Forensic Science International: Genetics 18 (2015) 78–89

Contents lists available at ScienceDirect

Forensic Science International: Genetics

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

Next generation sequencing and its applications in forensic genetics

Claus Børsting *, Niels Morling

Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark

ARTICLE INFO

Article history:
Received 24 October 2014
Received in revised form 12 January 2015
Accepted 11 February 2015

Keywords:
Next generation sequencing
Single-molecule sequencing
Forensic genetics
Review

ABSTRACT

It has been almost a decade since the first next generation sequencing (NGS) technologies emerged and quickly changed the way genetic research is conducted. Today, full genomes are mapped and published almost weekly and with ever increasing speed and decreasing costs. NGS methods and platforms have matured during the last 10 years, and the quality of the sequences has reached a level where NGS is used in clinical diagnostics of humans. Forensic genetic laboratories have also explored NGS technologies and especially in the last year, there has been a small explosion in the number of scientific articles and presentations at conferences with forensic aspects of NGS. These contributions have demonstrated that NGS offers new possibilities for forensic genetic case work. More information may be obtained from unique samples in a single experiment by analyzing combinations of markers (STRs, SNPs, insertion/deletions, mRNA) that cannot be analyzed simultaneously with the standard PCR-CE methods used today. The true variation in core forensic STR loci has been uncovered, and previously unknown STR alleles have been discovered. The detailed sequence information may aid mixture interpretation and will increase the statistical weight of the evidence. In this review, we will give an introduction to NGS and single-molecule sequencing, and we will discuss the possible applications of NGS in forensic genetics.

1. Introduction

DNA sequencing has a long history in forensic genetics. In the late 1980's and early 1990's, sequencing of mitochondrial DNA (mtDNA) was evaluated and used for case work at a time when restriction fragment length polymorphism (RFLP) analysis was the state of the art for human identification and years before the first short tandem repeat (STR) assays were developed. Successful RFLP analysis required micrograms of preferably intact DNA and that made the sensitive PCR-based mtDNA sequencing method the preferred tool for typing of low amounts of degraded sample materials, e.g., hair shafts and old bones [1–3]. Sequencing of the mtDNA control region was used extensively and the European DNA Profiling (EDNAP) Group's mitochondrial DNA population database project (EMPOP) was initiated in 1999 with the purpose of creating a common forensic standard for mtDNA sequencing and an on-line mtDNA database with high quality mtDNA population data [4,5]. Laboratories that qualified by successful participation in EMPOP collaborative exercises submitted mtDNA sequences into EMPOP and with release 11 (October 2013), the EMPOP database contained 34,617 mtDNA sequences from populations all over the world.

Sequencing was conducted with the Sanger dideoxynucleotide (ddNTP) chain terminating method [6], where the incorporation of a ddNTP to a growing DNA chain prevented further extension by the DNA polymerase (Fig. 1A). Early on, the synthesized DNA fragments were separated by slab gel electrophoresis and detected by either radioactively or fluorescently labeled deoxynucleotides (dNTPs) incorporated into the DNA fragments. Subsequent introduction of fluorescently labeled ddNTPs and capillary electrophoresis (CE) platforms [7] increased sensitivity and throughput, and decreased the cost of Sanger sequencing to a level where sequencing of complete genomes became possible. The improvements in CE technology and the development of highly sensitive PCR-based STR assays gradually reduced the need for mtDNA sequencing in forensic genetics during the 1990's. However, the Sanger sequencing method was used continuously for verification and identification of, e.g., STR alleles (see references in STRbase, http://www.cstl.nist.gov/strbase/). The ddNTP chain terminating method was also used for the so-called mini-sequencing or single base extension (SBE) reaction that was used for typing of single nucleotide polymorphisms (SNPs) [8]. SBE was a post-PCR cyclic reaction where SBE primers hybridized to the PCR products and were extended with a labeled ddNTP complimentary to the nucleotide in the SNP position (Fig. 1B). The SBE products

* Corresponding author at: Section of Forensic Genetics, Department of Forensic Medicine, Faculty of Health and Medical Sciences, University of Copenhagen, Frederik V's Vej 11, DK-2100, Copenhagen, Denmark. Tel.: +45 3532 6225; fax: +45 3532 62891.
E-mail addresses: claus.boersting@sund.ku.dk, http://retsmedicin.ku.dk (C. Børsting).

http://dx.doi.org/10.1016/j.fsigen.2015.02.002
1872-4973/

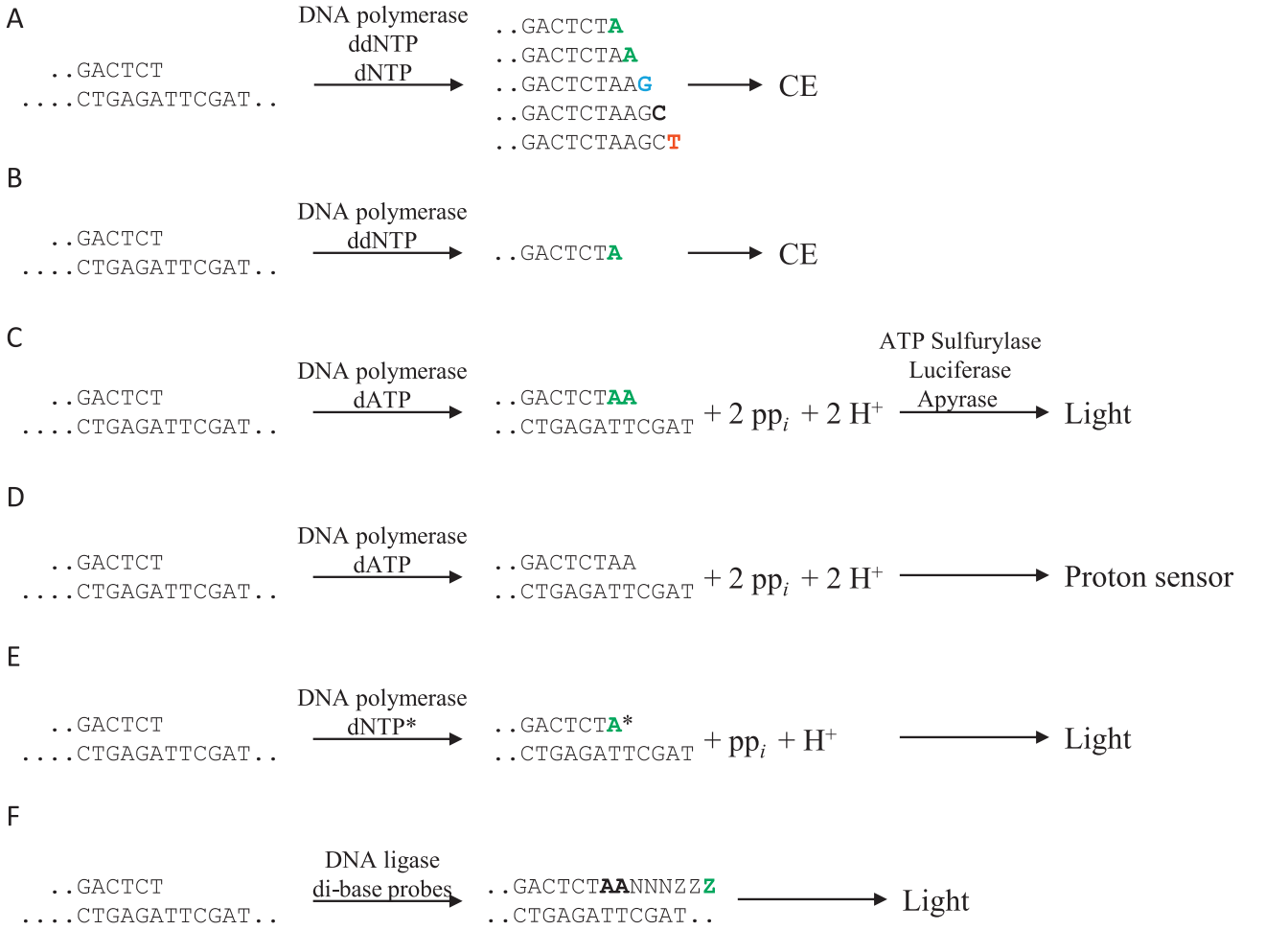

Fig. 1. Sequencing methods. (A) Sanger sequencing. DNA is synthesized in the presence of fluorescently labeled ddNTPs. The differently sized fragments are separated by CE and the sequence of fluorescently labeled nucleotides is detected by a camera. (B) Single base extention. The SBE primers are extended with a fluorescently labeled ddNTP complimentary to the nucleotide in the SNP locus. The extended SBE primers are detected by CE. (C) Pyrosequencing. Nucleotides are added sequentially to the sequencing reaction. Incorporation of one or more nucleotide(s) to the growing strand release one or more pyrophosphate(s) that are used in secondary enzymatic reactions to generate light. The light emission is detected by a camera. (D) Semi-conductor sequencing. Nucleotides are added sequentially to the sequencing reaction. Incorporation of one or more nucleotide(s) to the growing strand release one or more hydrogen ion(s) that are detected by an ion sensor. (E) Sequencing by synthesis. DNA synthesis is performed with fluorescently labeled dNTPs with reversible 3′ terminators (marked by an asterisk). Each addition of a nucleotide to the growing strand is detected by a camera. The terminator is chemically removed allowing for the next nucleotide to be incorporated. (F) Sequencing by ligation. The sequencing primer is hybridized to the target DNA and four sets of four fluorescently labeled di-base probes (all the 16 possible combinations) are added sequentially to the ligase reaction. Successful ligation of a probe to the sequencing primer is detected by a camera. The probes are cleaved (between the N and Z nucleotides) and another cycle of ligations can begin.

were detected by capillary electrophoresis, where the length of the extended SBE primer identified the SNP locus, and the ddNTP label identified the SNP allele. Panels of SNPs for human identification, pigmentary traits and ancestry information were identified, and SBE assays were validated and used in actual case work [9–14].

