample enter the capillary and are drawn towards the positive charge of the anode located at the opposite end. Larger molecules are

impeded by the polymer medium, and thus smaller molecules move towards the anode more rapidly. As the molecules migrate through a specific point, laser-excited fluorescent labels attached to the PCR primers emit light of a particular wavelength. The resultant data are visualized as a series of peaks in an electropherogram. Each peak within the electropherogram is assigned an integer value indicative of the number of complete and incomplete repeat motifs separated by a decimal point [29]. As such, alleles are defined operationally and not biologically [16].

### **Hidden variation**

Although many of the forensically-relevant STR loci exhibit a high degree of diversity using the fragment length approach, allelic designations are based on the variation generated by CE systems rather than the true variation of the underlying DNA sequence. Early studies with mass spectrometry [8,9] demonstrated that the discriminatory power of microsatellite loci could be increased using nucleotide-level variations to differentiate between alleles of identical size. The presence of polymorphisms within STR loci can have a substantial impact on the concept of allele sharing by enabling the detection of sequence variation in alleles of the same length. These changes within a particular unit of a given motif can also trigger a complex process of evolution that alters both the diversity and distribution of alleles within a population [16].

Several publications [7,11–14] have revealed that nucleotide variations occurring within some STR markers disrupt the typical repeat pattern. For instance, Gettings et al. [11] reported a doubling of the number of alleles identified by sequence analysis in comparison to CE at six autosomal STR markers, observing repeat region sequences that had not been previously reported. Expanding upon the limited sample size initially observed, Novroski et al. [13] and Wendt et al. [14] confirmed the presence of potentially informative variation within STR repeat motifs using larger scale population datasets. In all of these studies, nucleotide variations were readily observed

within the repeat regions of some forensically-relevant markers but were notably absent in others. The presence of single nucleotide polymorphisms (SNPs) in the sequences adjacent to STR repeat motifs could provide an additional level of resolution in human identify testing [10,15]. Inclusion of flanking region deep-sequencing data in forensic genetic analyses could further increase allelic diversity, improve kinship analyses, and aid in the deconvolution of mixed DNA profiles.

Collectively, these data indicate that the potential discriminatory power of both auSTRs and Y-STRs is limited not by a lack of variability within the regions, themselves, but rather by the fragment length approach currently utilized in forensic DNA examinations. The text string of sequenced nucleotides would provide the most comprehensive representation of the genetic variability at forensically-relevant markers. This would enable differentiation of alleles at the sequence level, and thus expand upon the resolution achieved using current STR typing techniques.

### **Mainstay DNA sequencing**

The dideoxynucleotide terminator method has been the mainstay approach to DNA sequencing for over 30 years. Significant improvements since the introduction of traditional Sanger sequencing in 1977 [30] have increased efficiency of the reactions and detection techniques currently implemented in research laboratories [17]. During this process, oligonucleotide primers anneal to specific regions of the denatured template DNA. The bound primers are then elongated by DNA polymerase using the available pool of nucleotides. In addition to the usual deoxynucleotide triphosphates (dNTPs), the Sanger sequencing reaction mixture consists of a relatively low concentration of dideoxynucleotide triphosphates (ddNTPs) [30]. These modified bases contain a hydrogen on the 3' carbon of the pentose sugar as opposed to the hydroxyl group found in dNTPs. Given that a phosphodiester bond cannot form between two nucleotides in the absence of this hydroxyl group, random incorporation of ddNTPs into the nascent strand terminates extension

[17]. The resultant pool contains DNA fragments of varying lengths, terminated at each nucleotide position of the template strand by a ddNTP [30]. Attachment of a unique fluorescent label to each of the four ddNTPs enables subsequent separation and detection via CE and laser excitation [28].

Over the years, Sanger sequencing techniques have been employed to achieve a better understanding of forensically-relevant STR loci. The high-quality data obtained have led to the characterization of normal, variant, and null alleles and also revealed the molecular basis of observed discordances between CE-based STR typing kits [31–34]. Although still considered the gold standard for DNA sequencing, routine application of this method in forensic casework is impractical because it requires the physical separation of heterozygous alleles prior to sequencing and suites of loci cannot be multiplexed [11]. The advent of massively parallel sequencing (MPS) presented the ability to sequence many genomic regions in a single reaction with relative ease. Consequently, MPS platforms, such as the Illumina MiSeq<sup>®</sup> (San Diego, CA, USA) and Life Science Technologies Ion Torrent<sup>™</sup> (Carlsbad, CA, USA), have largely supplanted the dideoxynucleotide terminator method for the generation of DNA sequence data. The potential to access all of the information available within STR loci has persuaded numerous research groups to assess the applicability of MPS technologies to forensically-relevant markers. In addition to the detection of hidden variation described above, genotypes that are concordant with size-based allelic designations have been derived from MPS deep-sequencing data [11,13–15]. These results indicate that MPS platforms are capable of providing not only reliable STR profiles, but also elucidate an additional level of variation present in the underlying sequence, increasing allelic diversity and maximizing the discriminatory power of current marker systems.

The most common MPS platform currently employed for research and forensic applications is the Illumina MiSeq system. This sequencing by synthesis method relies on the reversible incorporation and subsequent detection of fluorescently-labeled terminator nucleotides

via bridge amplification [35]. MiSeq library preparation consists of DNA template fragmentation followed by ligation of forward and reverse adapter sequences to both fragmented ends of the molecule. Complementary binding between the attached adapters and the primers present on the surface of the flow cell results in immobilization of the denatured template strands. Following unidirectional elongation from the primer on the flow cell by DNA polymerase, the double-stranded molecule is denatured and the original template is washed away. The extended molecule bends and hybridizes to the reverse PCR primer on the flow cell surface. Denaturation of the bridged strands results in two copies of the DNA molecule that are tethered to the flow cell surface. This process is repeated numerous times to produce clusters containing millions of copies of the template fragment. The flow cell is then flooded with fluorescently-labeled terminator dNTPs and the incorporated nucleotide is determined based on the characteristic signal emitted. The fluorescent labels and elongation-terminating dNTPs are removed, allowing elongation and sequencing to continue.

Ion semiconductor sequencing differs from the dideoxynucleotide terminator method and other MPS platforms in that modified nucleotides and optics are not utilized for detection and determination of sequence composition, placing it in a category between second- and third-generation technologies [17]. The Ion Torrent system is based on the detection of hydrogen ions released during emulsion PCR [36]. The surface of the three-layer sequencing array chip consists of millions of micromachined wells each possessing a single copy of the DNA template. DNA sequencing is accomplished by sequentially flooding the array chip with each of the four nucleotides. Introduction of a dNTP complementary to the leading template nucleotide results in elongation of the nascent strand. Incorporation of a given nucleotide is accompanied by a release of a hydrogen atom that alters the pH of the solution within the ion-sensitive layer beneath the wells. This change is detected by the highly sensitive pH meter at the base of the chip. The specific

nucleotide incorporated during each sequencing cycle is determined and recorded for the individual wells on the array chip.

The MiSeq and Ion Torrent offer deep coverage with unparalleled read accuracy in high-complexity genomic regions [17,37]. Despite significant advances in library preparation methods and technological improvement in sequencing chemistries, depth of coverage and strand completion is greatly reduced in low-complexity regions, such as the microsatellites targeted in the current project [17,37]. These platforms are also restricted to reliable read lengths ranging from approximately 150 base pairs to 400 base pairs [17]. Although sufficient for capturing the STR repeat motifs, the ability to generate longer read data encompassing the surrounding sequences would improve subsequent alignment strategies and enable accurate identification of nucleotide variations within the surrounding sequences. Therefore, the entire repeat motif as well as the flanking regions must be sequenced as a whole.

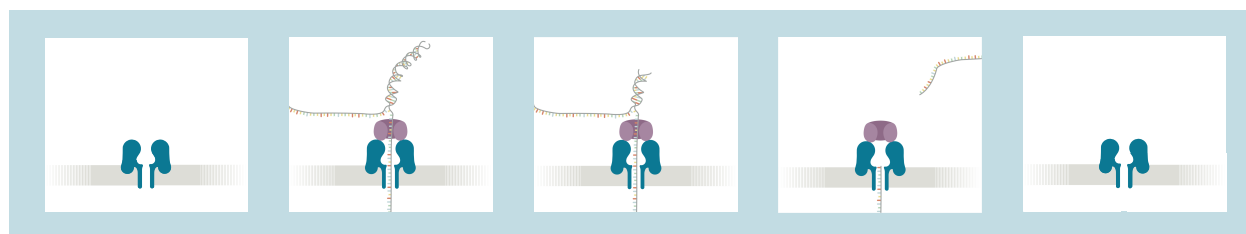
In addition to technological limitations, the cost of mainstay MPS platforms presents a major obstacle to implementation in routine casework. Laboratories conducting forensic DNA analyses have invested a significant amount of resources acquiring and validating CE-based STR typing kits and instrumentation. Although industry competition has resulted in a substantial decrease in price, a vast majority of forensic laboratories cannot allocate the funding needed to simultaneously maintain current STR typing workflows and implement MPS platforms. Another factor hindering widespread adoption is the challenges sequencing and subsequent data interpretation pose to the forensic DNA analysts. The development and validation of forensic-specific library preparation kits and platforms have aimed to streamline efficiency and facilitate adoption in routine casework. However, MPS platforms have a much more involved workflow and steeper learning curve than that of CE-based STR typing.



The DNA sequencing approaches discussed herein are not feasible for routine casework due to the technological restrictions and high cost of implementation. Widespread adoption of sequence-based STR typing in forensic laboratories would require a cost-effective alternative to MPS platforms capable of producing longer read lengths data for accurate and reliable genotyping.

## Nanopore sequencing

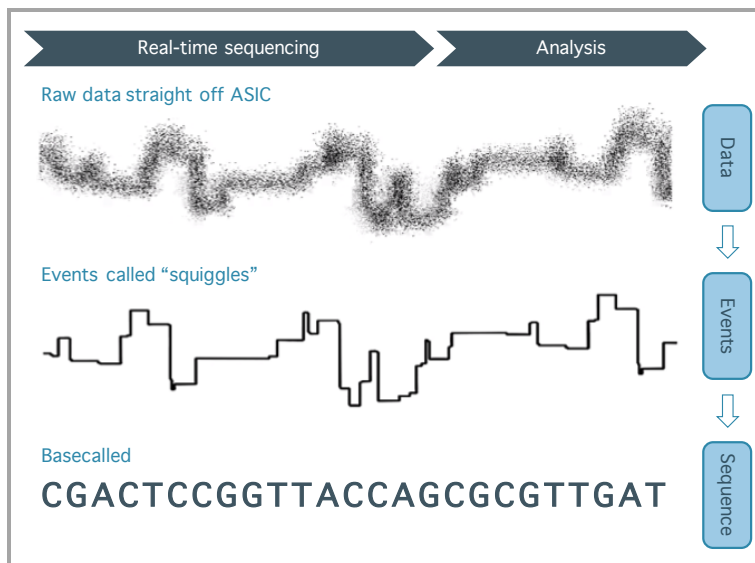
In 1989, David Deamer proposed a DNA sequencing method that relies on a voltage-bias to draw a single-stranded molecule through a nanoscopic pore [38]. Roughly 25 years later, Oxford Nanopore Technologies (ONT) introduced the MinION™ device as the first nanopore-based sequencing platform [39]. DNA sequencing on an array of nanopores enables long-read data to be generated in real-time. This technology is fully scalable and available at a relatively low cost, making deep-sequencing data accessible at various levels of funding and environmental conditions. Deamer's seemingly implausible idea has led to one of the most revolutionary deep-sequencing platforms to date.



**Figure 1.** How nanopore sequencing works. The DNA is directed to an available nanopore (blue) by an attached motor protein (purple). Upon application of an electric voltage across the synthetic lipid bilayer (grey), one strand is pulled through the pore as the incoming double-stranded DNA is unzipped by the helicase activity of the motor protein. Following translocation of the strand and disassociation of the motor protein, the nanopore becomes available for sequencing additional strands. Adapted from [40].

The core unit of ONT's sequencing technology consists of a nanoscopic hole created by a pore-forming protein (**Figure 1**) [38,39,41]. Nanopore sensing refers to the detection of molecules coming into contact with this protein [42]. A constant voltage applied across the membrane

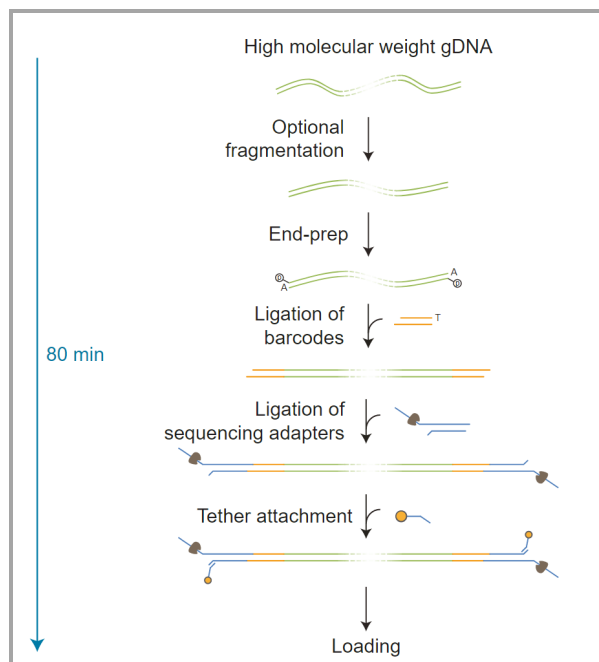
produces an ionic current through the pore. Initial disruption of the baseline current occurs upon contact between the DNA strand and nanopore. The helicase activity of the motor protein unwinds the double-stranded DNA while ratcheting the nucleotides through the small pore diameter, base-by-base [38]. Translocation of the molecule through the nanopore causes a conformational change in the protein that produces a series of current measurements, referred to as squiggles (**Figure 2**). [43]. Each nucleotide structure causes a unique current disruption that can be decoded to produce the sequence of the DNA using the base caller integrated into the MinKNOW™ software. The resultant nucleotide sequences can then be used in further data analyses.



**Figure 2.** Nanopore data structure. The raw data depicted in this figure represents a direct measurement of changes in the baseline ionic current as a molecule is translocated through a given pore. MinKNOW™ not only records, but also processes the raw signals from squiggles to the string of nucleotides in the read. Adapted from [44].

Prior to data collection, a DNA library must be prepared for sequencing (**Figure 3**). The process of library preparation begins with optional DNA repair and fragmentation followed by the addition of a dA-tail. The DNA repair and fragmentation steps may be bypassed for amplicons generated via PCR because these sequences are undamaged and of optimal length. Ligation of

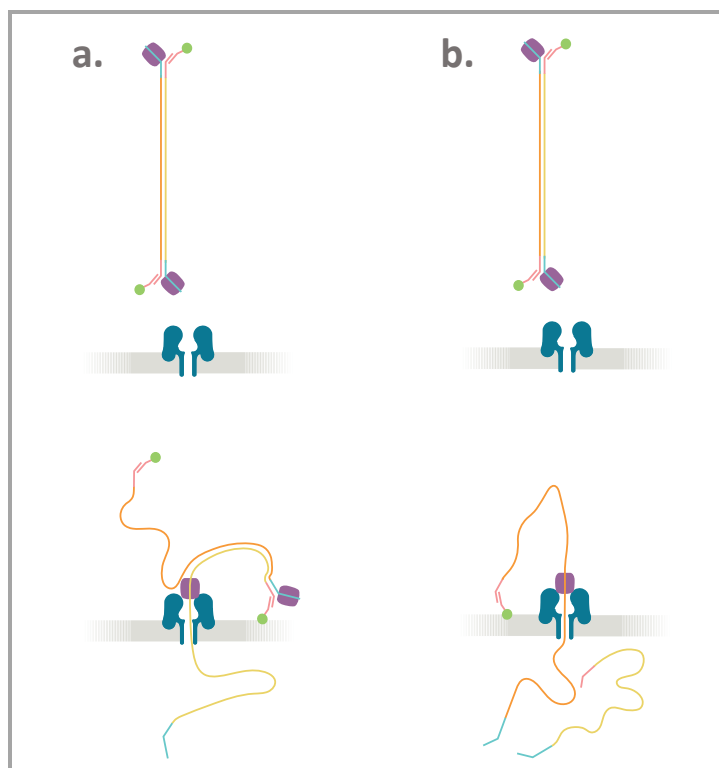
unique barcode adapters after dA-tailing enables multiple samples to be sequenced simultaneously. Sequencing adapters, which facilitate strand capture and loading of the molecular motor protein, are then ligated to both ends of a DNA molecule [39].



**Figure 3.** Library preparation workflow. Following PCR and purification, the amplicons were prepared to produce 1D reads. This protocol involves ligations of unique sample barcodes and Y-sequencing adapters onto end-repaired, A-tailed fragments. Attachment of tethers immediately prior to loading facilitates DNA capture during nanopore sequencing. Figure from [45].

Upon capture, the motor protein begins processing along the template strand of the DNA molecule. For the 1D reads (**Figure 4a**) utilized in the current project, the tether disassociates after the template strand is translocated through the pore, making the pore available to sequence another molecule. In this mechanism, the nanopore reads only one strand, and thus produces only template reads. In an alternate library preparation method (**Figure 4b**), the complement strand is tethered to the electrically-resistant membrane as the template strand is sequenced. Following translocation of the template strand through the pore, the complementary strand is directed to the pore and the

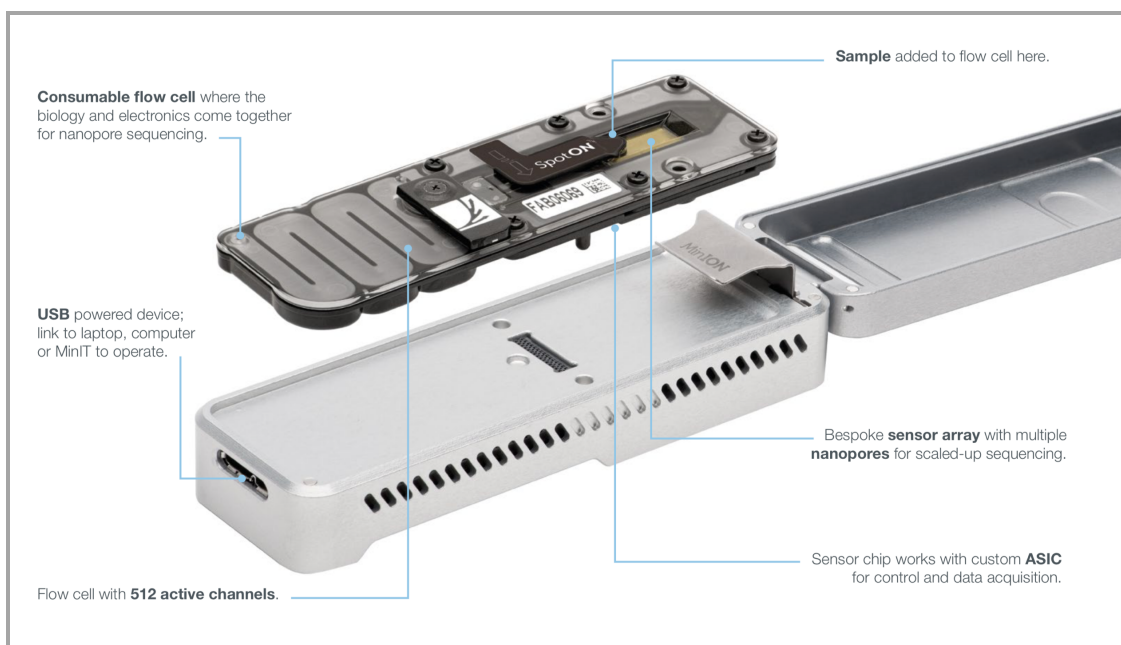
sequencing process is repeated. This 1D<sup>2</sup> library preparation results in reads with a higher accuracy, but a lower throughput than that of 1D reads [39].