Pyrosequencing was presented as a real-time sequencing alternative to Sanger sequencing in 1996 [15]. Nucleotides were added sequentially to the DNA synthesis reaction, and the released pyrophosphate was used to generate light via a cascade of enzymatic reactions involving the three enzymes; ATP Sulfurylase, Luciferase and Apyrase (Fig. 1C). The light was detected in real-time by a CCD camera and thus, electrophoresis of the sequencing products was not necessary. Pyrosequencing was cheap and fast compared to Sanger sequencing, and the method was applied to mtDNA sequencing [16,17] and later also used for STR sequencing [18]. However, the short sequencing length and especially the limited multiplexing capability of the instruments were not compatible with the low amounts of DNA usually recovered from trace samples, and the method was never used in case work.

Even though the first pyrosequencing instruments never found a strong foothold in science, the pyrosequencing technology itself and the idea of real-time sequencing became the foundation on which the ongoing revolution in DNA sequencing was made. The first commercial high throughput sequencing platform, the Genome Sequencer 20 from 454 Life Sciences, used pyrosequencing [19], and it was possible to sequence the human genome in five months at a cost of $1.5 million with this technology [20]. In comparison, the first human genome was sequenced with Sanger sequencing technology during a period of 13 years and a cost of $2,700 million [21]. Several high throughput sequencing methods and platforms have since then been introduced. Most of them have been acquired by larger companies and sometimes the instruments have changed names, e.g., Solexa was changed to Illumina. Some have come and gone again, e.g., the HeliScope platform from Helicos BioSciences [22,23], and Roche has recently announced that the production of the highly successful 454 pyrosequencers will be terminated in 2015. Early on, these platforms were usually referred to as next generation sequencing or massively parallel sequencing platforms. However, with the introduction of single-molecule sequencing, some platforms were referred to as second generation sequencers and the single-molecule sequencer sometimes referred to as third generation sequencers or the next–next

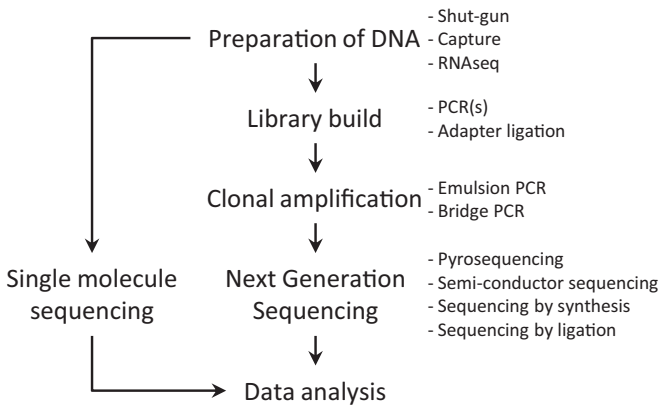

Fig. 2. Work flow for high throughput sequencing.

generation sequencers. We will use the general term next generation sequencing (NGS) in this review to cover all sequencing methods, except for single-molecule sequencing, that have been developed after Sanger sequencing and the early pyrosequencing methods. Single-molecule sequencing will be addressed in a separate section. There are many excellent reviews in the literature [21,24–31] that describes the various platforms in this rapidly changing field. In this review, we will focus on the possible applications of NGS in forensic genetics and only give an introduction to high throughput sequencing (Fig. 2).

2. The basics of NGS

The capability of some NGS platforms is so large that the aim of the sequencing assay may simply be to sequence every double stranded DNA molecule in the sample material. Sequencing without any prior selection of targets is known as shotgun sequencing and requires fragmentation of micrograms of DNA into short fragments of 50–500 base pairs either by mechanical force, enzymatic digestion or random insertion of transposons [21,24,25,31]. If the sample material is cDNA, shotgun sequencing may generate a gene expression profile of the sample which is known as RNA sequencing or RNAseq [32,33].

The alternative to shotgun sequencing is usually called targeted (re-) sequencing and involves an initial enrichment step that either amplifies the selected regions by PCR, or uses probes to capture the regions or uses a combination of probes and enzymatic reactions [34–36]. The probes may be attached to a solid surface (e.g., a slide or a bead) or they may be biotinylated and hybridize to their targets in solution. Either way, the purpose is to capture the selected genomic regions and eliminate the unwanted fragments of DNA. The DNA is subsequently eluted and used for sequencing. If the probes are used in combination with a DNA polymerase, the captured fragments are used as templates for DNA synthesis. The probes may be biotinylated and the primer extended probes captured by streptavidin beads, or the DNA synthesis is followed by a DNA ligase reaction that generates a circular product resistant to subsequent exonuclease treatment. The newly synthesized DNA is isolated from the genomic DNA and used in the downstream NGS reaction. Large portions of the genome may be captured with these technologies, e.g., all the known coding regions (known as exome sequencing) or panels of relevant genes related to particular diseases. There are countless numbers of commercial capture assays available from different companies, e.g., the SureSelect Human All Exon Kit (Agilent), the HaloPlex Exome Kit (Agilent), the Ion AmpliSeq™ Exome RDY Kit (Life Technologies), the Nextera Rapid Capture Exome Kit (Illumina), the SeqCap EZ Human Exome Library (Roche NimbleGen), the TruSight One Sequencing Panel (Illumina) that targets 4800 genes or the Ion AmpliSeq™ Cancer Panel (Life Technologies) that targets 400 genes by PCR. The above mentioned companies also provide services for generation of customized panels defined by the user for specific projects or purposes. The major advantage of capture methods is that the majority of the sequencing capacity is focused on the regions of interest. This allows for more efficient sequencing experiments either by allowing more individuals to be sequenced at the same time or that the number of sequences for each nucleotide position (known as the coverage or sequencing depth) is higher. The PCR-based capture method is by far the most sensitive and requires <10 ng DNA per multiplex reaction whereas the probe based methods typically requires 50–500 ng DNA. In contrast, the PCR-based captures are limited by the level of multiplexing capability (up to 6144 amplicons with the Ion AmpliSeq™ technology).

Once the DNA has been prepared for either shotgun or capture sequencing, the fragments are used to generate a library. The library is constructed by ligating adapters to the fragments or by one or two PCR reactions where the PCR primers are tagged with sequences needed for the downstream reactions. The adapters or the PCR primer tags may include specific sequences for clonal amplification of the library (see below), target sequences for the NGS reaction, a key sequence with 4–8 nucleotides used for quality control of the NGS reaction and a 6–10 nucleotide barcode for identification of the sample. The various sequence elements are combined in various ways depending on the assay and the NGS platform, e.g., the barcode can be left out entirely if only one sample is sequenced (typical for shotgun sequencing), or barcodes may be placed in both ends or in only one end of the library, or different barcodes may be used in either end to have a high number of combinations when many samples are sequenced in the same experiment. The construction of the library is the critical step of the experimental design. The choice of barcodes dictates how many samples can be sequenced. The choice of key and tag sequences for the clonal amplification and NGS sequencing partly dictates the choice of NGS platform. Three examples of library constructions are described in Fig. 3.

Multiple libraries may be pooled in equal amounts prior to the clonal amplification. The number of samples that can be analyzed in the same experiment depends on a number of factors: (1) The number of available barcodes, (2) the sequencing capacity of the NGS platform, (3) the numbers and sizes of targeted regions and (4) the desired sequencing depth.