**Figure 4.** Reads generated during nanopore sequencing. **(a)** For 1D reads, only the template strand (yellow) is translocated, and thus sequenced. The complement strand (orange) is released on the *cis* side of the pore. **(b)** In contrast, 1D<sup>2</sup> reads enable both strands to be sequenced by tethering the complement strand to the membrane. Following translocation of the template strand through the pore, the complement strand is drawn in and the sequence process is repeated. Adapted from [46].

Each individual nanopore channel is controlled and measured by an Application-Specific Integrated Circuit (ASIC) [41]. The sensory array chip along with all reagents for sequencing are contained in a flow cell cartridge. Following preparation, the DNA library is loaded directly onto the sensory array chip through the SpotON™ port and the flow cell is inserted into a node for sequencing and subsequent data collection. ONT currently offers nodes in four different forms – the SmidgION™, MinION™, GridION™, and PromethION™ [40]. The available devices reflect

the scalability nanopore sequencing solutions ranging from miniature to high-throughput installations, respectively. The MinION™, which was utilized in this project, is a pocket-sized device that weighs less than 100 grams and can be connected to a computer via USB port (**Figure 5**) [39]. To increase the scalability of the MinION™ device itself, ONT recently introduced a flow cell adapter, known as the Flongle™, for smaller tests and experiments [47].



**Figure 5.** MinION™ node and flowcell anatomy. The key components of the pocket-sized sequencing device are described in the image above. The MinION™ weighs under 100 grams and plugs into a laptop computer via high-speed USB cable. Figure from [40].

ONT nanopore-based sequencing requires reagents and instrumentation amounting to a small fraction of the cost required for implementation of mainstay MPS platforms [17]. The range of devices available provide a scalable method capable of producing deep-sequencing data that is not restricted by instrument-specified read lengths, input concentrations, or structure of the underlying nucleotide sequence. These features make DNA sequencing on an array of nanopores promising for application to forensic STR profiling conducted in both the typical laboratory settings and on-site at crime scenes.

Despite significant advances since the initial 2014 release, the relatively high raw read error rate of the nanopore sequencing platform is often cited as a major obstacle to subsequent data analysis [43,48,49]. Improvements in the library preparation reagents, flow cell sensors, and base recognition software have resulted in an overall increase in raw read accuracy rate, but it still falls short in comparison to that of some short-read MPS platforms [43,48]. Furthermore, particularly high error rates have been observed in homopolymeric and low-complexity sequences, such as those targeted in this project [48]. Cornelis et al. recently investigated the applicability of nanopore sequencing to forensic STR [50] and SNP [51] profiling using the MinION™ device. Although the results obtained confirm the potential usefulness for SNP detection, only partial STR profiles could be extracted from the data using both a sequence-based and an amplicon length-based approach. The research group attributes failure to produce conclusive forensic STR profiles to the high error rate of nanopore sequencing, indicating that deep-sequencing data produced via the MinION™ device is subpar in comparison to that of other MPS platforms for STR typing purposes. In an earlier publication, Zascavage et al. [49] reported a consensus call error rate of less than 1% in the mitochondrial DNA sequencing data obtained with the MinION™ device. The error rate was further reduced to 0.30% using a modified base-calling algorithm for homopolymeric regions. In addition to dismantling the common misconception that the nanopore sequencing platform is highly error-prone, these results indicate that a comparable level of accuracy can be achieved for low-complexity regions, such as STRs.

Discordance between the results obtained for genomic regions posing similar challenges to nanopore sequencing may stem from the experimental design employed by Cornelis et al. rather than the device itself. The STR loci of interest were targeted for multiplex PCR amplification using primer sets originally designed for detection via CE. Therefore, amplicons generated for the one sample tested were approximately 150 to 250 base pairs in length. As discussed above, key

advantages of nanopore platforms include the ability to generate long read data for multiple samples in a single sequencing run. Failure to harness the unrestricted read length capability of this platform may have complicated subsequent alignment attempts. This shortcoming, along with the success of SNP profiling and low error rate observed in the homopolymeric regions of the mitochondrial genome, warrant further investigation into the application of this platform to forensically-relevant STR loci.

## CHAPTER III

### MATERIALS & METHODS

#### **Sample preparation**

The results presented in this paper are based on sequencing data from twenty unrelated individuals, one control DNA sample, and three NIST-traceable standards (female,  $n = 11$ ; male,  $n = 13$ ). Human genomic DNA was extracted from twenty buccal swab samples (**Appendix A – Table 5**) with the QIAamp® DNA Mini Blood Kit (Qiagen, Hilden, Germany) as per manufacturer's protocol [52]. DNA extracts were eluted to a final volume of 50  $\mu\text{L}$  and quantified on the Applied Biosystems® 7500 Real-Time PCR System using the Quantifiler™ Trio DNA Quantification Kit (Thermo Fisher Scientific) [53]. Samples were then normalized to a concentration of 0.1 ng/ $\mu\text{L}$  according to the values obtained for the small autosomal and Y targets (**Appendix A – Table 5**). Buccal swab samples used in this study were collected and maintained under an Institutional Review Board of the University of North Texas Health Science Center approved protocol (#2010-106).

Extracted DNA for single contributor reference samples were purchased directly from the respective manufacturer. Male positive control DNA 007 (Thermo Fisher Scientific) was received at the desired concentration of 0.1 ng/ $\mu\text{L}$ . Components A, B, and C of NIST Standard Reference Material® 2391c (SRM 2391c, Gaithersburg, MD, USA) were quantified on the Qubit® 2.0 fluorometer using the Qubit® dsDNA BR Assay Kit (Thermo Fisher Scientific) [54] and diluted



to a concentration of 1.0 ng/μL. The same methods were used to verify the final concentration of the traceable standards prior to downstream applications.

### **Length-based profiling**

All twenty normalized samples were amplified using the GlobalFiler™ PCR Amplification Kit (Thermo Fisher Scientific) [55]. The ten male samples normalized according to Y target quantification values were also amplified with the YFiler™ Plus PCR Amplification Kit (Thermo Fisher Scientific) [56]. Half-reactions, which have been previously validated for the purpose of genotype analyses, were ran on the GeneAmp® PCR System 9700 and typed via capillary electrophoresis (CE) on the Applied Biosystems® 3130xl Genetic Analyzer (Thermo Fisher Scientific) [55,56]. Length-based genotypes were visualized using GeneMapper® *ID-X* Software v1.4 (Thermo Fisher Scientific) [57]. Resultant autosomal and Y allele designations for the twenty samples and control DNA 007 processed are reported in **Appendix A – Tables 6 through 8**. **Appendix A – Table 8** also contains length-based genotypes data obtained from the manufacturer for the NIST traceable-standards at STRs interrogated.

### **Primer design**

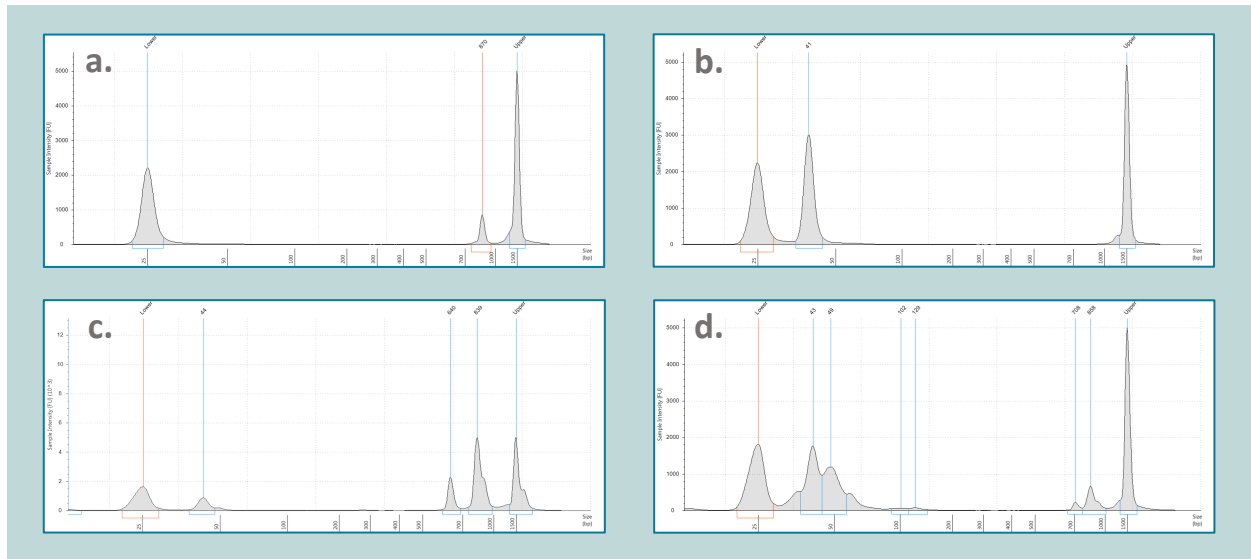
Oligonucleotide primers targeting 22 autosomal and 22 Y-chromosome common forensic STRs as well as Amelogenin were designed. Map positions of the markers in the February 2009 human reference sequence (GRCh37/hg19) were acquired from [58] and [59], respectively (**Appendix B – Tables 9 & 10**). Reference sequence fasta files containing the repeat motif with approximately 500 base pairs (bp) of flanking sequence upstream and downstream were obtained from the UCSC Genome Browser [60]. Potential primer sets for the individual fasta files were identified using the NCBI Primer-BLAST online tool [61] with the following modifications to the default parameters.

The desired PCR product size was roughly 800 bp, but amplicons between 650 and 950 bp in length were considered acceptable. The range for primer melting temperatures ( $T_m$ ) was 58.5 to 60.5°C, with a maximum difference of 1°C between two primers in a given pair. The GC-clamp was initially set to 3 and adjusted if necessary. Primer sets specific to the target genomic region in the primary reference assembly were preferred. Low complexity and repeat filters were switched off for searches that failed to return sets specific to the target region. Primer pair and STR repeat motif positions within the fasta files were visualized with Primer3Plus [62]. The web-based version of the AutoDimer software [63] was used to screen the selected autosomal and Y primer panels for primer-dimer and hairpin interactions. The primer sets and amplicon length for the long amplicon microsatellite loci targeted in this project are described in **Appendix B – Tables 11 and 12**.

### **PCR amplification**

Female (19) and male (5, 7 & 20) samples containing the highest quantity of DNA (**Appendix A – Table 5**) along with control DNA 007 were used to test the designed primers and optimize the PCR parameters. Half-reactions were prepared and amplified using the TaKaRa LA PCR Kit™ (TaKaRa Bio, Otsu, Japan) according to the manufacturer's recommendations [64]. The amount of each component within the 25 µL reaction mixture can be found in **Appendix C – Table 13**. Singleplex PCR amplification with 0.5 ng input DNA was performed on a Mastercycler® Pro S (Eppendorf, Hamburg, Germany) using the following thermal cycling conditions: initial denaturation at 94°C for 1 min, followed by 30 cycles of 98°C for 10 s, 57°C to 59°C for 30 s, 68°C for 1 min 15 s, then a final extension at 72°C for 7 min (**Appendix C – Table 14**). Both the amplification efficiency and specificity of each primer pair were assessed by running the products on a D1000 ScreenTape using the Agilent® TapeStation 4200 (Agilent Technologies, Santa Clara,

USA) [65]. As depicted in **Figure 6a**, a single peak of approximately 800 bp was indicative of successful amplification of the target region. Primer sets that exhibited a significant amount of non-specific amplification or complete amplification failure were deemed unsuccessful (**Figure 6b**). PCR parameters were adjusted on a primer-specific basis according to the results obtained. Primer sets that failed to produce successful results despite adjustments to the thermal cycling conditions were redesigned and tested as described previously.



**Figure 6.** Amplicon quality assessment. Agilent® D1000 ScreenTape electropherograms depicting successful and unsuccessful singleplex and multiplex amplification with designed primer sets. Singleplex amplification reactions were deemed successful if one peak ~800 bp was observed (**a**). In contrast, a single peak at ~40 bp indicated significant interaction between the primers in the set that resulted in complete amplification failure (**b**). Although a primer-dimer peak is present in (**c**), this multiplex reaction was considered successful because the amplicon concentration for the 7 loci pooled exceed that of the primer-dimer peak. However, this peak was significantly higher than that of the expected amplicons in (**d**), and thus the multiplex reaction was considered unsuccessful.

Multiplex PCR reactions were created by sequentially pooling primer sets and amplifying the same three samples used in the singleplex reactions. The three autosomal and four Y multiplexes developed during this project are described in **Table 2**. PCR reaction components and thermal cycling parameters employed for multiplex amplification were consistent with those reported above with the exception of an increase in template DNA from 0.5 to 1.0 ng. Amplicon quality was assessed using the methodology previously described. Successful and unsuccessful

multiplex amplification reactions are shown in **Figure 6c** and **6d**, respectively. Rather than adjusting the conditions for PCR products exhibiting a significant amount of non-specific or inhibited amplification, the previously added primer set was replaced and the multiplex was reassessed.

**Table 2.** Autosomal and Y primer sets in each multiplex PCR panel.

Autosomal			Y			
1	2	3	1	2	3	4
D1S1656	FGA	D12S391	DYS19	DYS385a/b	DYS392	DYS389I/ II
TPOX	D5S818	D13S317	DYS438	DYS391	DYS435	DYS390
D2S441	CSF1PO	D16S539	DYS448	DYS393	DYS439	DYS437
D2S1338	D7S820	D21S11	DYS456	DYS549	DYS481	DYS576
D3S1358	D8S1179	D22S1045	DYS458	DYS533	DYS570	GATAH4
TH01	D10S1248	Amelogenin			DYS643	
Penta E	vWA					
D18S51	Penta D					
D19S443						

Singleplex and multiplex products were merged by sample and purified using the QIAquick® PCR Purification Kit (Qiagen) [66] to remove remaining primers and PCR reaction components. The eluents (50 µL) were then assessed with a D1000 ScreenTape on the Agilent TapeStation 4200.

### Nanopore library preparation

The purified amplification products were prepared for nanopore sequencing using the 1D native barcoding genomic DNA kit with EXP-NBD103 & SQK-LSK108 (Oxford Nanopore Technologies (ONT), Oxford, England) [67]. Library preparation was performed with the following modifications to the standard protocol (**Appendix D**). The optional DNA repair and

fragmentation steps were bypassed because amplicons were of optimal length and undamaged. For autosomal loci generated via singleplex amplification (samples 19, 20 & 007), barcode ligation was terminated by a 10-minute incubation at 65°C, thereby eliminating a previously required Agencourt® AMPure® XP (Breckman Coulter, Brea, CA, USA) bead cleanup known to result in significant DNA loss. The pooled and barcoded samples were then concentrated using a Microcon® DNA Fast Flow Filter Device (MilliporeSigma, Burlington, MA, USA) [68] and eluted to a volume of 50 µL. Due to issues with this step (see Results & Discussion), bead purification was performed as per manufacturer's protocol in all subsequent library preparations with an additional 2.5 × wash after pooling 700 ng of the barcoded samples to be sequenced in a single run. Quantification checkpoints were performed on a D1000 ScreenTape using the Agilent® TapeStation 4200 and sample input was based on amplicon rather than total DNA concentration from Qubit® readings.

## **1D sequencing**

Prepared libraries were loaded in a drop-wise fashion into the SpotON™ port of primed vR9.4/R9.4D flow cells (FLO-MIN106/FLO-MIN106D, ONT). Flow cells were placed in the MinION™ device and sequenced for a total of 48 hours using the ONT MinKNOW™ software. The specific versions utilized varied between runs but can be determined by inspection of files in the data folders. Basecalling was performed in real-time using the local base caller integrated into the MinKNOW™ software program.

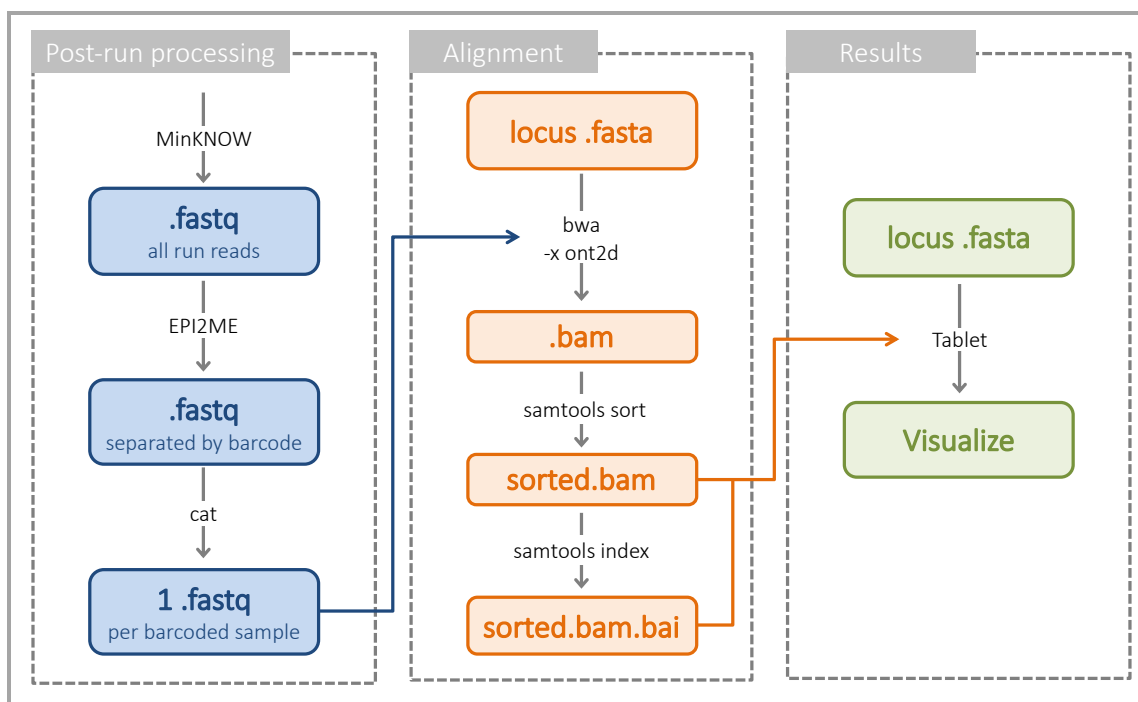
## Data analysis

Preliminary evaluation of the data obtained was accomplished using the pipeline described in **Figure 7**. Base-called fastq files generated via MinKNOW™ were separated by barcode in EPI2ME™ (ONT) [69] and merged by sample using the concatenate command. The merged fastq files were individually aligned to the locus reference sequence by the Burrows-Wheeler Aligner (bwa) mem v0.7.15 [70] supplemented with the -x ont2d setting [71]. Samtools v1.3.1 was then used to sort and index the bam files [72] and resultant alignments were visualized with Tablet v1.17.08.17 [73]. The depth of coverage at a given STR locus was visually assessed to determine successful sequencing for initial data analysis purposes. Satisfactory preliminary results for the singleplex amplification products warranted continuation with the development of autosomal and Y multiplex reactions. Individual multiplex sequencing data were processed and evaluated in the same manner as described above for singleplex amplicons. Following successful development of multiplex reactions, the remaining 18 DNA extracts and three NIST-traceable standards were amplified and sequenced on the MinION™ device.