The library pool is used as target for the clonal amplification step. Individual DNA molecules are hybridized to a primer on a solid surface, and each molecule is amplified by PCR in a reaction that is isolated from the other DNA molecules in the library pool (thus, the name clonal amplification). The physical separation of the molecules is secured by hybridization of one DNA molecule to one bead and generation of an oil–water emulsion with one bead per droplet (emulsion PCR) [19,37] or by hybridizing the DNA molecules to a slide (bridge PCR) [38]. Millions of individual DNA molecules (clones) may be amplified simultaneously and, after amplification, thousands of copies of each original DNA molecule form an immobilized “cluster of DNA” on the bead or the slide. Each DNA cluster forms an ideal target for sequencing and all the clusters may be efficiently sequenced in parallel on a NGS platform (thus, the name massively parallel sequencing). The beads from the emulsion PCR (emPCR) are placed in picoliter-sized wells with one bead per well whereas the slide from the bridge PCR is used directly for sequencing.

The DNA sequence of each cluster is determined in real-time by one of four methods (Fig. 1C–F); pyrosequencing (Roche 454 sequencing), semi-conductor sequencing (Thermo Fisher Scientific Ion Torrent™), sequencing by synthesis (Illumina®) or sequencing by ligation (Thermo Fisher Scientific SOLiD™ and BGI-Shenzhen

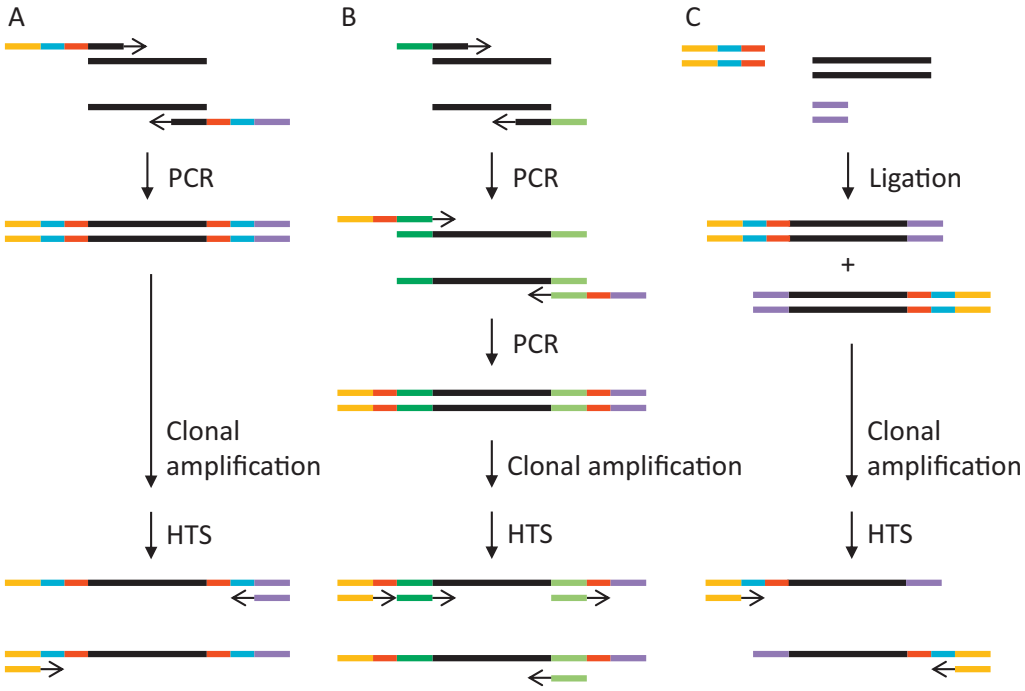

Fig. 3. Examples of library building and sequencing strategies. (A) The library is generated by one PCR reaction. The PCR primers include five elements; the target sequence (in black), the barcode for sample identification (in red), the key sequence for sequence quality control (in blue) and sequencing targets (in orange and purple). One of the sequencing targets is also used to hybridize the library to the solid surface during the clonal amplification step. With two sequencing targets, it is possible to perform directional sequencing of only one strand by choosing a sequencing primer complementary to either the orange or the purple sequencing target. With only one sequencing target (when the orange and the purple sequences are the same), both strands would be sequenced in the NGS reaction. (B) The library is generated by two PCRs. In the first PCR, the primers include the target sequence (in black) and the sequencing targets (in two shades of green). In the second PCR, the primers hybridize to the sequencing targets and include tags with the barcode (in red) and sequences for hybridization to the solid surface used for the clonal amplification. The target sequence (in black) is sequenced via the two sequencing targets (in green) whereas the barcodes are sequenced in separate reactions. (C) The library is generated by ligation of adapters to the fragmented genomic DNA. One adapter includes the barcode for sample identification (in red), the key sequence for sequence quality control (in blue) and the sequencing target (in orange). The second adapter includes the sequence for hybridization to the solid surface used for the clonal amplification. Four different products will be generated by the ligation; the two products shown in the figure, where two different adapters are ligated to the DNA fragment, and two products where the same adapter ligates to both ends. The later products cannot be used in the downstream reactions. Sequencing is conducted from hybridization of a sequencing primer complimentary to the sequencing target (in orange). Both strands will be sequenced because the adapter with the sequencing target ligates to either the forward or the reverse strand in equal numbers. HTS (high throughput sequencing).

Complete Genomics). In pyrosequencing and semi-conductor sequencing, the nucleotides are added sequentially to the reaction. In some clusters, no DNA synthesis will take place because the added nucleotide cannot extend the growing strand. In others, one or more nucleotides will be added, and the generated light signal or the number of protons released will be detected and interpreted. With the sequencing by synthesis method, all four fluorescently labeled nucleotides are present in the reaction, and one nucleotide is added to the growing DNA strand in all clusters. The nucleotides are reversibly blocked in the 3′ position, which prevents incorporation of more than one nucleotide at the time. This is an advantage when the sequence contains stretches of the same nucleotide (homopolymer stretches). In pyrosequencing and semi-conductor sequencing, several nucleotides will be incorporated at homopolymer stretches and more light and protons will be detected, respectively. However, if the stretch is longer than five nucleotides, it may be difficult to deduce the correct number of nucleotides in the homopolymer. Sanger sequencing suffered from a similar problem when similar sized fragments with the same ddNTP were difficult to separate by electrophoresis. Sequencing by ligation is very different from the other methods as it involves ligation of fluorescently labeled probes to a primer. The probes may vary in one (BGI-Shenzhen Complete Genomics) or two (Thermo Fisher Scientific SOLiD™) positions. Only probes that are perfectly complementary to the target sequence will ligate to the growing chain of primer and probes. Consecutive rounds of ligations and cleavage reactions generate a patchwork of fluorescent signals from the same cluster that may be combined into a complete sequence [21,24,25,31].

The maximum number of bases that may be sequenced in each cluster (known as the read length) varies between the NGS methods. The read length has been an important focus point for the commercial competition and considerable improvements in read length has been achieved on the pyrosequencing and the sequencing by synthesis platforms. Today, pyrosequencing generates read lengths that are comparable to the read lengths of the Sanger sequencing method (600–1000 bps) and the sequencing by synthesis platforms have reached a read length of 300 bps. The first semi-conductor platform was launched in 2011, and the read lengths are now up to 400 bps. In contrast, sequencing by ligation generates very short read lengths (<75 bp), and it has not changed much since the platforms were released in 2006.

Another major focus point for the commercial competition between NGS platforms has been the overall sequencing capacity or the total number of clusters (or reads) that are sequenced per run. The sequencing capacity is sometimes calculated as the number of sequenced bases per run, however, this may be misleading since the read lengths varies between platforms. The HiSeq 2500 System (Illumina®) has the largest capacity and can sequence 4000,000,000 clusters in one experiment. It takes 5–11 days to complete a run, however, the experiment may generate enough sequences to cover 5–10 (almost) complete human genomes with an average coverage of more than thirty. The smallest of the commercial high throughput sequencing platforms

is the GS Junior System (Roche). It can sequence 100,000 clusters in 10 h and belongs to one of the so-called bench-top sequencers that also include the MiSeq (Illumina®), the NextSeq® 500 (Illumina®), the Ion Proton™ System (Thermo Fisher Scientific Ion Torrent™) and the Ion PGM™ System (Thermo Fisher Scientific Ion Torrent™). The capacity of these instruments is in the range from 5 to 400,000,000 clusters, and the run times range from 2 to 55 h.

A high degree of flexibility in the experimental design is possible on the sequencing by synthesis and semi-conductor platforms. There are different sizes of flow cells (Illumina®) or chips (Ion Torrent™) available that vary in the number of clusters that may be sequenced, and there are different reagent kits available that vary in the number of nucleotide cycles. The number of nucleotide cycles regulates the read lengths and also affects the run time of the instrument (more cycles generate longer reads and takes longer time). In general, the run time on the pyrosequencers and the semi-conductor platforms are relatively short because signal detection is performed in real-time, whereas signal detection on the sequencing by synthesis and sequencing by ligation platforms is done by imaging which makes the run times longer. However, the manual preparation of the samples for the sequencing by synthesis platform is short compared to the other platforms because the cluster generation by bridge PCR and the sequencing reaction are an automated protocol on the flow cell performed by the NGS instrument. The other three sequencing methods use libraries that are clonally amplified by emPCR which involves many pipetting steps and considerable hands-on time. The large flexibility of the bench-top instruments makes them highly suitable for capture-based sequencing experiments where the numbers of samples or the sizes of the captured regions may vary from project to project or from experiment to experiment. This makes the platforms ideal for research. However, the scalability also makes the platforms interesting for diagnostic laboratories that perform routine genetic investigations.