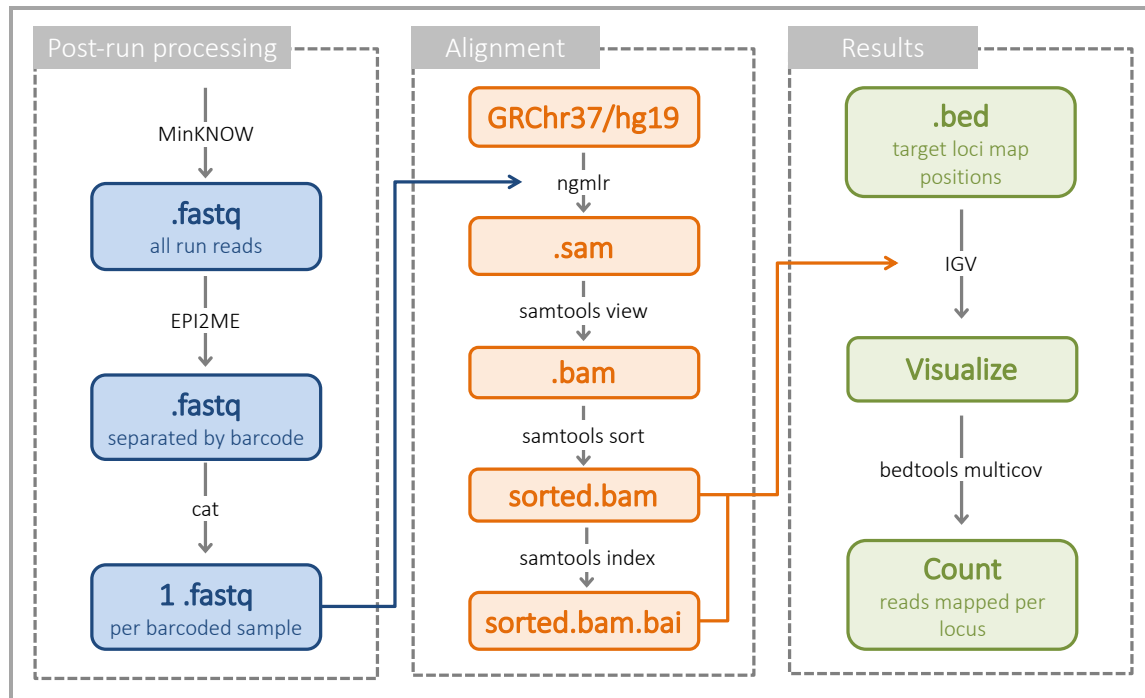
Although sufficient for preliminary evaluations, this bioinformatics pipeline was inefficient and unreliable (see Results & Discussion). Analysis of the of the resultant reads was ultimately accomplished using a customized pipeline (**Figure 8**) developed in collaboration with Fritz Sedlazeck and his bioinformatics team at Baylor College of Medicine Human Genome Sequencing Center. The fastq files merged above were realigned to the GRCh37/hg19 human reference sequence assembly with NextGenMap-LR (ngmlr) [46]. The sam files generated by ngmlr were converted to bam format, sorted, and indexed using the samtools view, sort, and index functions, respectively [72]. Resultant alignments were visualized with Integrative Genome Viewer (IGV) [74] and number of reads per locus were obtained using the bedtools multicov function [75]. Structural variations (SVs) in the ONT data were detected using Sniffles [46]. The

output files were manually assessed to determine the efficacy of this method for identifying insertions and deletions at STR loci relative to the reference genome used for alignment.

A coherent data analysis pipeline for STR allele determination specific to nanopore sequencing data has yet to be developed. Although Sniffles will likely be implemented in future studies, use of the short tandem repeat allele identification tool (STRait Razor) was explored herein. Length-based and sequence-based allele designations were evaluated with v3.0 of this bioinformatics suite using both the Forenseq and Powerseq configuration files [76]. The output was then analyzed using an Excel-based workbook provided by the developers [76].



**Figure 7.** Preliminary data analysis pipeline. The command or software utilized at each step of the process is indicated on the arrows connecting the output file boxes. The final step of the preliminary pipeline depicted in this figure was to visually assess the coverage at each locus.



**Figure 8.** Modified data analysis pipeline. The command or software utilized at each step of the process is indicated on the arrows connecting the output file boxes. Note that the fastq files resulting from post-run processing were used as the input for STRait Razor 3.0 analyses.



## CHAPTER IV

### RESULTS & DISCUSSION

#### **Designed primers**

Primer sets targeting 22 autosomal STRs, 22 Y-STRs, and the sex-determining marker, Amelogenin, were designed herein. Six of the 45 original pairs were redesigned due to significant non-specific amplification or complete amplification failure. Primer sets were successfully created for 100 percent of the loci targeted at the outset of this project. The length of resultant amplicons ranged from 684 bp to 904 bp with most falling within 35 bp of 812 bp. The melting temperatures of all customized primers were between 58.5°C and 60.5°C. While the overall difference between primer pairs in the panels was 2°C, the maximum difference between individual primers in a given pair did not exceed 1°C. Hairpin formation and primer-dimer interactions for the autosomal and Y panels assessed using the web-based AutoDimer software program suggested minimal interactions within and between the designed primer pairs. Weak interactions along with similar melting temperatures enabled development of multiplex PCR amplification reactions.

#### **Singleplex reactions**

Designed primer pairs were tested in singleplex amplification reactions. Sets yielding sufficient PCR product of the desired length with minimal amplification of non-target regions were considered successful. The results obtained following redesign of problematic sets indicate that

each primer pair amplified its respective target. All target loci were successfully amplified in singleplex reactions for six samples.

### **Multiplex development**

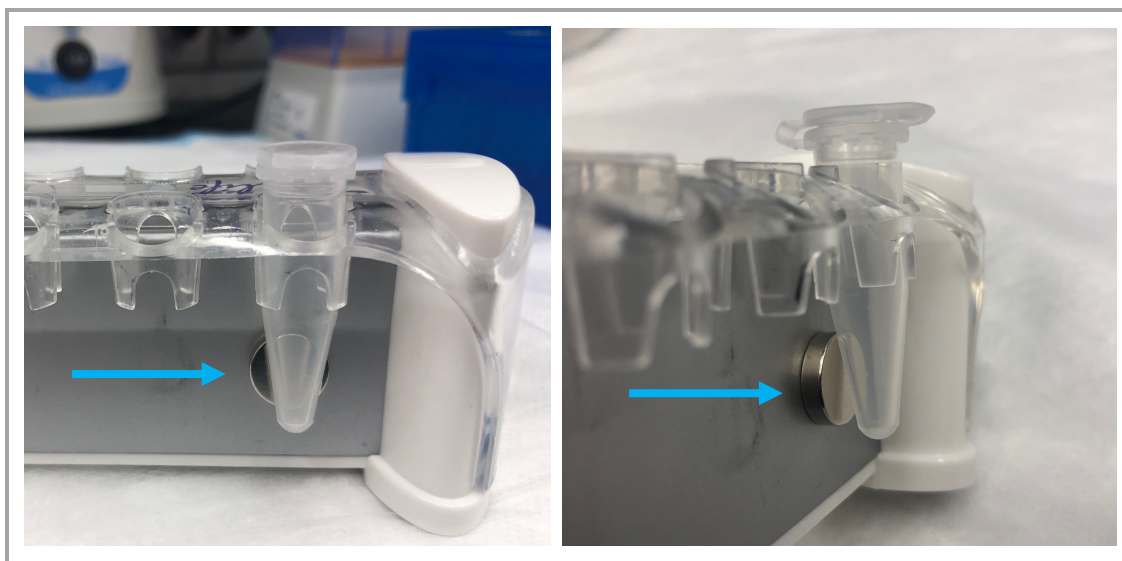
Multiplex PCR reactions for the autosomal and Y STR panels were successfully created in this project. Due to the relatively uniform size of products amplified by the primer sets designed, amplicons for each target could not be resolved in the electropherograms generated using the Agilent TapeStation 4200. Therefore, larger peaks in the representative set of autosomal and Y multiplex reactions generally encompass amplicons from three or more of the target loci. Following amplification, individual multiplexes were barcoded and sequenced on the MinION™ device to ensure that all target loci were successfully amplified. The raw read counts obtained from these sequencing runs indicates that each locus within the three autosomal and four Y multiplex panels were adequately represented (**Tables 3 & 4**, respectively). Amplification via multiplex reactions reduced the number of reactions per sample from 44 to 7, or roughly 84 percent, and thus significantly decreased the amount of time and money spent on this step.

### **Workflow optimization**

Optimization of the steps required to process a DNA extract through nanopore sequencing for forensically-relevant STR loci was initialized during this project. All PCR components and thermal cycling conditions were consistent between amplification reactions with the exception of input DNA and annealing temperature, respectively. The amount of DNA included in each reaction was increased from 0.5 ng to 1.0 ng for multiplex reactions due to the increased number of loci undergoing amplification. In order to decrease interactions between primer pairs and amplification of non-specific targets, annealing temperature was increased from 57°C to 59°C for the autosomal

multiplex 3 and all of the Y multiplexes. Amplicons were merged by sample and purified to remove remaining primers and PCR reaction components. Inclusion of this additional step resulted in the elimination of all non-target products less than 100 bp in length and residual primer/primer-dimer content from the amplification process, yielding high quality target amplicons of sufficient concentration for all of the loci.

As mentioned above, termination of barcode ligation during preparation of the singleplex autosomal amplicon library was accomplished by a 10-minute incubation at 65°C rather than the AMPure® XP bead cleanup step described in the protocol. After equal amounts of each barcoded sample was pooled to 700 ng, the sequencing library was concentrated using a Microcon® DNA Fast Flow Filter Device. Quantification of the pooled and barcoded samples on the Qubit® 2.0 fluorometer using the Qubit® dsDNA BR Assay Kit indicated roughly 65 percent sample loss. The results obtained were confirmed on a D1000 ScreenTape using the Agilent® TapeStation 4200. In order to recover DNA presumably retained by the filter, nuclease-free water was heated to 98°C, applied directly to the membrane, and incubated at room temperature for 10 minutes. Following a brief vortex at full speed, the filter device was inverted into a clean collection tube and centrifuged at 1000 ×g for 3 minutes. Comparison of the results obtained before and after attempted recovery suggests that the barcoded samples were, in fact, trapped within the membrane. In order to decrease loss during these steps, the magnetic force of the DynaMag™-2 (Thermo Fisher Scientific) was increased by addition of a circular magnet at the point of tube contact (**Figure 9**). Barcode ligation was terminated by bead purification in all subsequent library preparations with this modification to the magnet used for pelleting steps.



**Figure 9.** DynaMag™-2 modifications. In order to increase magnetic force, a circular magnet (blue arrow) was added to the device at the point of tube contact. This resulted in tighter pelleting, and thus reduced DNA loss during AMPure® XP bead purification steps.

### Depth of coverage

Amplification and nanopore-based sequencing of STR loci was assessed based on the number of reads mapping to each locus (**Appendix E – Tables 15-18**). The depth of coverage for barcoded, PCR-enriched samples ranged from 24 to 182160 ×. The lowest coverage at a single locus in any of the samples was at TH01. Generally, the TH01 primer sets did not amplify the target region as efficiently as other designed pairs. Although accurate genotype determination is unlikely with 24 × coverage, this is an outlier in the data set at hand. The number of reads mapping to TH01 were lower than that of the other loci (up to 369 ×) for multiplex amplification, but comparable for data generated via singleplex PCR reactions. These results suggest that the relatively low coverage observed at TH01 was due to interactions during multiplex amplification. It is possible that this PCR bias caused preferentially amplified loci to outcompete the relatively low number of TH01 amplicons for pore access during nanopore sequencing. While multiplex amplification and sequencing reduced the amount of resources expended, it also contributed to between-run variability in the number of reads mapping to each locus.

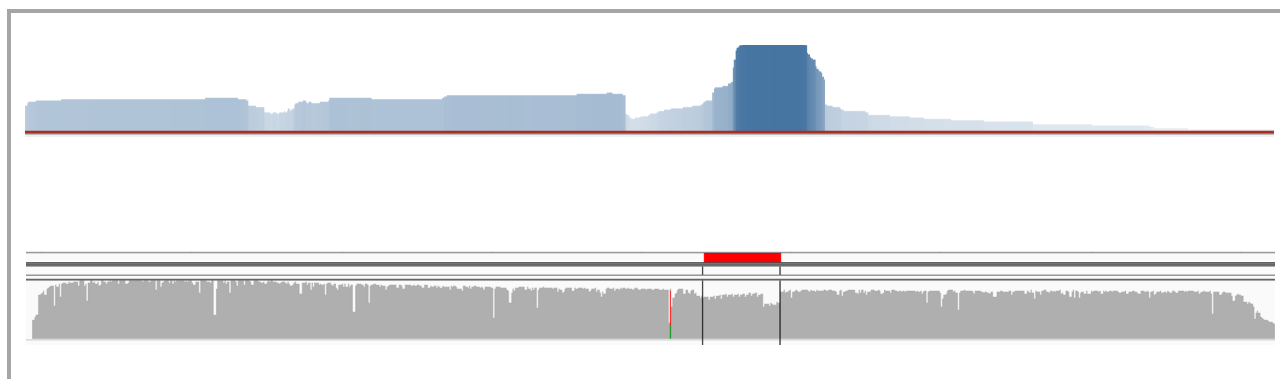
**Table 3.** Autosomal STR multiplex read counts. Number of reads aligning to each locus in the three autosomal multiplexes grouped by color.

Locus	Multiplex 1	Multiplex 2		Multiplex 3	
	19	19	20	007	007
D1S1656	12672	8	6	9	40
TPOX	9683	10	3	8	30
D2S441	5767	11	9	15	52
D2S1338	8552	3	3	3	31
D3S1358	11078	7	3	13	34
FGA	1003	0	0	3	5
D5S818	11499	8	6	13	26
CSF1PO	10230	10	3	12	36
D7S820	10857	6	7	19	32
D8S1179	7	7243	3675	13512	52
D10S1248	9	9270	4390	18954	40
TH01	3	11626	6476	21600	40
vWA	1	17267	10185	30744	36
D12S391	3	7800	4742	20420	44
D13S317	5	17583	10918	34676	37
Penta E	13	11364	7385	20997	59
D16S539	14	19580	9761	32490	46
D18S51	6	6	2	6	17076
D19S433	6	8	2	14	19910
D21S11	4	7	3	9	22950
Penta D	5	4	3	14	10735
D22S1045	1	5	1	4	5279
AMEL X	2	5	1	2	9112
AMEL Y	1	3	0	3	8371

**Table 4.** Y-STR multiplex read counts. Number of reads aligning to each locus in the four Y multiplexes grouped by color.

Locus	Multiplex 1			Multiplex 2			Multiplex 3			Multiplex 4		
	5	7	007	5	7	007	5	7	007	5	7	007
DYS19	15147	7348	28718	2	1	2	0	9	2	17	10	5
DYS438	57273	39845	77050	7	11	2	2	19	4	26	28	22
DYS448	19907	12598	38794	9	9	10	2	13	5	29	21	13
DYS456	31388	33789	58998	6	6	4	4	15	5	27	26	18
DYS458	15738	6321	25727	11	4	3	1	9	5	15	18	4
DYS385a	5	3	6	13094	10077	7499	3	5	0	10	12	7
DYS385b	6	3	1	14710	12151	8867	2	4	4	12	13	7
DYS391	3	8	13	38722	51966	25829	1	13	0	37	22	27
DYS393	17	6	12	39166	57682	35715	4	32	8	27	29	26
DYS549	8	6	8	51475	43769	28408	1	19	3	24	28	20
DYS533	8	3	8	13958	20571	10410	2	14	6	15	9	15
DYS392	19	13	21	17	13	7	18348	33847	49299	32	22	18
DYS435	10	9	16	5	5	13	12824	22500	50745	31	30	17
DYS439	8	10	11	8	11	7	4117	13866	22135	25	27	7
DYS570	11	4	17	9	4	8	12278	27350	50518	35	27	13
DYS643	9	8	13	8	9	9	9166	14784	38551	13	31	19
DYS389I	6	7	15	10	10	4	4	34	11	58611	48433	41755
DYS389II	6	7	15	10	11	4	4	35	11	58916	48697	41994
DYS390	5	6	6	8	5	7	2	23	5	40856	49454	13636
DYS437	5	9	9	11	6	3	2	28	7	5491	35214	1128
DYS576	10	7	7	6	11	2	4	12	11	48927	48085	25727
GATAH4	5	3	8	14	14	3	1	23	4	45859	51397	34181

In order to ensure that the regions of interest were not only amplified but also sequenced during a run, coverage was assessed using the two data analysis pipelines described above. Although full coverage plots were obtained for some loci interrogated by aligning all reads for a given sample to the individual reference fasta file with bwa, most resultant coverage plots visualized in Tablet were variable (**Figure 10**, top). Because amplicons spanned the entire reference sequence used for alignment and visualization, full and relatively uniform coverage was expected. In contrast to the preliminary data analysis pipeline, the expected coverage plots were obtained for all loci aligned to the entire GRChr27/hg19 human genome reference assembly using ngmlr (**Figure 10**, bottom). It should be noted that the comparison made herein is between the data analysis pipelines and not the aligners themselves.