3. Single-molecule sequencing

Detection of the sequence of a single DNA molecule instead of a cluster of clonally amplified DNA is often referred to as third generation sequencing [26,29,30]. With these methods, the original DNA or RNA molecules may be analyzed, and any biases generated by the capture and clonal amplification steps are eliminated (Fig. 2).

The HeliScope platform (Helicos Biosciences) was the first commercial single-molecule sequencing platform to be launched [22,23]. However, the company had a very short life time, and the HeliScope platform is no longer produced. Soon after, the PacBio platform (Pacific Biosciences®) was launched [39,40] and recently, a beta-version of the MinION™ sequencer (Oxford Nanopore) was released for testing by members of the MinION access program. The HeliScope and PacBio platforms use variations of the sequencing by synthesis technology (Fig. 1E). For the HeliScope platform, the original DNA is fragmented and a poly(A) tail is added to the fragments. This library is subsequently hybridized to anchored poly(T) probes on a slide, and the sequences are determined by primer extension using cycles of sequential addition of Cy5-labeled nucleotides and fluorescence imaging. The nucleotides are reversibly blocked in the 3′ end to ensure that only one nucleotide is incorporated per cycle. The run time on the HeliScope platform is 2–9 days, and the read lengths are only 50 nucleotides. However, up to 1500,000,000 reads may be generated. On the PacBio platform, DNA polymerase/DNA template complexes are immobilized at the bottom of zero-mode waveguide (ZMW) wells that are zeptoliter (10^{-21} L) sized wells only nanometers in diameter. One DNA polymerase/DNA template complex fits into one ZMW well and the template is sequenced using four different fluorescently labeled nucleotides. The fluorophore is linked to the terminal phosphate, and it is released when the nucleotide is incorporated into the growing strand. The fluorophores are excited by multiple lasers, and the pulse of fluorescence is monitored by a camera. The duration and intensity of the pulse determines the identity of the incorporated nucleotide. All four nucleotides are added to the reaction, and the generated signals are detected in real-time with a speed of approximately 5 nucleotides per second. That makes the run time on the PacBio platform short and it usually last less than 1 h. The PacBio may generate read lengths of more than 15,000 nucleotides, and the maximum capacity is currently 150,000 ZMWs. Furthermore, the PacBio platform uses the ϕ29 DNA polymerase that is capable of multiple displacement amplification. Thus, a circular DNA template may be sequenced several times in one experiment and a single ZMW well may generate a sequencing depth of ten or more depending of the size of the template DNA and the read length.

The MinION™ platform uses a completely different technology based on the transport of DNA molecules through a nanopore embedded in a lipid bilayer or a synthetic polymer [30,41,42]. An electric field across the membrane will drive the DNA molecules through the pore and the current of (other) ions through the pore will be partly blocked as the DNA pass through. The decrease in current amplitude is detected and used to determine the nucleotide sequence of the DNA passing through the pore. The pore used by the MinION™ is not known, however, many channel proteins, synthetic solid-state nanopores and even scaffold structures of DNA or proteins have been used for nanopore sequencing. On the MinION™, a protein/DNA complex binds reversibly to the pore. The undisclosed protein unfolds the double stranded DNA and single stranded DNA pass through the pore. A hairpin is attached to the DNA molecule during sample preparation, and when the hairpin sequence has passed through the pore, the reverse strand follows and the sequence of the complementary strand may be determined. The MinION access program was initiated in the spring of 2014 and peer-reviewed results are limited [43]. However, in our own experience, the MinION™ may generate read lengths that are longer than 70,000 nucleotides, and the total numbers of reads are 2000–10,000 [Jill Olofsson, unpublished results].

The long read lengths of the PacBio and MinION™ platforms will make it possible to determine the haplotype of an individual in long stretches of DNA and will simplify the assembly of genome regions with multiple duplications and repeats [44,45]. Furthermore, it may be possible to determine epigenetic modifications of the DNA in real-time since modified nucleotides, e.g., 5-methylcytosine, 5-hydroxymethylcytosine and N6-methyladenine, give off another fluorescent pulse on the PacBio and decrease the current amplitude differently in nanopores than unmodified nucleotides [46–49].

Both platforms suffer from very high error rates because current methods of signal detection are inadequate. The base call error rate is estimated to be >15% on the PacBio [40] and >30% on the MinION™ [[43], Jill Olofsson, unpublished results]. The errors on the PacBio seem random and the quality of the consensus sequence generated in each ZMW may therefore be improved by increasing the sequencing depth using circular DNA templates and ϕ29 DNA polymerase (see above). In contrast, insertion/deletion errors are very frequent on the MinION™ which makes it very difficult to align the generated sequences accurately and exploit the sequence information in the long reads. Another disadvantage is the relatively large amount of input DNA required. Even though individual DNA molecules are sequenced on these platforms, the amount of input DNA is 250–5000 ng on the PacBio and >1000 ng on the MinION™.

4. NGS solutions in forensic genetics

The idea of sequencing every DNA (and/or RNA) molecule in the sample is very intriguing to a forensic geneticist, who is used to dealing with the challenge of obtaining sufficient information from trace samples that often contain DNA from more than one contributor. However, shotgun sequencing requires micrograms of DNA and is not applicable for many of the forensic samples. Also, reproducibility may be impossible, since the shotgun experiment is not directed toward specific targets but generate sequences from random positions in the genome. Thus, two shotgun sequencing experiments of the same sample, or e.g., a trace sample and a reference sample from a suspect, will generate different results. Furthermore, shotgun sequencing requires exhaustive data analyses that are both time-consuming and may generate different "DNA profiles" depending on the choice of NGS platform and alignment software. Concordance studies between platforms have demonstrated that as many as 20% of the SNPs and 80% of the insertion/deletions (indels) called by one platform were not reproduced by typing the same sample on another platform [50–53]. Large portions of the inconsistencies were seen in regions of low (or no) sequencing depth for one or both platforms or caused by systematic errors introduced by the different methods of alignment and variant calling. Even though the capacity of some NGS platforms is huge, they are only just able to sequence genomes the size of the human genome and large portions of the genome are only covered by low number of reads in typical shotgun sequencing experiments. This leads to a high risk of mis-interpretation of the sequencing data and to the lack of reproducibility observed in concordance studies. Finally, full genome sequencing seems excessive in most forensic genetic cases where the purpose is primarily to establish the identity of the individual(s) contributing to the sample and possibly estimate any phenotypical characteristics of the individual(s) or identify the specific tissue type(s) in the sample. This requires relatively few markers and a capture based approach will be much more economical and require less sample material.

Today, the core forensic markers are typed with PCR-CE and there are individual assays for autosomal STRs, Y-chromosome STRs, X-chromosome STRs, indels, mtDNA SNPs, autosomal SNPs, Y-chromosome SNPs, ancestry informative markers (AIMs), phenotypical markers, mRNA, etc. PCR-CE may be performed in one work day, whereas NGS takes minimally 2–3 days. However, one of the major advantages of NGS is that all (or most) of the PCR-CE assays may be combined into a single NGS assay if it is possible to develop a capture for the relevant loci. One NGS assay with many different markers will save time in cases where supplementary investigations are needed and reduce the overall time a sample is processed in the laboratory. Among the various capture methods, PCR continues to be the most sensitive and currently, it is the only method that approach the level of sensitivity required for forensic genetic case work. Another important advantage is that the fragments do not need to be separated by lengths in CE and thus, all the analyzed fragments can be designed to be as short as possible which will improve the chance of typing degraded DNA/RNA.

Combining nuclear markers with mtDNA or mRNA markers in a sequencing assay may prove to be difficult. DNA, mtDNA and RNA may be co-extracted and the RNA converted to cDNA in a separate Reverse Transcriptase reaction. However, the large variation in target copy numbers will make construction of a combined multiplex PCR very difficult. Also, it is important to keep in mind that mRNA sequencing needs to be semi-quantative to allow tissue identification of the sample and that genomic DNA is unwanted in cDNA analyses. Nevertheless, it may be possible to pool PCR products from separate PCR captures of nuclear DNA, mtDNA and cDNA prior to the library build (adapter ligation or the second PCR) or, more likely, to pool DNA, mtDNA and cDNA libraries prior to the clonal amplification step. This way, it should be possible to sequence all relevant markers in a single sequencing reaction. However, it is uncertain whether this is a practical solution for case work, because the information obtained from mRNA and mtDNA is not needed in all cases and the case officer needs to process all the sequencing data if the information is generated even though the information is irrelevant to the case. Also, sequencing of mtDNA and cDNA will take up a large portion of the sequencing capacity which eventually will result in fewer samples per sequencing run and a higher cost of the investigation. Similarly, it may be argued that ancestry information, phenotypical traits or certain human identification markers are irrelevant in other case work scenarios. Therefore, flexible NGS solutions for various types of cases, including assays with large number of markers, will be preferred.