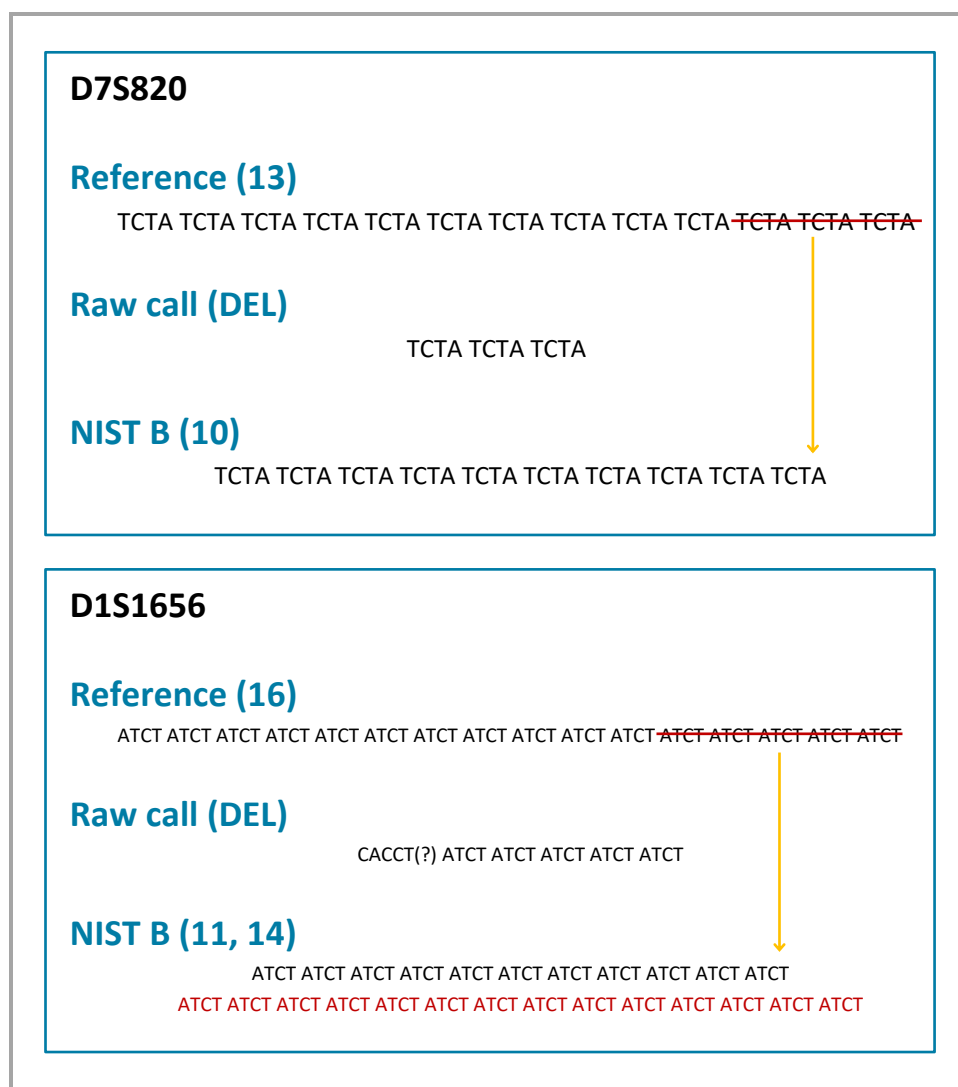
**Figure 10.** Alignment strategy comparison. The plots depict amplicon coverage at locus D7S820 for NIST B using bwa (top) and ngmlr (bottom) to align resultant reads. The height of the graph corresponds to the number of reads mapping to any position within the given region. Since amplicons spanned the entire region shown above, full, uniform coverage was expected as seen for ngmlr. Plots with variable coverage were obtained for most loci using the preliminary bwa alignment strategy.

### Allele designations

Given that a data analysis pipeline specific to STRs has yet to be developed, results presented herein are limited to algorithms designed for large chromosomal alterations (i.e., ngmlr & Sniffles) and shorter-reads generated via validated MPS platforms (i.e., STRait Razor). Generally, structural variants (SVs) detected using Sniffles corresponded to the expected insertions and deletions

(indels) relative to the reference genome. However, this tool failed to identify SVs smaller than 10 bp, and thus was not successful for indels within 2 repeat units of the reference sequence. Sniffles is not currently capable of detecting minor heterozygote alleles, so all loci appeared to be homozygous. Using STR Razor 3.0 with both Forenseq and Powerseq configuration files produced partial STR profiles for the autosomal and Y STR loci under investigation. Unlike Sniffles, which only generated raw data calls that were then manually assessed, STRait Razor 3.0 is capable of assigning both length- and sequence-based allele designations for the loci detected from MPS data. Although concordance between the expected and observed allele designations was obtained at a majority of the loci identified, the depth of coverage at each locus was significantly lower than the read count data from ngmlr, indicating that the algorithm filtered a substantial portion of potentially informative reads. Representative data for both Sniffles and STRait Razor 3.0 are described in **Figures 11 and 12**, respectively. Ultimately, the results obtained suggest that both tools evaluated may produce full profiles if optimized for STR data generated via nanopore sequencing.





**Figure 11.** Representative ngmlr & Sniffles results. Following ngmlr alignment of resultant reads for NIST B, Sniffles was used to detect insertions and deletions relative to GRChr27/hg19. Raw read calls for loci at which NIST B is homozygous (D7S820) and heterozygous (D1S1656) are depicted above. The [TCTA] motif of D7S820 (top) is repeated 13 times in GRChr27/hg19. The raw call obtained from Sniffles was identified as a deletion (DEL) of 3 [TCTA] repeat units. Removal of these repeats from the reference sequence (red line) results in 10 [TCTA] repeats, which is concordant with the certified designation for NIST B. At D1S1656 (bottom) the [ATCT] motif is repeated 16 times in the human genome reference assembly. The raw call obtained from Sniffles was identified as a [CACCT] [ATCT]<sub>5</sub> deletion (DEL). Although the [CACCT] detected is not part of the STR motif, removal of the 5 [ATCT] repeat units from the reference sequence (red line) corresponds to the 11 heterozygous allele of NIST B. The 14 allele (red sequence), which was presumably present in a lower number of reads than the 11, is completely ignored by Sniffles.

### D8S1179 (10, 13)

	Allele 1	Allele 2	
Length-based	10	13	✓
Sequence-based	10	13	✓

	Allele 1	Allele 2	
Length-based	149	130	?
Sequence-based	135	111	

	STRait Razor	ngmlr
Locus DoC	1482	6296

### D7S820 (10)

	Allele 1	Allele 2	
Length-based	9.3	10	✗ ✓
Sequence-based	9.3	9.2	✗ ✗

	Allele 1	Allele 2	
Length-based	48	34	?
Sequence-based	21	20	

	STRait Razor	ngmlr
Locus DoC	2140	6529

**Figure 12.** Representative STRait Razor 3.0 results. Following post-run processing, the merged fastq file for a given sample was assessed using STRait Razor 3.0. Output for loci at which NIST B is heterozygous (D8S1179) and homozygous (D7S820) are depicted above. The certified allele designation is listed in parenthesis. The first table under each locus contains the result allele calls, the middle table contains the number of reads supporting each call, and the bottom table contains the overall locus depth of coverage (DoC) for both Strait Razor and ngmlr. In contrast to Sniffles, STRait Razor 3.0 correctly identified heterozygote loci (top), but the number of reads supporting both the length- and sequenced-based allele designations are relatively low in comparison to both the locus depth of coverage and the number of reads mapped using ngmlr. As for D7S820 (bottom), STRait Razor 3.0 correctly identified the repeat as a 10 for only the length-based allele 2. Although a higher locus DoC is observed, the number of reads supporting the allele calls are extremely low.

## CHAPTER V

### LIMITATIONS

ONT nanopore sequencing platforms have the potential to produce reliable genotype determinations and reveal an additional level of variation within and around microsatellite loci, offering a feasible and cost-efficient alternative to MPS technologies for sequence-based STR typing. The challenges encountered during this project as well as the limitations of the technology utilized must be addressed to guide future research efforts and aid in implementation for routine forensic casework.

Although several hundred copies of autosomal and Y-chromosome markers are present within a typical 1 ng DNA sample used for typing, these sequences represent a very small fraction of the total amount of genetic material in a given sample. STR copy number in unenriched libraries is insufficient for accurate genotype determination and identification of SNPs because target strands are unable to outcompete background gDNA for pore access during sequencing runs. Amplification of the expanded microsatellite region increased the depth of coverage, and thus enabled detection of heterozygote alleles and underlying sequence variations, but the process of PCR causes additional issues in subsequent data analyses. *Taq* DNA polymerase can introduce errors during amplification at an average rate ranging from  $10^{-5}$  to  $10^{-6}$  [21]. In an attempt to decrease occurrence of PCR byproducts, TaKaRa Long and Accurate (LA) *Taq*<sup>®</sup> was utilized in all amplification reactions. This enzyme possesses 3'-to-5' proofreading activity that enables detection and removal of misincorporated bases, resulting in a fidelity that is 6.5-fold higher than conventional *Taq* polymerase [77]. Stutter artifacts, however, were still incorporated during the

thirty-cycle amplification reaction performed presumably due to polymerase slippage while copying these low- complexity repeat sequences [22,78]. Despite concordance between predicted and CE-based allele designations identified, the presence of stutter within the resultant reads required an additional level of filtering. While enrichment prior to nanopore sequencing is necessary, the use of PCR at thirty cycles is not ideal for the for the STR loci targeted in this project and given the extremely high level of coverage obtained (**Appendix E – Tables 15-18**) may be unnecessary. Future studies will evaluate the effect of reducing PCR cycling on stutter artifact formation while retaining sufficient coverage to accurately call STR loci.

As with the current forensic typing methods and MPS approaches described above, this project relied on PCR to generate a sufficient amount of the targeted regions for nanopore sequencing. Amplification increased depth of coverage, allowing for successful identification of heterozygote loci and single nucleotide changes within and around the STR markers interrogated. However, the errors and artifacts resulting from amplification of repetitive regions inherently prone to polymerase slippage complicated interpretation of the data, requiring a statistically greater number of reads to overcome these byproducts. The amount of data produced and the lack of commercially-available software packages for extraction and assessment of resultant reads is another challenge associated with this novel sequencing method. Previous attempts to generate reliable sequence-based STR typing results using the MinION™ device were confounded by the inability to properly interrogate long-read repetitive sequencing data with available alignment tools. Therefore, the project at hand required a customized data analysis pipeline that was developed by experts in the fields of bioinformatics and computational genetics. The pipeline described in this report was created by individuals who are experienced in analyzing long-read data generated via ONT platforms, up-to-date on the strengths and limitations of the various tools available, and capable of implementing new algorithms as needed. Although the process has been

clearly outlined, extraction of the desired read count information from base called files involves numerous computationally-intensive steps. Merely processing the nanopore sequencing data generated through the appropriate software pipelines requires a level of computational skill that is not common among forensic DNA analysts.

Other challenges encountered throughout the course of this project stemmed directly from the rapid evolution of the library preparation reagents, flow cell sensors, and base recognition software. As detailed in the materials and methods section, techniques developed by previous students to improve DNA recovery during library preparation actually complicated processing of singleplex amplification of autosomal loci for samples 19, 20, and 007. Even after optimization of current library preparation, the variability in the kit reagents as well as in the hardware and software utilized resulted in notable differences between sequencing runs, complicating comparison of the data generated. For instance, Revision C flow cells (FLO-MIN106) were replaced by Revision D (Rev D) flow cells (FLO-MIN106D) in the midst of data collection for this project. Rev D flow cells utilize an upgraded ASIC within the sensory array chip which extends the lifetime of a sequencing experiment. An increase in the amount of data generated was observed for some Rev D flow cells, but there was also a notable difference between Rev D flow cells received in the same shipment. With roughly 11.51 million reads, the Y-STR sequencing run for the extracted DNA samples (2, 9, 10, 12, 14, 16, 18 & 20) outperformed all other Rev D flow cell runs by over 7 million reads. Because the libraries were prepared and sequenced in the exact same manner, the disproportionately high number of reads obtained were likely due to inherent differences between the recently released flow cells utilized. In addition to flow cell issues, Native Barcode 03 (BC03) was not performing as expected. The relatively low number of reads for all samples barcoded with BC03 from the same kit suggests that the issue was caused by the barcode, itself, rather than DNA loss or other human errors during library preparation. These suspicions

were confirmed after ONT Customer Service reviewed the log files from the associated sequencing runs.

Although intended to reduce the amount of DNA required for sequencing and improve quality of the resultant data, unforeseen issues in the updates released can impact sequencing and even result in unrecoverable data. The rapid evolution of nanopore-based sequencing platforms could hinder near-term adoption in forensic laboratories given fragile nature of biological material collected from crime scenes, however, once the platform stabilizes, it may be of benefit for smaller laboratory systems and field applications.

## CHAPTER VI

### FUTURE RESEARCH

This project forms the foundation for future efforts aimed at the continued assessment and optimization of ONT nanopore-based sequencing technologies for STR typing. Studies will focus on lessening the impact of the limitations and challenges discussed above to facilitate seamless adoption of this deep-sequencing platform in forensic laboratories across the nation.

A key limitation of the current project is the use of PCR amplification to generate sufficient sample for nanopore sequencing and subsequent STR typing. Previous studies suggest that the autosomal and Y copy number present within unenriched gDNA samples are inadequate for accurate genotype determination and SNP identification. The results obtained herein demonstrate that amplification prior to library preparation ensures target loci outcompete background gDNA for pore access but also generates a considerable number of PCR-induced artifacts that complicate data interpretation. The deep coverage achieved from thirty PCR cycles suggests that reliable auSTR and Y-STR typing results can be ascertained from a considerably lower depth using the MinION™ device. In order to minimize the occurrence of PCR errors and stutter artifacts, future studies will aim to reduce the number of amplification cycles required to overcome background gDNA. Optimization of the enrichment process using the primer sets designed in this project will not only alleviate interpretational difficulties during subsequent data analyses, but also increase current knowledge of the stutter generation process and sensitivity of nanopore sequencing devices.

Another major challenge associated with the use of this novel deep-sequencing technology is the volume of data generated and the lack of available tools capable of correctly interrogating the microsatellite loci examined. The customized data analysis pipeline described in the materials and methods section was developed in order to properly align resultant reads. However, this workflow must be expanded upon to identify structural variation (e.g., STR motif repeats) and predict allelic designations at loci of interest, which is a goal of future studies. A better understanding of amplification and sequencing artifacts would provide the necessary information to develop a statistically-driven computational model capable of categorizing a given signal as a true biological allele, byproduct of PCR, or sequencing artifact. This model could then be utilized to further improve alignment of the target regions and identification of nucleotide variations within each sample, thereby aiding in identification of the contributor. Ultimately, continued development of the data analysis pipeline would alleviate some of the primary bioinformatic challenges associated with nanopore sequencing data, increasing the amount of usable data extracted from each run and streamlining efficiency for application to routine forensic casework.

Implementation of the nanopore sequencing strategies developed in future research to routine forensic casework would require validation in accordance with the Scientific Working Group on DNA Analysis Methods (SWGDAM) recommendations [79]. Protocols specific to ONT nanopore-based sequencing devices should be established and applied in developmental validation studies. As mentioned above, assessment of amplification at various cycle numbers and input DNA concentrations would provide important information about the sensitivity of nanopore sequencing devices and the occurrence of PCR-induced artifacts. Experiments involving degraded and mixed samples should also be conducted using the long amplicons targeted herein, as well as shorter amplicons, to evaluate the effectiveness of amplification approaches on resultant profile quality



and determine the value of SNPs identified in the deconvolution of mixed profiles. These experiments would provide the basis for the establishment of validated interpretation guidelines.

The application of nanopore-based sequencing to forensic investigations would make it possible to generate nucleotide-level data for the entire panel of genetic markers currently interrogated in routine human identity testing on a single, cost-effective platform. The significance of this deep-sequencing technology in forensic investigations is unquestionable. However, the challenges encountered while attempting to sequence and understand the genetic variation within microsatellite regions extends well beyond the crime laboratory. Nanopore sequencing allows for direct, long-range phasing of polymorphisms in gene regions, which could provide critical information in the context of genetic diseases and the genetic features of health disparities [80–83]. Application of the techniques developed in this and future forensic-centered projects to problems in other areas of biomedical research could provide optimized approaches to the detection of microsatellite expansion or contraction throughout the genome.

## CHAPTER VII

### CONCLUDING REMARKS

The ability to amplify and sequence autosomal and Y STR markers of forensic interest using the MinION™ device was evaluated in this proof-of-principle study using custom primer sets targeting 800 bp amplicons. The results presented herein suggest that PCR amplification followed by purification ensured that target microsatellite loci outcompeted background genomic DNA for pore access during the sequencing runs information at the 45 loci interrogated for all 24 samples assessed.

Significant progress developing and optimizing a workflow for processing DNA extracts from STR amplification through nanopore sequencing was achieved in this project. The longer-read data generated allowed for accurate alignment and will presumably enable identification of single nucleotide polymorphisms (SNPs) within and around microsatellite loci in future studies. However, both Sniffles and the short-read bioinformatic pipeline utilized to generate STR allele designations failed to produce conclusive profiles containing all of the loci interrogated. Through continued collaboration with the bioinformatics team at Baylor College of Medicine, the data analysis pipeline established for alignment and read count generation will be further developed to predict STR allele designations from the nanopore sequencing data generated. This will enable assessment of concordance between profiles obtained from traditional STR typing techniques and the longer-read nanopore sequencing data.

## REFERENCES

- [1] A. Edwards, A. Civitello, H.A. Hammond, C.T. Caskey, DNA typing and genetic mapping with trimeric and tetrameric tandem repeats., *Am. J. Hum. Genet.* 49 (1991) 746–56.
- [2] H.A. Hammond, L. Jin, Y. Zhong, C.T. Caskey, R. Chakraborty, Evaluation of 13 short tandem repeat loci for use in personal identification applications., *Am. J. Hum. Genet.* 55 (1994) 175–89.
- [3] J.M. Butler, Short tandem repeat analysis for human identity testing, in: *Curr. Protoc. Hum. Genet.*, John Wiley & Sons, Inc., Hoboken, NJ, USA, 2004. doi:10.1002/0471142905.hg1408s41.
- [4] M.A. Jobling, P. Gill, Encoded evidence: DNA in forensic analysis, *Nat. Rev. Genet.* (2004). doi:10.1038/nrg1455.
- [5] J.M. Butler, Short tandem repeat typing technologies used in human identity testing, *Biotechniques*. 43 (2007) Sii-Sv. doi:10.2144/000112582.
- [6] D.R. Hares, Expanding the CODIS core loci in the United States, *Forensic Sci. Int. Genet.* 6 (2012) e52–e54. doi:10.1016/j.fsigen.2011.04.012.
- [7] N.M.M. Novroski, A.E. Woerner, B. Budowle, Potential highly polymorphic short tandem repeat markers for enhanced forensic identity testing, *Forensic Sci. Int. Genet.* (2018). doi:10.1016/j.fsigen.2018.08.011.
- [8] F. Pitterl, K. Schmidt, G. Huber, B. Zimmermann, R. Delport, S. Amory, B. Ludes, H. Oberacher, W. Parson, Increasing the discrimination power of forensic STR testing by employing high-performance mass spectrometry, as illustrated in indigenous South African and Central Asian populations, *Int. J. Legal Med.* 124 (2010) 551–558. doi:10.1007/s00414-009-0408-x.
- [9] J. V. Planz, K.A. Sannes-Lowery, D.D. Duncan, S. Manalili, B. Budowle, R. Chakraborty, S.A. Hofstadler, T.A. Hall, Automated analysis of sequence polymorphism in STR alleles by PCR and direct electrospray ionization mass spectrometry, *Forensic Sci. Int. Genet.* 6 (2012) 594–606. doi:10.1016/j.fsigen.2012.02.002.
- [10] K.B. Gettings, R.A. Aponte, K.M. Kiesler, P.M. Vallone, The next dimension in STR sequencing: Polymorphisms in flanking regions and their allelic associations, *Forensic Sci. Int. Genet. Suppl. Ser.* 5 (2015) e121–e123. doi:10.1016/j.fsigss.2015.09.049.
- [11] K.B. Gettings, K.M. Kiesler, S.A. Faith, E. Montano, C.H. Baker, B.A. Young, R.A. Guerrieri, P.M. Vallone, Sequence variation of 22 autosomal STR loci detected by next generation sequencing, *Forensic Sci. Int. Genet.* 21 (2016) 15–21. doi:10.1016/j.fsigen.2015.11.005.
- [12] K.B. Gettings, L.A. Borsuk, C.R. Steffen, K.M. Kiesler, P.M. Vallone, Sequence-based U.S. population data for 27 autosomal STR loci, *Forensic Sci. Int. Genet.* 37 (2018) 106–115. doi:10.1016/j.fsigen.2018.07.013.
- [13] N.M.M. Novroski, J.L. King, J.D. Churchill, L.H. Seah, B. Budowle, Characterization of genetic sequence variation of 58 STR loci in four major population groups, *Forensic Sci. Int. Genet.* 25 (2016) 214–226. doi:10.1016/j.fsigen.2016.09.007.

- [14] F.R. Wendt, J.D. Churchill, N.M.M. Novroski, J.L. King, J. Ng, R.F. Oldt, K.L. McCulloh, J.A. Weise, D.G. Smith, S. Kanthaswamy, B. Budowle, Genetic analysis of the Yavapai Native Americans from West-Central Arizona using the Illumina MiSeq FGx™ forensic genomics system, *Forensic Sci. Int. Genet.* 24 (2016) 18–23. doi:10.1016/j.fsigen.2016.05.008.
- [15] F.R. Wendt, J.L. King, N.M.M. Novroski, J.D. Churchill, J. Ng, R.F. Oldt, K.L. McCulloh, J.A. Weise, D.G. Smith, S. Kanthaswamy, B. Budowle, Flanking region variation of ForenSeq™ DNA Signature Prep Kit STR and SNP loci in Yavapai Native Americans, *Forensic Sci. Int. Genet.* 28 (2017) 146–154. doi:10.1016/j.fsigen.2017.02.014.
- [16] J. V. Planz, T.A. Hall, Hidden variation in microsatellite loci: utility and implications in forensic DNA analysis, *Forensic Sci. Rev.* 24 (2012) 28–42. doi:doi.org/10.1201/b15361.
- [17] R.R. Zascavage, S.J. Shewale, J. V. Planz, Deep-Sequencing technologies and potential applications in forensic DNA testing, *Forensic DNA Anal.* 2397 (2013) 311–348. doi:https://doi.org/10.1201/b15361.
- [18] Oxford Nanopore Technologies, MinION brochure, Version: BR\_1002(EN)\_V1\_14Mar2019. (2019).
- [19] J.M. Butler, Genetics and genomics of core short tandem repeat loci used in human identity testing, *J. Forensic Sci.* 51 (2006) 253–265. doi:10.1111/j.1556-4029.2006.00046.x.
- [20] H. Fan, J.-Y. Chu, A brief review of short tandem repeat mutation, *Genomics. Proteomics Bioinformatics.* 5 (2007) 7–14. doi:10.1016/S1672-0229(07)60009-6.
- [21] J. Cline, PCR fidelity of pfu DNA polymerase and other thermostable DNA polymerases, *Nucleic Acids Res.* 24 (1996) 3546–3551. doi:10.1093/nar/24.18.3546.
- [22] D. Pumpernik, B. Oblak, B. Borštnik, Replication slippage versus point mutation rates in short tandem repeats of the human genome, *Mol. Genet. Genomics.* 279 (2008) 53–61. doi:10.1007/s00438-007-0294-1.
- [23] M. Kayser, R. Kittler, A. Erler, M. Hedman, A.C. Lee, A. Mohyuddin, S.Q. Mehdi, Z. Rosser, M. Stoneking, M.A. Jobling, A. Sajantila, C. Tyler-Smith, A comprehensive survey of human Y-chromosomal microsatellites, *Am. J. Hum. Genet.* 74 (2004) 1183–1197. doi:10.1086/421531.
- [24] K.N. Ballantyne, V. Keerl, A. Wollstein, Y. Choi, S.B. Zuniga, A. Ralf, M. Vermeulen, P. De Knijff, M. Kayser, A new future of forensic Y-chromosome analysis: Rapidly mutating Y-STRs for differentiating male relatives and paternal lineages, *Forensic Sci. Int. Genet.* (2012). doi:10.1016/j.fsigen.2011.04.017.
- [25] S. Willuweit, L. Roewer, Y chromosome haplotype reference database (YHRD): Update, *Forensic Sci. Int. Genet.* (2007). doi:10.1016/j.fsigen.2007.01.017.
- [26] K.B. Mullis, F.A. Faloona, Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reactions, in: *Methods Enzymol.*, 1987: pp. 270–279. doi:10.1016/0076-6879(93)17067-F.
- [27] J.M. Butler, E. Buel, F. Crivellente, B.R. McCord, Forensic DNA typing by capillary electrophoresis using the ABI Prism 310 and 3100 genetic analyzers for STR analysis, *Electrophoresis.* (2004). doi:10.1002/elps.200305822.
- [28] H. Drossman, J.A. Luckey, A.J. Kostichka, J. D’Cunha, L.M. Smith, High-speed separations of DNA sequencing reactions by capillary electrophoresis, *Anal. Chem.* 62 (1990) 900–903. doi:10.1021/ac00208a003.

- [29] K.B. Gettings, R.A. Aponte, P.M. Vallone, J.M. Butler, STR allele sequence variation: Current knowledge and future issues, *Forensic Sci. Int. Genet.* 18 (2015) 118–130. doi:10.1016/j.fsigen.2015.06.005.
- [30] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, *Proc. Natl. Acad. Sci.* 74 (1977) 5463–5467. doi:10.1073/pnas.74.12.5463.
- [31] C. Allor, D.D. Einum, M. Scarpetta, Identification and characterization of variant alleles at CODIS STR loci, *J. Forensic Sci.* 50 (2005) 1–6. doi:10.1520/JFS2005024.
- [32] M.C. Kline, C.R. Hill, A.E. Decker, J.M. Butler, STR sequence analysis for characterizing normal, variant, and null alleles, *Forensic Sci. Int. Genet.* 5 (2011) 329–332. doi:10.1016/j.fsigen.2010.09.005.
- [33] E.-M. Dauber, A. Kratzer, F. Neuhuber, W. Parson, M. Klintschar, W. Bär, W.R. Mayr, Germline mutations of STR-alleles include multi-step mutations as defined by sequencing of repeat and flanking regions, *Forensic Sci. Int. Genet.* 6 (2012) 381–386. doi:10.1016/j.fsigen.2011.07.015.
- [34] R.A.L. Griffiths, M.D. Barber, P.E. Johnson, S.M. Gilbard, M.D. Haywood, C.D. Smith, J. Arnold, T. Burke, A.J. Urquhart, P. Gill, New reference allelic ladders to improve allelic designation in a multiplex STR system, *Int. J. Legal Med.* 111 (1998) 267–272.
- [35] Illumina, MiSeq brochure, Version: BR\_1002(EN)\_V1\_14Mar2019. (2011).
- [36] Thermo Fisher Scientific, Ion Torrent next-generation sequencing technology, (n.d.). <https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/ion-torrent-next-generation-sequencing-technology.html> (accessed December 13, 2018).
- [37] V. Sharma, H.Y. Chow, D. Siegel, E. Wurmbach, Qualitative and quantitative assessment of Illumina’s forensic STR and SNP kits on MiSeq FGx™, *PLoS One.* (2017). doi:10.1371/journal.pone.0187932.
- [38] D. Deamer, M. Akeson, D. Branton, Three decades of nanopore sequencing, *Nat. Biotechnol.* 34 (2016) 518–524. doi:10.1038/nbt.3423.
- [39] M. Jain, H.E. Olsen, B. Paten, M. Akeson, The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community, *Genome Biol.* 17 (2016). doi:10.1186/s13059-016-1103-01.
- [40] Oxford Nanopore Technologies, Technology brochure, Version: BR\_1005(EN)\_V2\_25Mar2019. (2019).
- [41] Oxford Nanopore Technologies, How it works, (2012). <https://nanoporetech.com/how-it-works> (accessed December 2, 2018).
- [42] Oxford Nanopore Technologies, Nanopore Sensing-how it works, Version: NSTD\_5000\_v1\_revG\_04Apr2016. (2016).
- [43] F.J. Rang, W.P. Kloosterman, J. de Ridder, From squiggle to basepair: computational approaches for improving nanopore sequencing read accuracy, *Genome Biol.* 19 (2018) 90. doi:10.1186/s13059-018-1462-9.
- [44] Oxford Nanopore Technologies, DNA: nanopore sequencing, n.d. <https://nanoporetech.com/applications/dna-nanopore-sequencing>.
- [45] Oxford Nanopore Technologies, Chemistry Technical Document, Version: CHTD\_500\_v1\_revP\_07Jul2016. (2016).
- [46] F.J. Sedlazeck, P. Rescheneder, M. Smolka, H. Fang, M. Nattestad, A. Von Haeseler, M.C. Schatz, Accurate detection of complex structural variations using single-molecule sequencing, *Nat. Methods.* (2018). doi:10.1038/s41592-018-0001-7.
- [47] Oxford Nanopore Technologies, Flongle, Version: FLG\_5000\_V1\_REVA\_16OCT2018. (2018).

- [48] F.J. Sedlazeck, H. Lee, C.A. Darby, M.C. Schatz, Piercing the dark matter: bioinformatics of long-range sequencing and mapping, *Nat. Rev. Genet.* 19 (2018) 329–346. doi:10.1038/s41576-018-0003-4.
- [49] R.R. Zascavage, K. Thorson, J. V. Planz, Nanopore sequencing: An enrichment-free alternative to mitochondrial DNA sequencing, *Forensic Sci. Int. Genet.* 40 (2019) 272–280. doi:10.1002/elps.201800083.
- [50] S. Cornelis, S. Willems, C. Van Neste, O. Tytgat, J. Weymaere, A.-S. Vander Plaetsen, D. Deforce, F. Van Nieuwerburgh, Forensic STR profiling using Oxford Nanopore Technologies' MinION sequencer, *BioRxiv.* (2018). doi:10.1101/433151.
- [51] S. Cornelis, Y. Gansemans, A.S. Vander Plaetsen, J. Weymaere, S. Willems, D. Deforce, F. Van Nieuwerburgh, Forensic tri-allelic SNP genotyping using nanopore sequencing, *Forensic Sci. Int. Genet.* 38 (2019) 204–210. doi:10.1016/j.fsigen.2018.11.012.
- [52] Qiagen, QIAamp® DNA Mini and Blood Mini Handbook, Version: 1102728. 5 (2016).
- [53] Thermo Fisher Scientific, Quantifiler™ HP and Trio DNA Quantification Kits, Version: 4485354. (2017).
- [54] Thermo Fisher Scientific, Qubit® dsDNA BR Assay Kits, Version: MP32850. (2015).
- [55] Thermo Fisher Scientific, GlobalFiler™ PCR Amplification Kit User Guide, Version: 4477604. 4477604 (2016).
- [56] Thermo Fisher Scientific, Yfiler™ Plus PCR Amplification Kit User Guide, 4484678 (2016).
- [57] Thermo Fisher Scientific, GeneMapper® ID-X software v1.4, (2014).
- [58] K.J. van der Gaag, P. de Knijff, Forensic nomenclature for short tandem repeats updated for sequencing, *Forensic Sci. Int. Genet. Suppl. Ser.* (2015). doi:10.1016/j.fsigs.2015.09.214.
- [59] M. Gymrek, A.L. McGuire, D. Golan, E. Halperin, Y. Erlich, Identifying personal genomes by surname inference, *Science* (80-. ). (2013). doi:10.1126/science.1229566.
- [60] W.J. Kent, C.W. Sugnet, T.S. Furey, K.M. Roskin, T.H. Pringle, A.M. Zahler, a. D. Haussler, The Human Genome Browser at UCSC, *Genome Res.* 12 (2002) 996–1006. doi:10.1101/gr.229102.
- [61] J. Ye, G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen, T.L. Madden, Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction., *BMC Bioinformatics.* (2012). doi:10.1186/1471-2105-13-134.
- [62] A. Untergasser, H. Nijveen, X. Rao, T. Bisseling, R. Geurts, J.A.M. Leunissen, Primer3Plus, an enhanced web interface to Primer3, *Nucleic Acids Res.* (2007). doi:10.1093/nar/gkm306.
- [63] P.M. Vallone, J.M. Butler, AutoDimer: a screening tool for primer-dimer and hairpin structures, *Biotechniques.* 37 (2004) 226–231. doi:10.2144/04372ST03.
- [64] Takara Bio, TaKaRa LA PCR™ Kit v2.1 product manual, Version: 201701Da. (2017).
- [65] Agilent Technologies, Agilent D1000 ScreenTape Assay Quick Guide for 4200 TapeStation System, Version: G2991-90030. (2017).
- [66] Qiagen, QIAquick® Spin Handbook, Version: HB-1196-004. (2018).
- [67] Oxford Nanopore Technologies, 1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108), Version NBE\_9006\_v103\_revQ\_21Dec2016. (2018).
- [68] Millipore, User guider Microcon® centrifugal filter devices, 2018.
- [69] Oxford Nanopore Technologies, EPI2ME, Version: METD\_5000\_v1\_revAM\_29Feb2016. (2016).
- [70] H. Li, R. Durbin, Fast and accurate long-read alignment with Burrows–Wheeler transform, *Bioinformatics.* 26 (2010) 589–595. doi:10.1093/bioinformatics/btp698.

- [71] H. Lu, F. Giordano, Z. Ning, Oxford Nanopore MinION Sequencing and Genome Assembly, *Genomics, Proteomics Bioinforma.* (2016). doi:10.1016/j.gpb.2016.05.004.
- [72] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, The Sequence Alignment/Map format and SAMtools, *Bioinformatics.* (2009). doi:10.1093/bioinformatics/btp352.
- [73] I. Milne, M. Bayer, L. Cardle, P. Shaw, G. Stephen, F. Wright, D. Marshall, Tablet--next generation sequence assembly visualization, *Bioinformatics.* 26 (2010) 401–402. doi:10.1093/bioinformatics/btp666.
- [74] J.T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E.S. Lander, G. Getz, J.P. Mesirov, Integrative genomics viewer, *Nat. Biotechnol.* (2011). doi:10.1038/nbt.1754.
- [75] A.R. Quinlan, I.M. Hall, BEDTools: A flexible suite of utilities for comparing genomic features, *Bioinformatics.* (2010). doi:10.1093/bioinformatics/btq033.
- [76] A.E. Woerner, J.L. King, B. Budowle, Fast STR allele identification with STRait Razor 3.0, *Forensic Sci. Int. Genet.* (2017). doi:10.1016/j.fsigen.2017.05.008.
- [77] Takara Bio, FAQs : PCR Polymerases from Takara Bio, (2018). [https://www.takarabio.com/assets/documents/Frequently Asked Questions/PCRPolymerases\\_FAQ.pdf](https://www.takarabio.com/assets/documents/Frequently Asked Questions/PCRPolymerases_FAQ.pdf) (accessed December 10, 2018).
- [78] L.A. Clarke, PCR amplification introduces errors into mononucleotide and dinucleotide repeat sequences, *Mol. Pathol.* 54 (2001) 351–353. doi:10.1136/mp.54.5.351.
- [79] SWGDAM, Validation Guidelines for DNA Analysis Methods, (2016).
- [80] J.D. Cleary, L.P.W. Ranum, Repeat associated non-ATG (RAN) translation: new starts in microsatellite expansion disorders, *Curr. Opin. Genet. Dev.* 26 (2014) 6–15. doi:10.1016/j.gde.2014.03.002.
- [81] Q. Liu, P. Zhang, D. Wang, W. Gu, K. Wang, Interrogating the “unsequenceable” genomic trinucleotide repeat disorders by long-read sequencing, *Genome Med.* (2017). doi:10.1186/s13073-017-0456-7.
- [82] H. Tang, E.F. Kirkness, C. Lippert, W.H. Biggs, M. Fabani, E. Guzman, S. Ramakrishnan, V. Lavrenko, B. Kakaradov, C. Hou, B. Hicks, D. Heckerman, F.J. Och, C.T. Caskey, J.C. Venter, A. Telenti, Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes, *Am. J. Hum. Genet.* (2017). doi:10.1016/j.ajhg.2017.09.013.
- [83] Ł.J. Sznajder, J.D. Thomas, E.M. Carrell, T. Reid, K.N. McFarland, J.D. Cleary, R. Oliveira, C.A. Nutter, K. Bhatt, K. Sobczak, T. Ashizawa, C.A. Thornton, L.P.W. Ranum, M.S. Swanson, Intron retention induced by microsatellite expansions as a disease biomarker, *Proc. Natl. Acad. Sci.* 115 (2018) 4234–4239. doi:10.1073/pnas.1716617115.

APPENDIX A  
SAMPLE INFORMATION



**Table 5.** Sample information and quantification values for the twenty buccal swabs.

Assigned #	Biobank #	Sex	Population group	Target quantity (ng/ $\mu$ L)	
				Small autosomal	Y
1	80601	F	Caucasian	2.13840	-
2	80602	M	Caucasian	0.50149	0.55583
3	80670	F	Caucasian	3.54023	-
4	80779	F	Caucasian	2.74461	-
5	80780	M	Caucasian	5.17191	4.73328
6	81034	F	Caucasian	2.75087	-
7	81035	M	Caucasian	5.34180	4.25454
8	81083	F	African American	1.88590	-
9	81084	M	African American	6.82478	5.86549
10	81122	M	African American	2.18844	1.89377
11	81133	F	African American	1.15427	-
12	81134	M	African American	4.74295	4.13525
13	81139	F	African American	1.37530	-
14	81140	M	African American	5.26598	3.43180
15	80468	F	Hispanic	4.56372	-
16	80469	M	Hispanic	0.62719	0.37649
17	80511	F	Hispanic	4.48034	-
18	80512	M	Hispanic	4.08780	3.03162
19	80545	F	Hispanic	9.47398	-
20	80546	M	Hispanic	9.42883	9.47739

**Note:** Red and blue shading indicates samples used to test designed primer sets and optimize PCR parameters for autosomal and Y STR loci, respectively.

**Table 6.** Sample genotype data. Length-based genotypes for the twenty buccal swab samples at autosomal STR loci interrogated in this project and Amelogenin.