It is generally accepted that sequencing by ligations has the lowest error rate among the NGS methods followed by sequencing by synthesis, semi-conductor sequencing and pyrosequencing, in that order [29,50,52,53]. However, these error rates are from genome sequencing studies and can be misleading, because the errors are unevenly distributed and typically related to specific sequence elements, e.g., sequencing of homopolymer regions (see above), and it would be much too simple to state that sequencing by ligations is the best platform for forensic genetic applications (it is not because of the short read lengths). In order to evaluate the quality of a given NGS platform/assay, it is necessary to properly validate the genotypes against existing methods. In the last 1–2 years, there have been numerous reports in the literature from molecular diagnostics laboratories where NGS results have been compared to mainly Sanger sequencing [54–60]. The conclusions drawn from these studies are that probe or PCR capture based NGS analyses have matured sufficiently to be used in clinical diagnostics and will gradually replace Sanger sequencing as the gold standard. The main concerns are the data analyses and the overwhelming number of (new) variants that are detected. Each variant must be evaluated and classified as benign or disease related, and this may be difficult for the local clinician. Another concern has been the detection of variants that are not related to the disease under investigation. Large capture based investigations, e.g., exome sequencing, may reveal variants in other genes that may be disease related. How to handle this information raises ethical considerations for the clinician. These discussions are interesting because forensic genetics face many of the same challenges and will have to consider many of the same questions on how much and which loci to sequence, what to report and whether ignoring sequence information is prudent, and certainly, there will be a number of ethical and legislative considerations by introducing NGS in forensic genetics.

4.1. STR sequencing

STRs are essential to crime case work and will continue to be so, because of the large national DNA databases with STR profiles from criminal offenders and irreplaceable trace samples from old cases. Consequently, any NGS assay designed for forensic genetics must be able to sequence the core STR loci. However, most NGS studies focus on SNPs, small indels and copy number variations whereas repeats have not attracted much attention even though repeats cover almost half, and STRs alone 15%, of the human genome [61]. In the years after the first NGS platforms were launched, the read lengths of most instruments were too short to span many repeat structures which made it difficult to align reads with repetitive sequences and often these reads were simply ignored. The pyrosequencers were the only platforms with sufficient read length to sequence the core STR loci used in forensic genetics and to date, most of the forensic literature with NGS STR data were

Table 1
Examples of D12S391 alleles with the same length.

D12S391[21]AGAT[11]AGAC[9]AGAT[1]
D12S391[21]AGAT[11]AGAC[10]
D12S391[21]AGAT[12]AGAC[8]AGAT[1]
D12S391[21]AGAT[12]AGAC[9]
D12S391[21]AGAT[13]AGAC[7]AGAT[1]
D12S391[21]AGAT[13]AGAC[8]
D12S391[21]AGAT[13]GGAC[1]AGAC[7]
D12S391[21]AGAT[14]AGAC[6]AGAT[1]

produced with pyrosequencing technology [62–70]. In a few reports, sequencing by synthesis [71–73] and recently also semiconductor sequencing [70,74] were used. The libraries were constructed directly by PCR or by adapter ligation as described in Fig. 3A and C, respectively.

In contrast to fragment length analysis by PCR-CE, sequencing reveals the true variation of STR loci. Previously unknown STR alleles and more overall variability has been found by NGS of mainly complex and compound STRs [65,66,69], whereas few new alleles have been detected by sequencing of simple STRs [62,69,74]. Complex and compound STRs consist of different sub-repeats and, if the individual sub-repeats are polymorphic, the number of possible alleles will be much higher than in simple repeats (it is equivalent to having two STRs in a haplotype). In one study of 197 Danes, 53 different alleles in D12S391 were detected by NGS whereas only 15 different alleles were detected by PCR-CE. Many alleles in D12S391 have the same length, but different sequence compositions in the repeat region. So far, the top score is eight different alleles with a length equivalent to 21 repeats (Table 1) [66]. Sequence variations may also be found in the flanking regions of STRs [62,64,75], e.g., the rs6736691 SNP near D2S1338 [75]. One SNP with a minor allele frequency of a reasonably size may generate two families of SNP–STR haplotypes and increase the number of SNP–STR alleles by a factor of two.

Discovery of many new STR and SNP–STR alleles with the same sizes makes the old PCR-CE based nomenclature for STR alleles inadequate. A new transparent description of STR sequences is much in demand and the International Society of Forensic Genetics (ISFG) has initiated a working group with the purpose of finding a common definition for naming sequenced STR alleles. In this review (Table 1,Figs. 4 and 5), we have used the nomenclature of Gelardi et al. [66], where the name is divided into four elements:

A

BarcodeId	AMELX	AMELY	CSF1PO	D16S539	D3S1358	D5S818	D7S820	D8S1179	TH01	TPOX	vWA
IonXpress_049	1	-	11	12	15, 16	12	10, 11	8, 11	6, 9.3	8, 11	14, 16
IonXpress_050	1	-	11, 13	10, 11	15, 17	11	9, 10	10, 14	9, 9.3	8	15
IonXpress_051	1	-	11, 12	11	16, 18	12	10, 11	9, 14	9	8, 10	14, 19
IonXpress_052	1	-	10, 11	11	16, 17	10, 11	9, 12	10, 13	6, 9.3	8	16
IonXpress_053	1	-	10, 11	12, 14	17	11, 12	11	10, 14	6, 8	8	16, 20
IonXpress_054	1	-	9, 13	11, 12	16, 17	12	10	*14*	9.3	8, 9	15, 18
IonXpress_055	1	1	12	12	15, 17	11, 12	8, 10	10, 13	6, 7	8, 11	17, 18
IonXpress_056	1	1	10, 11	12, 13	15, 18	11, 12	9, 11	12, 13	7, 9	8	17, 18
IonXpress_057	1	-	12, 13	9, 12	14, 15	12	10, 12	10, 13	6, 9.3	11, 12	14, 16
IonXpress_058	1	1	11	11	15, 16	11, 12	8	13	6	8, 11	14, 18
IonXpress_059	1	1	10, 11	11, 12	16, 17	11, 12	11, 13	13, 14	6, 7	8, 9	15, 18
IonXpress_060	1	1	10, 12	9, 12	15, 18	11, 13	9, 11	13	9	8	17, 18
IonXpress_061	1	1	11	9, 11	17, 18	11	9	13, 15	6	8	16, 17
IonXpress_062	1	1	12, 14	10, 13	17, 18	11	10, 10.1	11, 12	6, 9.3	8	18
IonXpress_063	1	-	12	10, 11	15, 16	13	8, 12	14	7	8, 11	15, 17
IonXpress_064	1	1	10, 12	11, 12	16, 19	12	7, 12	12, 18	7, 9.3	11	17

Locus	Allele	Counts	Total cov	Sequence
D8S1179	14	3995	8655	TATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC
D8S1179	14	3938	8655	TATCTATCTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATC

B

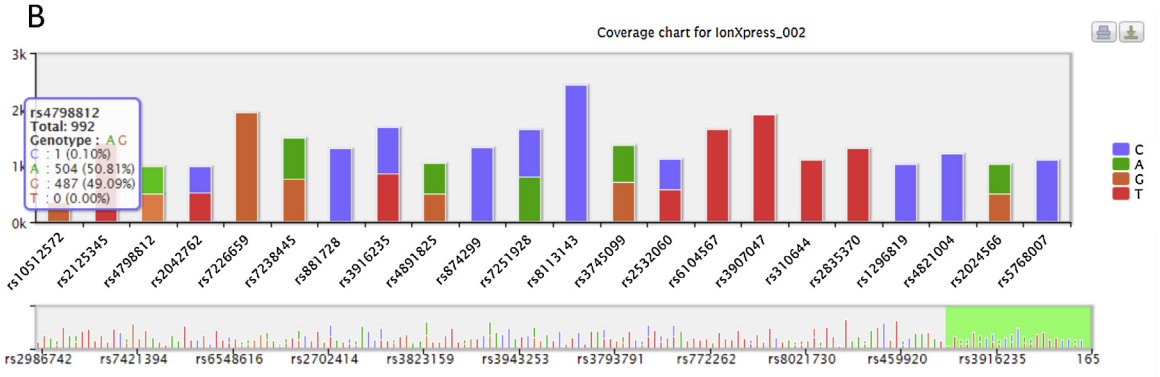

Fig. 4. Analysis of STRs and SNPs using the Torrent Suite Server. (A) Profile summaries of 15 samples typed with the Ion Torrent™ HID STR 10-plex. The D8S1179 locus in the individual labeled with the IonXpress 054 barcode is highlighted, and some details of the NGS reaction for this locus are shown below. The individual has two D8S1179 alleles of the same length, and the single nucleotide difference is highlighted. The allele balance of the two alleles was almost 1:1 (3995 reads of D8S1179[14]TCTA[14] versus 3938 reads of D8S1179[14] TCTA[1]TCTG[1]TCTA[12]). Please note that the Torrent Suite Server displays the repeat as TATC repeats whereas D8S1179 is defined as TCTA repeats in the forensic genetic literature. (B) Coverage chart of an individual typed with the HID-Ion AmpliSeq™ Ancestry Panel. Only 22 of the 173 SNPs in the panel are shown in the chart as indicated by the green bar below the chart. Details of the NGS reaction for the rs4798812 locus are shown in a box on the left. The allele balance of the two alleles was almost 1:1 (504 reads with an A and 487 reads with a G). One read was called as a C. This base call was most likely an error caused by the PCR.

Fig. 5. Analysis of STRs and SNPs using the ForenSeq™ Universal Analysis Software. Results from two different samples typed with the ForenSeq™ DNA Signature Prep Kit are shown in A and B. Genotype calls for each locus are shown in small boxes. (A) The NGS result for D9S1122 is highlighted, and some details are shown in the large box. The individual has two D9S1122 alleles of the same length. The allele balance was approximately 1.2:1 (an intensity of 866 for the D9S1122[12]TAGA[1]TCGA[1]TAGA[10] allele versus an intensity of 704 for the D9S1122[12]TAGA[12] allele). (B) The NGS result for rs2399332 is highlighted, and some details are shown in the large box. This sample was a 1:10 mixture and the ForenSeq™ Universal Analysis Software correctly designated the sample as a possible mixture (not shown). For rs2399332, the genotypes of the major and minor contributor are C and AC, respectively. The allele balance was approximately 1:20 (A versus C) as expected. Please note that the A allele was not called by the ForenSeq™ Universal Analysis Software.

(1) The locus name, (2) the length of the repeat region divided by the length of the repeat unit, (3) the sequence(s) of the repeat unit (s) followed by the number of repeats and (4) variations in the flanking regions.

More variable loci also mean more statistical power of the investigations and will reduce the number of loci that needs to be typed to solve a case. In the largest population study to date, the match probability decreased from 0.0001 to 0.000005, and the typical paternity index increased from 59 to 415 when three STRs D3S1358, D12S391 and D21S11 were typed by NGS in 197 Danes, and the results were compared to PCR-CE results [66]. These three loci are highly polymorphic STRs with more than one sub-repeat and the number of detected alleles was almost tripled by sequencing compared to PCR-CE. For simple STRs, the number of detected alleles is not expected to increase by a factor of two or three and the difference in statistical power between PCR-CE and NGS results will be smaller.

Another interesting observation from the same study was that approximately 30% of the homozygous genotype calls by PCR-CE turned out to be heterozygous when the individuals were sequenced. This demonstrates another important advantage of NGS of STRs. Sequencing of complex and compound STRs with many alleles of the same size may simplify mixture interpretation, if the contributors have alleles of the same size with different sequence compositions or if the true allele of the minor contributor has a different sequence than the stutter artifact of the major contributor. It was recently demonstrated that sequences from the minor contributor in 1:100 and 1:50 mixtures were detectable by NGS [74,76] – something that is not possible with the current PCR-CE technology. In these types of mixtures, the reads from the minor contributor will be difficult to separate from stutters and noise sequences, however, the mere fact that they could be identified opens up for new possibilities in mixture interpretation and it is certainly something that should be explored further.

4.2. The first commercial NGS kits for forensic genetics

Thermo Fisher Scientific launched two SNP typing assays in 2014 designed for the Ion PGM™ System: (1) the HID-Ion AmpliSeq™ Identity Panel for human identification [76,77] that amplifies 124 autosomal SNPs, including most of the SNP*for*ID [78] and Individual Identification SNPs (IISNPs) [79], and 34 Y-chromosome SNPs, and (2) the HID-Ion AmpliSeq™ Ancestry Panel for ancestry estimation that include most of the Ancestry Informative Markers (AIMs) in the Seldin [80] and Kidd laboratory selection panels [81]. Furthermore, Thermo Fisher Scientific is working on a panel of core forensic STRs, and an early version of this panel, the Ion Torrent™ HID STR 10-plex, has been tested by forensic laboratories [74]. The assays use the Ion AmpliSeq™ technology, where the PCR primers are partly degraded prior to adapter ligation. This reduces the lengths of the fragments to be sequenced, and it removes most of the primer-dimers from the library which makes the sequencing more efficient. The PCR fragments are clonally amplified by emPCR and sequenced in both directions using semi-conductor sequencing technology (Fig. 1D). The library construction and sequencing strategy are described in Fig. 3C. The sequence data are analyzed using the analysis softwares on the Torrent Suite Server (Fig. 4). The sensitivity of the two SNP typing assays, that both amplify >120 SNPs in one PCR, was 0.5–1 ng [[76], unpublished results], whereas full STR profiles were obtained from only 50 pg with the Ion Torrent™ HID STR 10-plex [74]. It was also noteworthy, that the STR assay generated full STR profiles from degraded samples whereas PCR-CE typing of the same samples resulted in partial profiles. This was most likely due to the short <170 bp PCR products generated by the Ion Torrent™ HID STR 10-plex.

So far, the strategy of Thermo Fisher Scientific has been to develop assays that may be used as supplement to PCR-CE typing. In contrast, Illumina® has announced that their strategy is to replace PCR-CE with PCR-NGS. At the time of writing of this review, Illumina® is conducting beta-tests of their new ForenSeq™ DNA Signature Prep Kit together with selected forensic laboratories and Illumina® plans to launch the kit in the fall of 2014. The ForenSeq™ DNA Signature Prep Kit amplifies 27 autosomal STRs, 8 X-STRs, 25 Y-STRs, 95 autosomal human identification SNPs, 56 autosomal AIMs and 24 autosomal SNPs associated with pigmentary traits, all in one multiplex PCR. The multiplex includes, among others, all of the STR loci in the CODIS and European standard set, most of the SNP*for*ID [78] and IISNPs [79] and all of the HIrisPlex loci [82]. The ForenSeq™ DNA Signature Prep Kit will be introduced together with the MiSeq FGx platform, a MiSeq developed specifically for forensic genomics. The MiSeq FGx platform is supported by the ForenSeq™ Universal Analysis Software that manages both the experimental set-up and data analysis (Fig. 5). The PCR fragments are clonally amplified by bridge PCR and sequenced with the sequencing by synthesis technology (Fig. 1E). The library construction and sequencing strategy are described in Fig. 3B. The maximum read length of the standard

MiSeq reagent kits is 300 nucleotides. However, for the ForenSeq™ DNA Signature Prep Kit 350 cycles of sequential addition of labeled nucleotides and fluorescence imaging is performed on the forward strand to allow sequencing of the longest STRs. The reverse strand and the barcodes are sequenced in short separate reactions. Thus, the reverse strand is only sequenced to confirm the sequence of the reverse PCR primer and the first few nucleotides of the amplicon. In our hands, the ForenSeq™ DNA Signature Prep Kit generated results from 50 pg–10 ng of input DNA. However, detailed evaluation of the sequence data is required and has not been completed at the time of writing. Proper analyses of genotype concordance, degraded samples and mixtures are pending.

One of the major challenges will be to develop a forensic NGS tool for analysis and reporting of the sequence data. With NGS data, it is not possible to analyze the sequences manually or even to analyze every genotype call manually. Therefore, the software solution must be completely trustworthy and thoroughly validated before they can be used in real case work. Thermo Fisher Scientific and Illumina® have developed software solutions for analysis of their kits, however, they are not sufficiently sophisticated for forensic genetics. There is a tendency to analyze all STRs or all SNPs with the same criteria and there are very few or no options for the user to alter the parameters for analysis. It is well known that different loci must be analyzed with different criteria, and historically, individual laboratories have defined different analysis parameters based on in-house validation studies and according to different standards of accreditation [83,84]. This will also be necessary with NGS kits and the software must be able to accommodate this demand from the forensic community. A "black box" for analysis is not acceptable and will probably not be used by many forensic laboratories [85]. For each STR locus, there should be user defined options for (1) the minimum number of reads used to call the genotype, (2) acceptable stutter ratios (reads with the same sequence as the genotype except for the number of repeats), (3) acceptable noise ratios (reads that are not identical to the genotype), and (4) acceptable allele balances. The later may depend on the lengths of the two alleles because of the tendency to generate more reads of the shortest allele [64,74], thus, the acceptable allele balance may vary depending on the length difference between the alleles. The software should also be able to identify two STR alleles of the same size but with different sequences, and it should be able to name the alleles according the nomenclature suggested by the ISFG. Figs. 4A and 5A show two examples where the heterozygous individual is called as homozygous because the primary report is too simplified, and the consequence is that manual intervention is required to analyze the results properly. The software should be able to identify SNP–STR variants by analyzing the flanking sequences for variants which is not possible with the current software solutions from Thermo Fisher Scientific and Illumina. For each SNP locus, the same user defined options should be available except for the acceptable stutter ratio and the length dependent allele balance which is not relevant for SNPs. In addition, there should be an option to define a maximum threshold of reads with base calls of the second known SNP allele in the case of a homozygous genotype call. For bi-allelic SNPs, the only way to identify mixtures is to look for allelic imbalances and this is an important quality assurance for SNP typing assays [9,10,76]. Fig. 5B shows an example where a mixture sample was called as homozygous even though 5% of the reads had sequences with the second allele and clearly indicated that the sample was heterozygous.