Locus	Sample									
	1	2	3	4	5	6	7	8	9	10
D1S1656	11, 15	15, 15.3	11, 16.3	16, 17.3	16, 17.3	14, 18.3	14, 16	13, 14	15, 16	15, 15.3
TPOX	8, 11	8, 11	8, 11	8, 10	8	11	8, 11	10, 11	9	8, 11
D2S441	11	11, 14	10, 11	12, 14	10	10, 12	13, 14	14, 15	14	11, 12
D2S1338	17, 24	22, 25	17, 25	24, 25	18, 20	17, 19	18, 19	17, 19	20, 23	19, 21
D3S1358	17	15, 18	16, 18	15, 16	16, 17	16, 18	16	9, 13	15	17, 18
FGA	20, 27	22	22, 23	21, 23	21, 23	21, 26	20, 22	22, 23	24, 25	22, 24
D5S818	11	12	11, 14	11, 14	10, 13	11	11, 12	8, 13	11, 12	11
CSF1PO	10, 11	10, 13	9, 10	12	9, 11	12, 15	13, 14	12, 13	8, 9	12
D7S820	11	7, 11	8, 11	12	8, 11	9	10, 12	9, 10	10, 11	11
D8S1179	12, 13	11, 15	14	12, 13	12, 14	10, 14	10, 13	12, 16	14, 15	15, 16
D10S1248	15	13, 15	12, 13	13	14, 15	13, 14	13, 15	15, 16	13, 15	12, 14
TH01	6	8, 9	6, 9.3	7, 9	7	8, 9.3	7, 9.3	7, 8	7	7
vWA	17	16, 19	14, 16	14, 15	16, 17	17, 18	16	15	18, 20	13, 16
D12S391	19, 20	18, 20	17, 20.3	18, 22	18, 21	17, 18	17	20	18, 22	17, 21
D13S317	12, 13	11, 12	9, 13	8, 12	8, 14	11	9, 12	11, 12	12	11, 12
D16S539	9, 12	13, 14	11, 12	11, 12	12	11, 12	11, 12	10, 13	9	9, 12
D18S51	14, 15	10, 17	12, 15	14, 15	12, 14	18, 19	13, 19	15, 16	13, 15	16
D19S433	14, 16	13, 14	14, 15	13, 14	13, 14	12	15, 16	12, 14.2	12, 13	14, 14.2
D21S11	30, 31	24.3, 28	30.2, 31.2	27, 28	27, 30	29, 31.2	29, 30	30, 32	30, 31.2	29, 30
D22S1045	11, 17	15, 16	16	16	16, 17	11, 16	15	16, 17	11	10, 16
AMEL	X	X, Y	X	X	X, Y	X	X, Y	X	X, Y	X, Y

**Note:** Penta E & Penta D are not queried by the GlobalFiler™ PCR Amplification Kit but were included in the panel of target loci.

**Table 6. (continued)** Sample genotype data. Length-based genotypes for the twenty buccal swab samples at autosomal STR loci interrogated in this project and Amelogenin.

Locus	Sample									
	11	12	13	14	15	16	17	18	19	20
D1S1656	11, 15	13, 18.3	14, 15	14, 16	14, 16	15.3, 16	17.3, 18.3	14, 16	14, 16	15.3, 17
TPOX	8	11	7, 11	6, 8	8, 11	8, 11	8	8	11	8, 12
D2S441	11, 14	14, 15	11, 14	11	10	11, 14	10, 11.3	10	10, 13	11, 14
D2S1338	20, 22	19, 22	19, 23	18, 19	22, 25	20, 25	18, 19	17, 21	19, 24	19, 20
D3S1358	14, 15	14, 16	15, 16	15, 16	15, 17	16	15	16	14, 15	18
FGA	20, 21	22, 26	24	20, 21	25	23, 26	19, 22	19, 25	19, 22	19, 24
D5S818	11, 12	12, 13	11	11, 14	11, 13	11, 13	12, 13	11, 14	10, 12	11, 12
CSF1PO	10, 12	11, 12	8, 10	8, 11	12, 13	12	10, 12	11	10, 12	10
D7S820	8, 10	8, 12	11	8, 9	10, 11	8, 11	11	8, 11	8, 9	8, 11
D8S1179	14, 16	12, 15	14	12, 14	14	10, 14	14	14, 15	13	11, 13
D10S1248	14, 16	13	13	13, 14	13	14, 15	15, 16	14, 15	13, 14	13
TH01	7, 9	7	7, 9	6, 7	6, 9.3	9.3	6, 7	6, 9.3	6, 7	6, 7
vWA	14, 18	13, 16	15, 17	16, 20	16, 17	16, 19	14, 19	18	14, 16	16, 17
D12S391	17, 20	19, 22	19, 23	16, 18	18, 19	17, 18	18, 23	19.3, 21	19	20
D13S317	12	8, 11	12	11	12, 13	10, 12	8, 12	12, 13	9, 12	9, 12
D16S539	13	9, 12	11, 13	9, 12	11, 12	11, 12	11, 12	10	11, 13	12, 13
D18S51	14, 17	17, 18	14, 16	17, 19	13, 14	13, 16	16, 18	14, 17	12, 13	13, 15
D19S433	10, 15	14, 15.2	11, 13	13, 13.2	15, 16	12.2, 14	13, 16.2	13	14, 15	15, 15.2
D21S11	29, 32.2	29	28, 36	30, 32.2	28, 29	30, 33.2	29, 30	29	31.2	30
D22S1045	16	15, 17	15, 17	11, 17	15, 16	12, 15	11	16	15, 16	15, 16
AMEL	X	X, Y	X	X, Y	X	X, Y	X	X, Y	X	X, Y

**Note:** Penta E & Penta D are not queried by the GlobalFiler™ PCR Amplification Kit but were included in the panel of target loci.

**Table 7.** Sample haplotype data. Length-based haplotypes for the ten male buccal swab samples at Y-STR loci interrogated in this project.

Locus	Sample									
	2	5	7	9	10	12	14	16	18	20
DYS19	15	14	14	14	15	16	14	14	13	14
DYS385a/b	11	12, 14	11, 15	12, 14	16	15, 19	14, 16	13, 18	15, 18	11, 14
DYS389I	13	13	13	14	13	13	13	12	12	13
DYS389II	29	29	29	30	31	31	30	29	28	29
DYS390	24	25	23	24	21	20	24	23	24	24
DYS391	10	10	11	10	10	10	9	11	9	11
DYS392	13	13	13	13	12	11	14	11	13	13
DYS393	13	13	13	13	13	14	12	13	13	13
DYS437	15	15	15	15	14	14	14	14	15	15
DYS438	12	12	12	12	11	11	11	10	11	12
DYS439	11	12	12	13	12	12	13	12	11	12
DYS448	19	20	19	19	22	21	21	19	20	19
DYS456	15	16	15	16	15	15	17	15	15	15
DYS458	18	17	17	18	18	17	15.1	19.2	19	18
DYS481	22	22	22	24	26	28	27	25	25	22
DYS533	12	11	12	12	11	11	11	11	11	12
DYS570	18	16	17	17	18	17	17	19	18	17
DYS576	17	17	18	18	17	14	18	21	20	19
GATAH4	12	12	12	12	11	12	12	11	12	12

**Note:** DYS435, DYS549 & DYS643 are not queried by the YFiler™ PCR Amplification Kit but were included in the panel of target loci.

**Table 8.** Control genotype/haplotype data. Length-based genotypes for the twenty buccal swab samples at autosomal STR loci interrogated in this project and Amelogenin.

Autosomal					Y			
Locus	007	A	B	C	Locus	007	A	B
D1S1656	11, 15	17.3, 17.3	11, 14	11, 15	DYS19	15	14	15
TPOX	8	8, 8	8, 11	11, 11	DYS385a	11	13	13
D2S441	11, 14	10, 10	10, 14	10, 10	DYS385b	141	17	15
D2S1338	20, 22	18, 23	17, 17	19, 19	DYS389I	13	13	12
D3S1358	14, 15	15, 16	15, 19	16, 18	DYS389II	29	31	27
FGA	20, 21	21, 23	20, 23	24, 26	DYS390	24	23	24
D5S818	11, 12	11, 12	12, 13	10, 11	DYS391	11	10	11
CSF1PO	10, 12	10, 10	10, 11	10, 12	DYS392	13	11	13
D7S820	8, 10	11, 11	10, 10	10, 12	DYS393	13	12	13
D8S1179	14, 16	13, 14	10, 13	10, 17	DYS437	15	14	16
D10S1248	14, 16	15, 16	13, 13	12, 16	DYS438	12	10	11
TH01	7, 9	8, 9.3	6, 9.3	6, 8	DYS439	12	11	12
vWA	14, 18	18, 19	17, 18	16, 18	DYS448	19	20	19
D12S391	17, 20	18.3, 22	19, 24	19, 23	DYS456	15	15	26
D13S317	12	8, 8	9, 12	11, 11	DYS458	17	17.2	15
Penta E	–	5, 10	7, 15	12, 13	DYS481	13	25	17
D16S539	13	10, 11	10, 13	10, 10	DYS533	13	11	10
D18S51	14, 17	12, 15	13, 16	16, 19	DYS549	–	12	13
D19S443	10, 15	13, 14	16, 16.2	13.2, 15.2	DYS570	17	18	20
D21S11	29, 32.2	28, 32.3	32, 32.2	29, 30	DYS576	19	17	16
Penta D	–	9, 13	8, 12	10, 11	DYS643	–	9	12
D22S1045	16	15, 15	15, 17	16, 16	GATAH4	13	11	11
AMEL	X, Y							

**Note:** Penta E & Penta D are not queried by the GlobalFiler™ PCR Amplification Kit but were included in the panel of target loci. DYS435, DYS549 & DYS643 are not queried by the YFiler™ PCR Amplification Kit but were included in the panel of target loci.

APPENDIX B

PRIMER DESIGN

**Table 9.** Autosomal loci genomic information. Map position, motif structure, and reference assembly allele designation for 22 autosomal STR loci of forensic interest and Amelogenin.

Locus	Map position (hg19)	Motif structure	Reference allele (hg19)
D1S1656	chr1: 230905351-230905426	[TAGA] <sub>4</sub> [TAG] <sub>0-1</sub> [TAGA] <sub>0-n</sub> [TAGG] <sub>0-1</sub> [TG] <sub>5</sub>	16
TPOX	chr2: 1493425-1493456	[AATG] <sub>n</sub>	8
D2S441	chr2: 68239079-68239126	[TCTA] <sub>4</sub> [TCA] <sub>0-1</sub> [TCTA] <sub>n</sub> [TTTA] <sub>0-1</sub> [TCTA] <sub>2</sub>	12
D2S1338	chr2: 218879582-218879673	[TGCC] <sub>n</sub> [TTCC] <sub>n</sub> [GTCC] <sub>0-1</sub> [TTCC] <sub>2</sub>	23
D3S1358	chr3: 45582231-45582294	TCTA [TCTG] <sub>1-4</sub> [TCTA] <sub>n</sub>	16
FGA	chr4: 155508888-155508975	[TTTC] <sub>4</sub> TTTTCT [CTTT] <sub>n</sub> [CTCG] <sub>0-5</sub> CTCC [TTCC] <sub>4</sub>	22
D5S818	chr5: 123111250-123111293	[AGAT] <sub>n</sub>	11
CSF1PO	chr5: 149455887-149455938	[AGAT] <sub>n</sub>	13
D7S820	chr7: 83789542-83789593	[GATA] <sub>n</sub>	13
D8S1179	chr8: 125907115-125907158	[TCTA] <sub>1-2</sub> [TCTG] <sub>1-2</sub> [TCTA] <sub>n</sub>	13
D10S1248	chr10: 131092508-131092559	[GGAA] <sub>n</sub>	13
TH01	chr11: 2192318-2192345	[AATG] <sub>3-5</sub> [ATG] <sub>0-1</sub> [AATG] <sub>n</sub>	7
vWA	chr12: 6093125-6093208	[TCTA] <sub>1-2</sub> [TCTG] <sub>1-6</sub> [TCTA] <sub>n</sub>	17
D12S391	chr12: 12449953-12450028	[AGAT] <sub>n</sub> [GAT] <sub>0-1</sub> [AGAC] <sub>n</sub> [AGAT] <sub>0-1</sub>	18
D13S317	chr13: 82722160-82722203	[TATC] <sub>n</sub>	11
Penta E	chr15: 97374245-97374269	[AAAGA] <sub>n</sub>	5
D16S539	chr16: 86386308-86386351	[GATA] <sub>n</sub>	11
D18S51	chr18: 60948900-60948971	[AGAA] <sub>n</sub>	18
D19S443	chr19: 30417141-30417204	[AAGG] AAAG [AAGG] TAGG [AAGG] <sub>n</sub>	14
D21S11	chr21: 20554291-20554417	[TCTA] <sub>n</sub> [TCTG] <sub>n</sub> [TCTA] <sub>3</sub> TA [TCTA] <sub>3</sub> TCA [TCTA] <sub>2</sub> TCCATA [TCTA] <sub>n</sub> TATCTA	29
Penta D	chr21: 45056086-45056150	[AAAGA] <sub>n</sub>	13
D22S1045	chr22: 37536327-37536377	[ATT] <sub>n</sub> ACT [ATT] <sub>2</sub>	17
Amelogenin	chrX: 11316133-11316931 chrY: 6736238-6737037	–	X, Y

**Table 10.** Y loci genomic information. Map position, motif structure, and reference assembly allele designation for 22 Y-STR loci of forensic interest.

Locus	Map position (hg 19)	Motif structure	Reference allele (hg19)
DYS19	chrY: 9521989-9522052	[TAGA] <sub>3</sub> TAGG [TAGA] <sub>n</sub>	15
DYS385a/b	chrY: 20842518-20842573 chrY: 20801568-20801824	[GAAA] <sub>n</sub>	14
DYS389I	chrY: 14612243-14612289	[TCTG] <sub>q</sub> [TCTA] <sub>r</sub>	12
DYS389II	chrY: 14612336-14612636	[TCTG] <sub>n</sub> [TCTA] <sub>p</sub> [TCTG] <sub>q</sub> [TCTA] <sub>r</sub>	29
DYS390	chrY: 17274947-17275042	[TCTG] <sub>n</sub> [TCTA] <sub>m</sub> [TCTG] <sub>p</sub> [TCTA] <sub>q</sub>	34
DYS391	chrY: 14102795-14102838	[TCTA] <sub>n</sub>	11
DYS392	chrY: 22633873-22633911	[TAT] <sub>n</sub>	13
DYS393	chrY: 3131152-3131199	[AGAT] <sub>n</sub>	12
DYS435	chrY: 14496298-14496333	[TGGA] <sub>n</sub>	9
DYS437	chrY: 14466994-14467057	[TCTA] <sub>n</sub> [TCTG] <sub>2</sub> [TCTA] <sub>4</sub>	16
DYS438	chrY: 14937824-14937873	[TTTTTC] <sub>n</sub>	10
DYS439	chrY: 14515312-14515363	[GATA] <sub>n</sub>	13
DYS448	chrY: 24365070-24365225	[AGAGAT] <sub>n</sub> N <sub>42</sub> [AGAGAT] <sub>n</sub>	19
DYS456	chrY: 4270960-4271019	[AGAT] <sub>n</sub>	15
DYS458	chrY: 7867880-7867943	[GAAA] <sub>n</sub>	16
DYS481	chrY: 8426378-8426443	[CTT] <sub>n</sub>	22
DYS533	chrY: 18393226-18393273	[ATCT] <sub>n</sub>	12
DYS549	chrY: 21520224-21520275	[GATA] <sub>n</sub>	13
DYS570	chrY: 6861231-6861298	[TTTC] <sub>n</sub>	17
DYS576	chrY: 7053359-7053426	[AAAG] <sub>n</sub>	16
DYS643	chrY: 17426012-17426066	[CTTTT] <sub>n</sub>	11
GATAH4	chrY: 18743553-18743600	[TAGA] <sub>n</sub> ATGGATAGATTA [GATG] <sub>p</sub> AA [TAGA] <sub>q</sub>	12



**Table 11.** Custom autosomal primer information. Designed primer sequences and amplicon lengths for 22 autosomal STRs and Amelogenin.

Locus	Orientation	Primer sequence (5' to 3')	Amplicon length
D1S1656	Forward	TCATGCCTACAGTGTAACGGG	797
	Reverse	CTAAAGCTGAACTCCGAATAGCC	
TPOX	Forward	CATGCCTGCACACACCTAC	800
	Reverse	CGCTCAAACGTGAGGTTGAC	
D2S441	Forward	GGGTTTGATTAATTTGCCAGCG	885
	Reverse	AGAGATAAGGTGTGCGTTACCC	
D2S1338	Forward	ACAACAGAATATGGGTTCTTGCG	801
	Reverse	CTGTGCTATGGAAAAGCCGTG	
D3S1358	Forward	GGTTTTGGTGGAAATGACTCCC	796
	Reverse	TCCCAGAGTGCTTCGTGC	
FGA	Forward	TCGTTTCATATCAACCAACTGAGC	795
	Reverse	TAGGAAACATTGCATTCACTCTGG	
D5S818	Forward	TTGGAAATGTTATCTACAACGCCTG	799
	Reverse	TCTAACTTTGAACTACACAACACGC	
CSF1PO	Forward	GCCACCTCCATCTCCCATAAC	799
	Reverse	GAGGAACATATGCAAGGCTCAAAG	
D7S820	Forward	AGGACTGGAAAGATCCAATTTTGC	829
	Reverse	CGCACCTGACCCCCTATG	
D8S1179	Forward	ATCCAAAGTGGATCCTCAGCC	794
	Reverse	CCGGCCCAAGTTTTTATTTTCC	
D10S1248	Forward	TGACAGTTCTGTTTTGCGGTG	802
	Reverse	AGTTATCTTATCAGTGGTCCAGGTC	
TH01	Forward	TCAGCTTCATCCTGAGCTTCC	806
	Reverse	TTGAATCTTAACGATCGGAATGTGG	
vWA	Forward	GACGTCCATGCAGAGTTCAG	889
	Reverse	AAGGTAGAGTTCCCACCTTCC	
D12S391	Forward	GAATATATGAAGTCGCCTGTAATCCC	795
	Reverse	CTCCTGGAGCTGCGACAC	
D13S317	Forward	ACTCCAAGCTCACAGTGCC	802
	Reverse	CCGAGGTTCTCTTCTGTGTC	
Penta E	Forward	CTCAATGCCAACATTACAGGGC	801
	Reverse	GCTTAAAGTTGACGTCTCATTGC	
D16S539	Forward	CTCAGTCCTGCCGAGGTG	812
	Reverse	CCAGATCAATAGGGCTGGGC	
D18S51	Forward	CTAACAATAGGCCAAGCGTGATG	813
	Reverse	ATTGAGTCAGGTAACATTTATGCCC	
D19S443	Forward	CATGAAACTGGACACAGAAACCAG	804
	Reverse	AGAAAAGTGCACAAATCCTTAGCG	
D21S11	Forward	CAAATTTCCCCTCTCACTTCTGG	810
	Reverse	AGAATTAGAGAGCTGTGTCGAGG	
Penta D	Forward	CAGTGGCTAATTGTACCTTGGG	684
	Reverse	CAAAGTGCTGGGATTACCATCG	
D22S1045	Forward	GAGCAGCTTAAGGCATCCTG	820
	Reverse	GCTTCTCACCGTTGCATTAGG	
Amelogenin	Forward	AGGCTGTGGCTGAACAGG	798/801
	Reverse	AGATGAAGAATGTGTGTGATGGATG	