The future software should have a specific module for mixture interpretation that may be used once a sample has been identified as a mixture. NGS offers new possibilities for analysis of mixtures, and since the forensic community is only beginning to explore the use of NGS, such tools have not been developed. Future software modules should also be used to estimate bio-geographic ancestry, mtDNA haplogroups, Y-chromosome haplogroups, tissue identification and phenotypes. It will be important for the development of these modules that the forensic genetic community and the manufacturers of commercial kits engage in a close collaboration and that the software algorithms are well described in the user manual or scientific papers to simplify future accreditation attempts of NGS in forensic laboratories. Preferably, the analyses should apply to recommendations from the ISFG and similar forensic standardization bodies.

4.3. New frontiers in forensic genetics

Besides analysis of classical forensic markers, NGS makes it possible to expand forensic genetic investigations to new areas related to forensic medicine. When a person dies unexpectedly and for no apparent reason, shotgun or exome sequencing may identify genetic variants associated with known diseases and assist the pathologist in finding the cause of death. In Denmark, it is estimated that approximately 20% of all deaths are caused by sudden unexpected cardiac arrest, and one third of these death remain unexplained even after autopsy [86]. It is assumed that many of these individuals have a genetic disorder and that sequencing of selected genes may identify disease-related variants. Genetic testing will not only improve the diagnostic rate and add important information to research in cardiac diseases, it will also allow for identification of relatives with the same genetic disorder and initiation of treatment of these relatives. In a similar way, NGS may be used to screen for variants in genes that are involved in the metabolism of particular drugs and supplement toxicology investigations of a deceased in order to assess whether an unexpected death was accidental or premeditated [87,88]. It will also be possible to investigate DNA from bacteria, viruses, phages and fungi from the deceased either to identify disease causing microorganisms or to look for imbalances in the microbial communities which may give clues to the cause of death [89–91].

Sequencing of the microbiome in swabs or soil samples have demonstrated large differences in the different taxi found at different locations [92–94]. This may be used to find similarities between trace and reference samples. However, it should be emphasized that perfect matches or even exclusions are unlikely since the microbiome is constantly changing under the influence of environmental factors such as temperature, humidity, time of sampling, etc. Also, it was found that samples taken a few meters apart at the same time only shared 50% of the microbiome diversity. Nevertheless, the variation between sampling sites was much higher [94].

5. Concluding remarks

High throughput sequencing has accelerated research in many areas of biology and applied science. In the last few years, the use of NGS in forensic genetics has been debated and now, we are beginning to see applications directed specifically for human identification and determination of phenotypical traits. The advantages of NGS compared to the traditional PCR-CE methods are many, and there is little doubt that NGS will be implemented and used in forensic laboratories in the future. Prices of instruments and kits will determine how fast the transition from CE to NGS will be and how large a fraction of cases will be investigated by NGS. Development and validation of software solutions will be other critical aspects of the introduction of NGS into forensic genetics. Both commercial companies and forensic laboratories [63,70,72,73] have initiated the process, however, the current software solutions are not sufficiently advanced and more work

and collaboration between the companies and the forensic community is necessary.

Among the various platforms, the bench-top sequencers seem to be applicable for case work in terms of daily through-put, flexibility, run time and instrument cost. Currently, PCR based capture methods combined with sequencing by synthesis and semi-conductor sequencing are the most promising technologies. However, high throughput sequencing has evolved dramatically in the last decade and there is reason to believe that the development will continue. PCR may be replaced with probe capture methods if the sensitivity can be improved, and single-molecule sequencers may render NGS platforms obsolete in the coming years if new landmarks in signal detection can be developed and the base call error rates can be reduced to an acceptable level.

Acknowledgments

We thank Vania Pereira, Jill Olofsson, Jeppe D. Andersen and Marie-Louise Kampmann for helpful discussions.