**Table 12.** Custom Y primer information. Designed primer sequences and amplicon lengths for 22 Y-STRs.

Locus	Orientation	Primer sequence (5' to 3')	Amplicon length
DYS19	Forward Reverse	TGTCATAGCTCTGAACATTTGGTTG AGTTTCATGAAGCATTCTCTACAGT	825
DYS385a/b	Forward Reverse	CAAGTTCTGTGTGCATCATTTTCTC TTGAACCTGAAATGTAAAGGGTGTC	791/803
DYS389I/II	Forward Reverse	ACCAGATGCGAGAACACCC CCCCAGATACAGATGCTGCC	796
DYS390	Forward Reverse	ACAAATAAAACCTACCTGCAGCC TACTGGAACACCTGTTGGAGAG	806
DYS392	Forward Reverse	CTCCCTGTGAGAATGAATTGTTCC ATTTGGGGTAGTGCTTAGTCCC	800
DYS391	Forward Reverse	ACCACAGATTAGCATTCTCTTGC ACCAGATGCCAGCAGTATCC	844
DYS393	Forward Reverse	CCCCTACACAAGTTCTCCTGC GTGGTGGTGCATGCTGTAA	799
DYS435	Forward Reverse	TGCAGTGCTGCATAATAAATGACC CAGCCCACCAGATATCTCTATCTAC	807
DYS437	Forward Reverse	AATCCATCCATTTCATCTCCCT CTGAGGTTGGAATTGCTTGAGTC	838
DYS438	Forward Reverse	ACCATTTTCTGATGAAAAGGAACCAG TACCTGGCTGCTCCCCTAG	802
DYS439	Forward Reverse	GTACTIONCTAGGTTTTCTTCTCGAGT TCTGGTGTCTCTTCCAACCTTG	839
DYS448	Forward Reverse	GGAGGATATGTCAAAGGATTCAAGG TTCTTCCCTTCTCTGACATTTTGG	796
DYS456	Forward Reverse	TTGCTTCCACAAGATTTACACTG TGGGGTCTCATTATGTTGTTTATGC	795
DYS458	Forward Reverse	CTTTAACAGTTTGGGACAGCTGAG CTTAAAAAGTTTCCCCATGTTGGTG	816
DYS481	Forward Reverse	AAGTCTGGCATCCAAACCTGC TCTCCTTGGTTTGGCCATAGG	904
DYS533	Forward Reverse	TCTCAAACCCCTACTCATTCAAAAC AGCTTTTGGAAACACCAGTTAGAG	802
DYS549	Forward Reverse	CAATGAACCCCAACACAAGC TGTGGATTAGTTTTTGACGTGCTAC	808
DYS570	Forward Reverse	TCCTTGGAATTGCAACTTGGC TCCATCTCAGAATCAAGAAGGGC	845
DYS576	Forward Reverse	AAAGGTGAAAATGTGCCTTCCC TACCTTTGGGTGTGTATGTTAGTCC	795
DYS643	Forward Reverse	CCAACTGAGTGGATATTTCTTGC ACGTTGCCTGTAACTCTGATAATAC	827
GATAH4	Forward Reverse	CTATGCCCAGAATATTGCATTAGGC AATGTCAAGCCTTCTGATTGCC	841

APPENDIX C

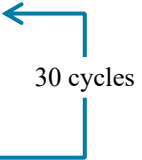
PCR AMPLIFICATION

**Table 13.** PCR components. PCR amplification half-reaction mixture (25µL).

Reaction component	Input
TaKaRa LA Taq (5 units/µL)	0.25 µL
10X PCR Buffer II (Mg <sup>2+</sup> )	2.5 µL
dNTP Mixture (2.5 mM each)	4.0 µL
Template DNA	0.5 – 1.0 ng
Primers	0.5 µM*
Sterile distilled water	up to 25 µL

**Table 14.** PCR parameters. PCR thermocycling conditions used to amplify target loci.

Step	Temperature	Time
Initial Denaturation	94°C	1 min
Denaturation	98°C	10 sec
Primer annealing	57 – 59°C	30 sec
Extension	65°C	1 min 15 sec
Final Extension	72°C	7 min
Hold	4°C	∞



APPENDIX D  
NANOPORE SEQUENCING

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

Before start checklist		
Materials	Consumables	Equipment
<input type="checkbox"/> Native Barcoding Expansion 1-12 (EXP-NBD103)	<input type="checkbox"/> NEBNext End repair / dA-tailing Module (E7546)	<input type="checkbox"/> Thermal cycler at 20° C and 65° C
<input type="checkbox"/> Ligation Sequencing Kit 1D (SQK-LSK108)	<input type="checkbox"/> NEB Blunt/TA Ligase Master Mix (M0367)	<input type="checkbox"/> Microfuge
<input type="checkbox"/> Library Loading Bead Kit (EXP-LLB001)	<input type="checkbox"/> Agencourt AMPure XP beads	<input type="checkbox"/> Vortex mixer
	<input type="checkbox"/> Freshly prepared 70% ethanol in nuclease-free water	<input type="checkbox"/> Magnetic rack
	<input type="checkbox"/> 1.5 ml Eppendorf DNA Lo-Bind tubes	<input type="checkbox"/> Heating block at 37° C capable of taking 1.5 ml tubes
	<input type="checkbox"/> 0.2 ml thin-walled PCR tubes	<input type="checkbox"/> Pipettes P2, P10, P20, P100, P200, P1000
	<input type="checkbox"/> Nuclease-free water (e.g. ThermoFisher, cat # AM9937)	
	<input type="checkbox"/> NEBNext Quick Ligation Module (E6056)	
	<input type="checkbox"/> Pipette tips P2, P10, P20, P100, P200, P1000	

INSTRUCTIONS	NOTES/OBSERVATIONS
<b>Preparing input DNA</b>	
<input type="checkbox"/> Record the quality, quantity and size of the DNA.	
<b>IMPORTANT</b> Criteria for input DNA <ul style="list-style-type: none"> <li><input type="checkbox"/> Purity as measured using Nanodrop - OD 260/280 of 1.8 and OD 260/230 of 2.0-2.2</li> <li><input type="checkbox"/> Average fragment size, as measured by pulse-field, or low percentage agarose gel analysis &gt;30 kb</li> <li><input type="checkbox"/> Input mass, as measured by Qubit - 1 µg (~ 1.5 µg if carrying out a DNA repair step)</li> <li><input type="checkbox"/> No detergents or surfactants in the buffer</li> </ul>	
Prepare the DNA in Nuclease-free water. <ul style="list-style-type: none"> <li><input type="checkbox"/> Transfer 1-1.5 µg genomic DNA into a DNA LoBind tube</li> <li><input type="checkbox"/> Adjust the volume to 46 µl with Nuclease-free water</li> <li><input type="checkbox"/> Mix thoroughly by inversion avoiding unwanted shearing</li> <li><input type="checkbox"/> Spin down briefly in a microfuge</li> </ul>	

**Figure 13.** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><b>Check your flow cell</b></p> <p><input type="checkbox"/> Set up the MinION, flow cell and host computer</p> <p>Once successfully plugged in, you will see a light and hear the fan.</p> <p>Open the MinkNOW GUI from the desktop icon and establish a local or remote connection.</p> <p><input type="checkbox"/> If running a MinION on the same host computer, plug the MinION into the computer.</p> <p><input type="checkbox"/> If running a MinION on a remote computer, first enter the name or IP address of the remote host under Connect to a remote computer (if running from the Connection page), or Connections (if running from the homepage) and click Connect.</p> <p><input type="checkbox"/> Choose the flow cell type from the selector box. Then mark the flow cell as "Selected":</p> <p>Click "Check flow cells" at the bottom of the screen.</p> <p><input type="checkbox"/> R9.4.1 FLO-MIN106</p> <p><input type="checkbox"/> R9.5.1 FLO-MIN107</p> <p><input type="checkbox"/> Click "Start test".</p> <p><input type="checkbox"/> Check the number of active pores available for the experiment, reported in the System History panel when the check is complete.</p>	
Flow cell check complete.	
<p><b>DNA fragmentation</b></p> <p><b>OPTIONAL</b></p> <p><input type="checkbox"/> Transfer each sample of &lt;1 µg genomic DNA in 46 µl to a Covaris g-TUBE.</p> <p>Spin the g-TUBE for 1 minute at RT at the speed for the fragment size required.</p> <p><input type="checkbox"/> Spin the g-TUBE for 1 minute</p> <p><input type="checkbox"/> Remove and check all the DNA has passed through the g-TUBE</p> <p><input type="checkbox"/> If DNA remains in the upper chamber, spin again for 1 minute at the same speed</p> <p>Invert the g-TUBE and spin again for 1 minute to collect the fragmented DNA.</p> <p><input type="checkbox"/> Remove g-TUBE, invert the tube and replace into the centrifuge</p> <p><input type="checkbox"/> Spin the g-TUBE for 1 minute</p> <p><input type="checkbox"/> Remove and check the DNA has passed into the lower chamber</p> <p><input type="checkbox"/> If DNA remains in the upper chamber, spin again for 1 minute</p> <p><input type="checkbox"/> Remove g-TUBE</p> <p><input type="checkbox"/> Transfer the 46 µl fragmented DNA to a clean 1.5 ml Eppendorf DNA LoBind tube.</p>	
Analyse 1 µl of the fragmented DNA for fragment size, quantity and quality.	
<1 µg fragmented DNA in 45 µl from each sample is taken into the next step.	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><b>DNA repair (optional)</b></p> <p><b>OPTIONAL</b></p> <p>Perform FFPE DNA repair treatment using NEB M6630.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 45 µl 1-1.5 µg fragmented** DNA</li> <li><input type="checkbox"/> 8.5 µl Nuclease-free water</li> <li><input type="checkbox"/> 6.5 µl FFPE Repair Buffer</li> <li><input type="checkbox"/> 2 µl FFPE Repair Mix</li> </ul> <p><input type="checkbox"/> Mix by pipetting and spin down.</p> <p><input type="checkbox"/> Incubate the reaction for 15 minutes at 20° C.</p> <p><input type="checkbox"/> Prepare the AMPure XP beads for use; resuspend by vortexing.</p> <p><input type="checkbox"/> Add 62 µl of the resuspended beads to the End-prep reaction and mix gently by pipetting.</p> <p><input type="checkbox"/> Incubate on a Hula mixer (rotator mixer) for 5 minutes at RT.</p> <p><input type="checkbox"/> Prepare 500 µl of fresh 70% ethanol in Nuclease-free water.</p> <p><input type="checkbox"/> Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette off the supernatant.</p> <p><input type="checkbox"/> Keep on magnet, wash beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.</p> <p><input type="checkbox"/> Repeat the previous step.</p> <p><input type="checkbox"/> Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.</p> <p><input type="checkbox"/> Remove the tube from the magnetic rack and resuspend pellet in 46 µl Nuclease-free water. Incubate for 2 minutes at RT.</p> <p><input type="checkbox"/> Pellet the beads on a magnet until the eluate is clear and colourless.</p> <p><input type="checkbox"/> Remove and retain 46 µl of eluate in a clean 1.5 ml Eppendorf DNA LoBind tube.</p> <p><input type="checkbox"/> Quantify 1 µl of fragmented and repaired DNA using a Qubit fluorometer - recovery aim &gt; 1 µg.</p>	
<p>Take 1 µg of FFPE repaired DNA in 45 µl into End-prep.</p>	
<p><b>Library preparation</b></p> <p>Perform end-repair / dA-tailing of fragmented DNA as follows:</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 45 µl &lt;1 µg end-repaired DNA</li> <li><input type="checkbox"/> 7 µl Ultra II End-prep reaction buffer</li> <li><input type="checkbox"/> 3 µl Ultra II End-prep enzyme mix</li> <li><input type="checkbox"/> 5 µl Nuclease-free water</li> </ul>	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].



1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><input type="checkbox"/> Mix gently by flicking the tube, and spin down.</p> <p><input type="checkbox"/> Transfer the sample to a 0.2 ml PCR tube, and incubate for 5 minutes at 20 °C and 5 minutes at 65 °C using the thermal cycler.</p> <p><input type="checkbox"/> Prepare the AMPure XP beads for use; resuspend by vortexing.</p> <p><input type="checkbox"/> Add 60 µl of resuspended AMPure XP beads to the end-prep reaction and mix by pipetting.</p> <p><input type="checkbox"/> Incubate on a Hula mixer (rotator mixer) for 5 minutes at RT.</p> <p><input type="checkbox"/> Prepare 500 µl of fresh 70% ethanol in Nuclease-free water.</p> <p><input type="checkbox"/> Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette off the supernatant.</p> <p><input type="checkbox"/> Keep on magnet, wash beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.</p> <p><input type="checkbox"/> Repeat the previous step.</p> <p><input type="checkbox"/> Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.</p> <p><input type="checkbox"/> Remove the tube from the magnetic rack and resuspend pellet in 25 µl Nuclease-free water. Incubate for 2 minutes at RT.</p> <p><input type="checkbox"/> Pellet the beads on a magnet until the eluate is clear and colourless.</p> <p><input type="checkbox"/> Remove and retain 25 µl of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube.</p> <p><input type="checkbox"/> Quantify 1 µl of end-prepped DNA using a Qubit fluorometer - recovery aim &gt; 700 ng.</p> <p><input type="checkbox"/> Thaw the Native Barcodes at RT, enough for one barcode per sample. Mix the barcodes by pipetting, and place them on ice.</p> <p><input type="checkbox"/> Select a unique barcode for every sample to be run together on the same flow cell, from the provided 12 barcodes. Up to 12 samples can be barcoded and combined in one experiment.</p> <p><input type="checkbox"/> Dilute 500 ng of each end-prepped sample to be barcoded to 22.5 µl in Nuclease-free water.</p> <p>Add the reagents in the order given below, mixing by flicking the tube between each sequential addition:</p> <p><input type="checkbox"/> 22.5 µl 500 ng end-prepped DNA</p> <p><input type="checkbox"/> 2.5 µl Native Barcode</p> <p><input type="checkbox"/> 25 µl Blunt/TA Ligase Master Mix</p> <p><input type="checkbox"/> Mix gently by flicking the tube, and spin down.</p> <p><input type="checkbox"/> Incubate the reaction for 10 minutes at RT.</p> <p><input type="checkbox"/> Prepare the AMPure XP beads for use; resuspend by vortexing.</p> <p><input type="checkbox"/> Add 50 µl of resuspended AMPure XP beads to the reaction and mix by pipetting.</p>	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><input type="checkbox"/> Incubate on a Hula mixer (rotator mixer) for 5 minutes at RT.</p> <p><input type="checkbox"/> Prepare 500 µl of fresh 70% ethanol in Nuclease-free water.</p> <p><input type="checkbox"/> Spin down the sample and pellet on a magnet. Keep the tube on the magnet, and pipette off the supernatant.</p> <p><input type="checkbox"/> Keep on magnet, wash beads with 200 µl of freshly prepared 70% ethanol without disturbing the pellet. Remove the 70% ethanol using a pipette and discard.</p> <p><input type="checkbox"/> Repeat the previous step.</p> <p><input type="checkbox"/> Spin down and place the tube back on the magnet. Pipette off any residual ethanol. Allow to dry for ~30 seconds, but do not dry the pellet to the point of cracking.</p> <p><input type="checkbox"/> Remove the tube from the magnetic rack and resuspend pellet in 26 µl Nuclease-free water. Incubate for 2 minutes at RT.</p> <p><input type="checkbox"/> Pellet the beads on a magnet until the eluate is clear and colourless.</p> <p>Remove and retain 26 µl of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube.</p> <p><input type="checkbox"/> Remove and retain the eluate which contains the DNA library in a clean 1.5 ml Eppendorf DNA LoBind tube</p> <p><input type="checkbox"/> Dispose of the pelleted beads</p> <p><input type="checkbox"/> Quantify 1 µl of each barcoded DNA sample using a Qubit fluorometer.</p> <p><input type="checkbox"/> Pool equimolar amounts of each barcoded sample into a DNA LoBind 1.5 ml Eppendorf tube, ensuring that sufficient sample is combined to produce a pooled sample of 700 ng total.</p> <p><input type="checkbox"/> Quantify 1 µl of pooled and barcoded DNA using a Qubit fluorometer.</p> <p><input type="checkbox"/> Dilute 700 ng pooled sample to 50 µl in Nuclease-free water.</p> <p>Thaw and prepare the kit reagents as follows:</p> <p><input type="checkbox"/> ABB Buffer (ABB) at RT</p> <p><input type="checkbox"/> Elution Buffer (ELB) at RT</p> <p><input type="checkbox"/> Barcode Adapter Mix (BAM) on ice</p> <p><input type="checkbox"/> Running Buffer with Fuel Mix (RBF) on ice</p> <p><input type="checkbox"/> NEBNext Quick Ligation Reaction Buffer (5x) on ice</p> <p><b>IMPORTANT</b></p> <p><input type="checkbox"/> Thoroughly mix the contents of each tube by flicking (vortexing or pipetting for RBF), and spin down very briefly before pipetting to ensure the contents of the tube can be aspirated accurately. For the Running Buffer FM (RBF) this should be carried out prior to use. For the Library Loading Beads (LLB), mix the beads by pipetting prior to adding to the library, and mix the DNA library with beads by pipetting prior to loading into the Flow Cell.</p>	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS						
<p>Check the contents of each tube are clear of any precipitate and are thoroughly mixed before setting up the reaction.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Mix the contents of each tube by flicking</li> <li><input type="checkbox"/> Check that there is no precipitate present (DTT in the Blunt/TA Master Mix can sometimes form a precipitate)</li> <li><input type="checkbox"/> Spin down briefly before accurately pipetting the contents in the reaction</li> </ul> <p>Taking the pooled and barcoded DNA, perform adapter ligation as follows, mixing by flicking the tube between each sequential addition.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> 50 µl 700 ng pooled barcoded sample</li> <li><input type="checkbox"/> 20 µl Barcode Adapter Mix (BAM 1D)</li> <li><input type="checkbox"/> 20 µl NEBNext Quick Ligation Reaction Buffer (5X)</li> <li><input type="checkbox"/> 10 µl Quick T4 DNA Ligase</li> </ul> <ul style="list-style-type: none"> <li><input type="checkbox"/> Mix gently by flicking the tube, and spin down.</li> <li><input type="checkbox"/> Incubate the reaction for 10 minutes at RT.</li> <li><input type="checkbox"/> Prepare the AMPure XP beads for use; resuspend by vortexing.</li> <li><input type="checkbox"/> Add 40 µl of resuspended AMPure XP beads to the adapter ligation reaction from the previous step and mix by pipetting.</li> <li><input type="checkbox"/> Incubate on a Hula mixer (rotator mixer) for 5 minutes at RT.</li> <li><input type="checkbox"/> Place on magnetic rack, allow beads to pellet and pipette off supernatant.</li> <li><input type="checkbox"/> Add 140 µl of ABB Buffer (ABB) to the beads. Close the tube lid, and resuspend the beads by flicking the tube. Return the tube to the magnetic rack, allow beads to pellet and pipette off the supernatant.</li> <li><input type="checkbox"/> Repeat the previous step.</li> <li><input type="checkbox"/> Remove the tube from the magnetic rack and resuspend pellet in 15 µl Elution Buffer. Incubate for 10 minutes at RT.</li> <li><input type="checkbox"/> Pellet the beads on a magnet until the eluate is clear and colourless.</li> </ul> <p>Remove and retain 15 µl of eluate into a clean 1.5 ml Eppendorf DNA LoBind tube.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Remove and retain the eluate which contains the DNA library in a clean 1.5 ml Eppendorf DNA LoBind tube</li> <li><input type="checkbox"/> Dispose of the pelleted beads</li> </ul> <ul style="list-style-type: none"> <li><input type="checkbox"/> Quantify 1 µl of adapter ligated DNA using a Qubit fluorometer - recovery aim ~430 ng.</li> </ul>							
<p>The prepared library is used for loading into the flow cell. Store the library on ice until ready to load.</p>							
<p><b>Before sequencing checklist</b></p> <table border="0"> <tr> <td><input type="checkbox"/> Prepared library on ice</td> <td><input type="checkbox"/> Computer set up to run MinKNOW</td> <td><input type="checkbox"/> Hardware check complete</td> </tr> <tr> <td><input type="checkbox"/> Sequencing device connected to computer with SpotON Flow Cell inserted</td> <td><input type="checkbox"/> Desktop Agent set up (if applicable)</td> <td><input type="checkbox"/> Flow cell check complete</td> </tr> </table>		<input type="checkbox"/> Prepared library on ice	<input type="checkbox"/> Computer set up to run MinKNOW	<input type="checkbox"/> Hardware check complete	<input type="checkbox"/> Sequencing device connected to computer with SpotON Flow Cell inserted	<input type="checkbox"/> Desktop Agent set up (if applicable)	<input type="checkbox"/> Flow cell check complete
<input type="checkbox"/> Prepared library on ice	<input type="checkbox"/> Computer set up to run MinKNOW	<input type="checkbox"/> Hardware check complete					
<input type="checkbox"/> Sequencing device connected to computer with SpotON Flow Cell inserted	<input type="checkbox"/> Desktop Agent set up (if applicable)	<input type="checkbox"/> Flow cell check complete					

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A  $2.5 \times$  AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><b>Priming and loading the SpotON flow cell</b></p> <p><b>IMPORTANT</b></p> <p><input type="checkbox"/> Thoroughly mix the contents of the RBF tube by vortexing or pipetting, and spin down briefly.</p> <p><input type="checkbox"/> Flip back the MinION lid and slide the priming port cover clockwise so that the priming port is visible.</p> <p><b>IMPORTANT</b></p> <p><input type="checkbox"/> Care must be taken when drawing back buffer from the flow cell. The array of pores must be covered by buffer at all times. Removing more than 20-30 µl risks damaging the pores in the array.</p> <p>After opening the priming port, check for small bubble under the cover. Draw back a small volume to remove any bubble (a few µls):</p> <p><input type="checkbox"/> Set a P1000 pipette to 200 µl</p> <p><input type="checkbox"/> Insert the tip into the priming port</p> <p><input type="checkbox"/> Turn the wheel until the dial shows 220-230 µl, or until you can see a small volume of buffer entering the pipette tip</p> <p>Prepare the flow cell priming mix in a clean 1.5 ml Eppendorf DNA LoBind tube.</p> <p><input type="checkbox"/> 576 µl RBF</p> <p><input type="checkbox"/> 624 µl Nuclease-free water</p> <p><input type="checkbox"/> Load 800 µl of the priming mix into the flow cell via the priming port, avoiding the introduction of air bubbles. Wait for 5 minutes.</p> <p><input type="checkbox"/> Thoroughly mix the contents of the RBF and LLB tubes by pipetting.</p> <p>Prepare the library for loading as follows:</p> <p><input type="checkbox"/> 35.0 µl RBF</p> <p><input type="checkbox"/> 25.5 µl LLB</p> <p><input type="checkbox"/> 2.5 µl Nuclease-free water</p> <p><input type="checkbox"/> 12 µl DNA library</p> <p>Complete the flow cell priming:</p> <p><input type="checkbox"/> Gently lift the SpotON sample port cover to make the SpotON sample port accessible.</p> <p><input type="checkbox"/> Load 200 µl of the priming mix into the flow cell via the priming port (not the SpotON sample port), avoiding the introduction of air bubbles.</p> <p><input type="checkbox"/> Mix the prepared library gently by pipetting up and down just prior to loading.</p> <p><input type="checkbox"/> Add 75 µl of sample to the flow cell via the SpotON sample port in a dropwise fashion. Ensure each drop flows into the port before adding the next.</p> <p><input type="checkbox"/> Gently replace the SpotON sample port cover, making sure the bung enters the SpotON port, close the priming port and replace the MinION lid.</p>	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<p><b>Starting a sequencing run</b></p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Double-click the MinKNOW icon located on the desktop to open the MinKNOW GUI.</li> <li><input type="checkbox"/> If your MinION was disconnected from the computer, plug it back in.</li> <li><input type="checkbox"/> Choose the flow cell type from the selector box. Then mark the flow cell as "Selected".</li> <li><input type="checkbox"/> Click the "New Experiment" button at the bottom left of the GUI.</li> </ul> <p>On the New experiment popup screen, select the running parameters for your experiment from the individual tabs.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> Output settings - FASTQ: The number of basecalls that MinKNOW will write in a single file. By default this is set to 4000</li> <li><input type="checkbox"/> Output settings - FAST5: The number of files that MinKNOW will write to a single folder. By default this is set to 4000</li> <li><input type="checkbox"/> Click "Begin Experiment".</li> </ul> <p>Allow the script to run to completion.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> The MinKNOW Experiment page will indicate the progression of the script; this can be accessed through the "Experiment" tab that will appear at the top right of the screen</li> <li><input type="checkbox"/> Monitor messages in the Message panel in the MinKNOW GUI</li> </ul> <p>The basecalled read files are stored in :data\reads</p>	
<p><b>Progression of MinKNOW protocol script</b></p> <p>The running experiment screen</p> <p>Experiment summary information</p> <p>Check the number of active pores reported in the MUX scan are similar (within 10-15%) to those reported at the end of the Platform QC</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> If there is a significant reduction in the numbers, restart MinKNOW.</li> <li><input type="checkbox"/> If the numbers are still significantly different, close down the host computer and reboot.</li> <li><input type="checkbox"/> When the numbers are similar to those reported at the end of the Platform QC, restart the experiment on the Connection page. There is no need to load any additional library after restart.</li> <li><input type="checkbox"/> Stopping the experiment is achieved by clicking "Stop experiment" button at the top of the screen.</li> </ul> <p><input type="checkbox"/> Check the temperature is approximately 34° C.</p> <p>Check pore occupancy in the channel panel at the top of the experimental view.</p> <ul style="list-style-type: none"> <li><input type="checkbox"/> A good library will be indicated by a higher proportion of light green channels in Sequencing than are in Pore. The combination of Sequencing and Pore indicates the number of active pores at any point in time. A low proportion of Sequencing channels will reduce the throughput of the run.</li> <li><input type="checkbox"/> Recovering indicates channels that may become available for sequencing again. A high proportion of this may indicate additional clean up steps are required during your library preparation.</li> <li><input type="checkbox"/> Inactive indicates channels that are no longer available for sequencing. A high proportion of these as soon as the run begins may indicate an osmotic imbalance.</li> <li><input type="checkbox"/> Unclassified are channels that have not yet been assigned one of the above classifications</li> </ul>	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

1D Native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108)

Version: NBE\_9006\_v103\_revQ\_21Dec2016  
Last update: 10/05/2018



Flow Cell Number: .....

DNA Samples: .....

INSTRUCTIONS	NOTES/OBSERVATIONS
<input type="checkbox"/> Monitor the pore occupancy  Duty time plots  <input type="checkbox"/> Monitor the development of the read length histogram.  Trace viewer	
<b>Onward analysis of MinKNOW basecalled data</b>	
<input type="checkbox"/> Open the Desktop Agent using the desktop shortcut.  <input type="checkbox"/> Click on the New Workflow tab in the Desktop Agent and select the FASTQ barcoding workflow.  Select the workflow parameters. <input type="checkbox"/> Select the quality score cut-off (this defaults to 7 unless changed) <input type="checkbox"/> Select "Yes" in answer to "Detect barcode?" <input type="checkbox"/> If you are working with human data, please tick "Yes" in answer to "Is the data you are about to upload a whole or partial human genome?", and confirm that you have consent from the subject to upload the data.  <input type="checkbox"/> Check the correct settings are selected in the Desktop Agent.  <input type="checkbox"/> Click "Start Run" to start data analysis.  <input type="checkbox"/> Follow the progression of upload and download of read files in the Desktop Agent.  Click on VIEW REPORT. <input type="checkbox"/> Click on VIEW REPORT to navigate to the Metrichor website, this can be done at any point during data exchange <input type="checkbox"/> Return to the Desktop Agent to see progression of the exchange	
<b>Close down MinKNOW and the Desktop Agent</b>	
<input type="checkbox"/> Quit Desktop Agent using the close x.  <input type="checkbox"/> Quit MinKNOW by closing down the web GUI.  <input type="checkbox"/> Disconnect the MinION.	
<b>Prepare the flow cell for re-use or return to Oxford Nanopore.</b>	
<input type="checkbox"/> If you would like to reuse the flow cell, follow the Wash Kit instructions and store the washed flow cell at 2-8 °C, OR  <input type="checkbox"/> Follow the returns procedure by washing out the MinION Flow Cell ready to send back to Oxford Nanopore.	

**Figure 13. (continued)** 1D native barcoding genomic DNA (with EXP-NBD103 and SQK-LSK108). Optional DNA repair and fragmentation steps were bypassed because amplicons were undamaged and of optimal length. A 2.5 × AMPure® XP bead purification was performed after pooling 700 ng of the barcoded samples [67].

## APPENDIX E

### READ COUNT DATA

**Table 15.** Autosomal STR sample read counts. Number of reads aligning to each autosomal STR locus for the twenty buccal swab samples.

Locus	Sample									
	1	2	3	4	5	6	7	8	9	10
D1S1656	8078	6880	5956	4344	5326	4901	5916	3354	2772	5292
TPOX	7091	7405	3557	3672	4906	4396	4113	2317	5641	4296
D2S441	5440	1755	1714	3663	5353	1119	4237	1369	529	748
D2S1338	5743	4501	3152	2987	4076	3604	3124	2021	1822	3376
D3S1358	4935	5195	2860	3027	4977	2435	3967	1555	4466	3195
FGA	2477	2387	1706	1468	2125	1314	1559	653	710	1765
D5S818	4394	3574	2103	2001	3504	2019	2303	1364	1609	2196
CSF1PO	5184	4158	2218	2013	3238	2031	1727	1136	2212	2599
D7S820	7506	7824	3419	3639	5706	3606	2989	1843	2326	4157
D8S1179	3355	2564	2243	1951	2773	2243	1813	1412	2596	2569
D10S1248	11591	7890	7768	8339	11259	8439	7270	5359	4468	8525
TH01	372	311	263	179	275	176	198	144	164	155
vWA	4196	3254	2201	1784	2690	2025	2224	1000	1663	2466
D12S391	6412	6751	3844	4698	5022	4636	4043	2722	4876	3477
D13S317	6742	7536	4963	5391	8559	4993	5240	2288	4634	4584
Penta E	6040	4788	3977	4977	6072	5434	5780	3351	5659	4167
D16S539	4938	5965	4546	5100	6666	4833	3774	2622	3121	2992
D18S51	6584	5849	4301	5231	6047	4466	5458	2997	2300	4556
D19S433	10773	11276	6209	7698	7759	7175	6627	4732	7330	5911
D21S11	3436	3552	2583	2164	2643	2115	2662	1007	2228	1990
Penta D	8083	10026	6163	6201	6884	5172	5182	2963	4110	5857
D22S1045	1210	1000	429	536	814	801	615	219	1708	778
AMEL X	11479	5809	6724	7690	4566	6731	4162	3335	2161	3041
AMEL Y	103	5240	65	71	4479	61	4637	33	2432	2899



**Table 15. (continued)** Autosomal STR sample read counts. Number of reads aligning to each autosomal STR locus for the twenty buccal swab samples.

Locus	Sample									
	11	12	13	14	15	16	17	18	19	20
D1S1656	5376	4390	1473	5558	4547	4362	4083	3386	62073	56890
TPOX	3493	4921	692	3773	4116	5646	3331	2798	63913	147367
D2S441	584	2102	42	3513	2030	2931	1188	2252	39822	22323
D2S1338	2996	3857	713	3579	4244	3268	3648	2186	43467	50646
D3S1358	2981	3655	392	3692	3975	3176	2749	1638	33960	30734
FGA	1384	1699	226	1841	1684	1464	1600	851	24121	23554
D5S818	3145	3265	514	3190	3132	2332	2880	1510	30134	57229
CSF1PO	2435	3346	403	4227	3372	2020	2412	1454	30882	53816
D7S820	3975	5064	639	3805	4116	4420	3695	2406	78628	68422
D8S1179	2002	2632	522	2590	2254	2768	2150	1596	21711	21772
D10S1248	7702	7916	1552	6528	6689	8498	7787	5013	87727	99271
TH01	187	369	31	268	261	207	315	126	8024	12664
vWA	1863	3563	279	3332	3482	1984	2188	999	21180	38063
D12S391	3628	2975	1089	4031	2872	3278	2401	2859	59618	43604
D13S317	6173	4387	880	5607	5905	4832	2700	3296	129954	182160
Penta E	3954	4063	1185	2558	4386	3466	3678	2753	42652	37961
D16S539	3610	3062	1051	3314	2732	3296	2005	3117	125127	163697
D18S51	4428	4729	1209	3231	5656	3143	4450	2611	56544	56194
D19S433	5939	7647	1135	6048	5472	6162	5363	4389	75787	91257
D21S11	2642	2137	378	2935	2421	2282	1445	1381	59634	81964
Penta D	5701	6531	882	6196	5726	6116	4301	3592	146141	164540
D22S1045	799	631	97	1984	1093	612	858	343	19441	15419
AMEL X	6104	3214	1312	2816	6718	3449	5272	2317	292420	162472
AMEL Y	53	2626	13	4223	53	4105	136	2195	2898	144479

**Table 16.** Y-STR sample read counts. Number of reads aligning to each Y-STR locus for the ten buccal swab samples.

Locus	Sample									
	2	5	7	9	10	12	14	16	18	20
DYS19	18502	11749	8504	11639	13036	19249	13027	16186	11060	10402
DYS385a	7959	1820	6613	10318	7292	10052	5604	6672	7151	3953
DYS385b	11084	2053	5311	11658	8668	11558	5902	7769	8420	5061
DYS389I	60507	45717	48722	67937	70906	60150	61758	44125	31992	31282
DYS389II	60965	46141	49166	68477	71482	60595	62223	44505	32281	31556
DYS390	30144	15859	13637	28336	20109	37077	19054	17048	24030	10180
DYS391	46912	130490	97534	50184	39812	47934	38306	35461	37898	26771
DYS392	33611	53859	36595	41580	27375	63794	36877	46169	32545	17354
DYS393	38149	24486	21760	22637	27458	33514	31086	25058	24851	14202
DYS435	35105	10180	16203	42819	33117	40495	35702	40809	34343	18232
DYS437	903	35827	2446	2788	16624	2469	1077	1003	2142	1293
DYS438	42189	40243	20941	41179	39274	52371	42096	47056	40644	22995
DYS439	10271	17674	7974	18019	8570	20833	11208	12516	11192	3070
DYS448	19051	6793	3435	16230	9476	17009	9946	11696	8155	5346
DYS456	41561	8235	7725	32179	28099	41846	24740	24455	26362	10052
DYS458	10386	21190	14158	8307	7835	11307	7646	8408	5528	5988
DYS481	52483	–	–	44689	25267	52631	28074	26115	25240	14052
DYS533	37073	6736	9025	16078	14565	20974	9817	15491	11092	6206
DYS549	34700	32223	32044	40477	45352	71485	38562	41428	36265	15781
DYS570	43317	12428	14193	40769	33693	58467	32282	36011	31913	21318
DYS576	35959	13930	20004	52874	24679	53948	35509	38288	27200	20239
DYS643	35116	11993	14140	27659	20597	38647	24929	20242	24635	12036
GATAH4	38419	73856	69428	46183	44582	54108	54246	46023	36320	17051

**Table 17.** Autosomal STR control read counts. Number of reads aligning to each autosomal STR locus for the four control DNA samples.

Locus	Sample				
	007.1	007.2	NIST A	NIST B	NIST C
D1S1656	9935	31767	11201	13440	2380
TPOX	14578	86579	6926	7114	894
D2S441	5181	66997	16255	11328	1045
D2S1338	5339	22361	4940	4249	745
D3S1358	4526	37648	8349	7202	923
FGA	2464	17736	4733	3113	478
D5S818	2843	19511	8162	5307	839
CSF1PO	3005	30044	9853	6469	875
D7S820	3422	34944	14885	6529	1080
D8S1179	2149	11076	9582	6296	1269
D10S1248	9768	49356	45381	26832	4627
TH01	1298	8013	289	201	24
vWA	2090	23079	13175	8067	980
D12S391	18439	62920	12562	13081	2327
D13S317	14244	102620	12251	8177	1389
Penta E	9212	37044	6835	7089	1546
D16S539	33016	103841	8651	9214	1668
D18S51	5668	21440	9641	12021	2044
D19S433	8820	44751	17294	18301	2407
D21S11	6375	55701	6922	4932	739
Penta D	8071	51622	11700	6932	1398
D22S1045	3151	33914	3296	1951	315
AMEL X	18833	104954	28367	12447	1687
AMEL Y	21031	119731	373	10888	1634

**Table 18.** Y-STR sample read counts. Number of reads aligning to each Y-STR locus for the three control DNA samples.

Locus	Sample		
	007	NIST B	NIST C
DYS19	11639	5431	3564
DYS385a	13656	2486	1635
DYS385b	17171	3039	2163
DYS389I	80542	20441	19626
DYS389II	81290	20632	19799
DYS390	17767	8075	5990
DYS391	109527	8489	10738
DYS392	71030	16536	13917
DYS393	29066	11178	6290
DYS435	24039	12377	7885
DYS437	4900	2561	3907
DYS438	61637	14493	13700
DYS439	21538	7548	4206
DYS448	6493	3514	2640
DYS456	19180	8419	8236
DYS458	27136	4350	2812
DYS481	—	14199	10931
DYS533	9643	3707	2502
DYS549	35194	19026	14476
DYS570	26955	11199	8636
DYS576	22480	10965	11951
DYS643	16618	9196	6546
GATAH4	114705	14988	13464