References

[1] R. Higuchi, C.H. von Beroldingen, G.F. Sensabaugh, H.A. Erlich, DNA typing from single hairs, Nature 332 (1988) 543–546.
[2] C. Ginther, L. Issel-Tarver, M.C. King, Identifying individuals by sequencing mitochondrial DNA from teeth, Nat. Genet. 2 (1992) 135–138.
[3] K.M. Sullivan, R. Hopgood, P. Gill, Identification of human remains by amplification and automated sequencing of mitochondrial DNA, Int. J. Legal Med. 105 (1992) 83–86.
[4] W. Parson, A. Brandstatter, A. Alonso, N. Brandt, B. Brinkmann, A. Carracedo, D. Corach, O. Froment, I. Furac, T. Grzybowski, K. Hedberg, C. Keyser-Tracqui, T. Kupiec, S. Lutz-Bonengel, B. Mevag, R. Ploski, H. Schmitter, P. Schneider, D. Syndercombe-Court, E. Sorensen, H. Thew, G. Tully, R. Scheithauer, The EDNAP mitochondrial DNA population database (EMPOP) collaborative exercises: organisation, results and perspectives, Forensic Sci. Int. 139 (2004) 215–226.
[5] W. Parson, H.J. Bandelt, Extended guidelines for mtDNA typing of population data in forensic science, Forensic Sci. Int. Genet. 1 (2007) 13–19.
[6] F. Sanger, S. Nicklen, A.R. Coulson, DNA sequencing with chain-terminating inhibitors, Proc. Natl. Acad. Sci. U. S. A. 74 (1977) 5463–5467.
[7] K.R. Mitchelson, The use of capillary electrophoresis for DNA polymorphism analysis, Mol. Biotechnol. 24 (2003) 41–68.
[8] C. Børsting, N. Morling, Single nucleotide polymorphisms, in: J.A. Siegel, S. Pekka (Eds.), Encyclopedia of Forensic Sciences, Elsevier, Maryland, USA, 2013, pp. 233–239.
[9] C. Børsting, E. Rockenbauer, N. Morling, Validation of a single nucleotide polymorphism (SNP) typing assay with 49 SNPs for forensic genetic testing in a laboratory accredited according to the ISO 17025 standard, Forensic Sci. Int. Genet. 4 (2009) 34–42.
[10] C. Børsting, M. Mikkelsen, N. Morling, Kinship analysis with diallelic SNPs – experiences with the SNPforID Multiplex in an ISO17025 accreditated laboratory, Transfus. Med. Hemother. 39 (2012) 195–201.
[11] S. Walsh, A. Lindenbergh, S.B. Zuniga, T. Sijen, P. de Knijff, M. Kayser, K.N. Ballantyne, Developmental validation of the IrisPlex system: determination of blue and brown iris colour for forensic intelligence, Forensic Sci. Int. Genet. 5 (2011) 464–471.
[12] S. Walsh, L. Chaitanya, L. Clarisse, L. Wirken, J. Draus-Barini, L. Kovatsi, H. Maeda, T. Ishikawa, T. Sijen, P. de Knijff, W. Branicki, F. Liu, M. Kayser, Developmental validation of the HIrisPlex system: DNA-based eye and hair colour prediction for forensic and anthropological usage, Forensic Sci. Int. Genet. 9 (2014) 150–161.
[13] C. Phillips, A. Salas, J.J. Sanchez, M. Fondevila, A. Gomez-Tato, J. Alvarez-Dios, M. Calaza, M.C. de Cal, D. Ballard, M.V. Lareu, A. Carracedo, S.N. Consortium, Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs, Forensic Sci. Int. Genet. 1 (2007) 273–280.
[14] C. Phillips, L. Prieto, M. Fondevila, A. Salas, A. Gomez-Tato, J. Alvarez-Dios, A. Alonso, A. Blanco-Verea, M. Brion, M. Montesino, A. Carracedo, M.V. Lareu, Ancestry analysis in the 11-M Madrid bomb attack investigation, PLoS One 4 (2009) e6583.
[15] M. Ronaghi, S. Karamohamed, B. Pettersson, M. Uhlen, P. Nyren, Real-time DNA sequencing using detection of pyrophosphate release, Anal. Biochem. 242 (1996) 84–89.
[16] H. Andreasson, A. Asp, A. Alderborn, U. Gyllensten, M. Allen, Mitochondrial sequence analysis for forensic identification using pyrosequencing technology, Biotechniques 32 (2002) 124–126, 128, 130–133.
[17] H. Andreasson, M. Nilsson, B. Budowle, S. Frisk, M. Allen, Quantification of mtDNA mixtures in forensic evidence material using pyrosequencing, Int. J. Legal Med. 120 (2006) 383–390.
[18] A.M. Divne, H. Edlund, M. Allen, Forensic analysis of autosomal STR markers using pyrosequencing, Forensic Sci. Int. Genet. 4 (2010) 122–129.
[19] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L. Alenquer, T.P. Jarvie, K.B. Jirage, J.B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G. T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature 437 (2005) 376–380.
[20] D.A. Wheeler, M. Srinivasan, M. Egholm, Y. Shen, L. Chen, A. McGuire, W. He, Y.J. Chen, V. Makhijani, G.T. Roth, X. Gomes, K. Tartaro, F. Niazi, C.L. Turcotte, G.P. Irzyk, J.R. Lupski, C. Chinault, X.Z. Song, Y. Liu, Y. Yuan, L. Nazareth, X. Qin, D.M. Muzny, M. Margulies, G.M. Weinstock, R.A. Gibbs, J.M. Rothberg, The complete genome of an individual by massively parallel DNA sequencing, Nature 452 (2008) 872–876.
[21] K.V. Voelkerding, S.A. Dames, J.D. Durtschi, Next-generation sequencing: from basic research to diagnostics, Clin. Chem. 55 (2009) 641–658.
[22] I. Braslavsky, B. Hebert, E. Kartalov, S.R. Quake, Sequence information can be obtained from single DNA molecules, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 3960–3964.
[23] T.D. Harris, P.R. Buzby, H. Babcock, E. Beer, J. Bowers, I. Braslavsky, M. Causey, J. Colonell, J. Dimeo, J.W. Efcavitch, E. Giladi, J. Gill, J. Healy, M. Jarosz, D. Lapen, K. Moulton, S.R. Quake, K. Steinmann, E. Thayer, A. Tyurina, R. Ward, H. Weiss, Z. Xie, Single-molecule DNA sequencing of a viral genome, Science 320 (2008) 106–109.
[24] E.R. Mardis, Next-generation DNA sequencing methods, Annu. Rev. Genomics Hum. Genet. 9 (2008) 387–402.
[25] M.L. Metzker, Sequencing technologies – the next generation, Nat. Rev. Genet. 11 (2010) 31–46.
[26] J. Korlach, K.P. Bjornson, B.P. Chaudhuri, R.L. Cicero, B.A. Flusberg, J.J. Gray, D. Holden, R. Saxena, J. Wegener, S.W. Turner, Real-time DNA sequencing from single polymerase molecules, Methods Enzymol. 472 (2010) 431–455.
[27] R. Nielsen, J.S. Paul, A. Albrechtsen, Y.S. Song, Genotype and SNP calling from next-generation sequencing data, Nat. Rev. Genet. 12 (2011) 443–451.
[28] A. Altmann, P. Weber, D. Bader, M. Preuss, E.B. Binder, B. Muller-Myhsok, A beginners guide to SNP calling from high-throughput DNA-sequencing data, Hum. Genet. 131 (2012) 1541–1554.
[29] F. Ozsolak, Third-generation sequencing techniques and applications to drug discovery, Expert Opin. Drug Discov. 7 (2012) 231–243.
[30] F. Haque, J. Li, H.C. Wu, X.J. Liang, P. Guo, Solid-state and biological nanopore for real-time sensing of single chemical and sequencing of DNA, Nano Today 8 (2013) 56–74.
[31] D. Goldman, K. Domschke, Making sense of deep sequencing, Int. J. Neuropsychopharmacol. (2014) 1–9.
[32] N. Cloonan, S.M. Grimmond, Transcriptome content and dynamics at single-nucleotide resolution, Genome Biol. 9 (2008) 234.
[33] M. Yano, T. Ohtsuka, H. Okano, RNA-binding protein research with transcriptome-wide technologies in neural development, Cell Tissue Res. 359 (2014) 135–144.
[34] L. Mamanova, A.J. Coffey, C.E. Scott, I. Kozarewa, E.H. Turner, A. Kumar, E. Howard, J. Shendure, D.J. Turner, Target-enrichment strategies for next-generation sequencing, Nat. Methods 7 (2010) 111–118.
[35] X. Liu, J. Wang, L. Chen, Whole-exome sequencing reveals recurrent somatic mutation networks in cancer, Cancer Lett. 340 (2013) 270–276.
[36] I.S. Hagemann, C.E. Cottrell, C.M. Lockwood, Design of targeted capture-based, next generation sequencing tests for precision cancer therapy, Cancer Genet. 206 (2013) 420–431.
[37] D. Dressman, H. Yan, G. Traverso, K.W. Kinzler, B. Vogelstein, Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations, Proc. Natl. Acad. Sci. U. S. A. 100 (2003) 8817–8822.
[38] M. Fedurco, A. Romieu, S. Williams, I. Lawrence, G. Turcatti, BTA a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies, Nucleic Acids Res. 34 (2006) e22.
[39] P.M. Lundquist, C.F. Zhong, P. Zhao, A.B. Tomaney, P.S. Peluso, J. Dixon, B. Bettman, Y. Lacroix, D.P. Kwo, E. McCullough, M. Maxham, K. Hester, P. McNitt, D.M. Grey, C. Henriquez, M. Foquet, S.W. Turner, D. Zaccarin, Parallel confocal detection of single molecules in real time, Opt. Lett. 33 (2008) 1026–1028.
[40] J. Eid, A. Fehr, J. Gray, K. Luong, J. Lyle, G. Otto, P. Peluso, D. Rank, P. Baybayan, B. Bettman, A. Bibillo, K. Bjornson, B. Chaudhuri, F. Christians, R. Cicero, S. Clark, R. Dalal, A. Dewinter, J. Dixon, M. Foquet, A. Gaertner, P. Hardenbol, C. Heiner, K. Hester, D. Holden, G. Kearns, X. Kong, R. Kuse, Y. Lacroix, S. Lin, P. Lundquist, C. Ma, P. Marks, M. Maxham, D. Murphy, I. Park, T. Pham, M. Phillips, J. Roy, R. Sebra, G. Shen, J. Sorenson, A. Tomaney, K. Travers, M. Trulson, J. Vieceli, J. Wegener, D. Wu, A. Yang, D. Zaccarin, P. Zhao, F. Zhong, J. Korlach, S. Turner, Real-time DNA sequencing from single polymerase molecules, Science 323 (2009) 133–138.
[41] D.H. Stoloff, M. Wanunu, Recent trends in nanopores for biotechnology, Curr. Opin. Biotechnol. 24 (2013) 699–704.
[42] L. Liang, Q. Wang, H. Agren, Y. Tu, Computational studies of DNA sequencing with solid-state nanopores: key issues and future prospects, Front. Chem. 2 (2014) 1–4.
[43] A.S. Mikheyev, M.M. Tin, A first look at the Oxford Nanopore MinION sequencer, Mol. Ecol. Resour. 14 (2014) 1097–1102.

[44] C.S. Chin, D.H. Alexander, P. Marks, A.A. Klammer, J. Drake, C. Heiner, A. Clum, A. Copeland, J. Huddleston, E.E. Eichler, S.W. Turner, J. Korlach, Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data, Nat. Methods 10 (2013) 563–569.
[45] J. Huddleston, S. Ranade, M. Malig, F. Antonacci, M. Chaisson, L. Hon, P.H. Sudmant, T.A. Graves, C. Alkan, M.Y. Dennis, R.K. Wilson, S.W. Turner, J. Korlach, E.E. Eichler, Reconstructing complex regions of genomes using long-read sequencing technology, Genome Res. 24 (2014) 688–696.
[46] B.A. Flusberg, D.R. Webster, J.H. Lee, K.J. Travers, E.C. Olivares, T.A. Clark, J. Korlach, S.W. Turner, Direct detection of DNA methylation during single-molecule, real-time sequencing, Nat. Methods 7 (2010) 461–465.
[47] E.V. Wallace, D. Stoddart, A.J. Heron, E. Mikhailova, G. Maglia, T.J. Donohoe, H. Bayley, Identification of epigenetic DNA modifications with a protein nanopore, Chem. Commun. (Camb.) 46 (2010) 8195–8197.
[48] T.A. Clark, I.A. Murray, R.D. Morgan, A.O. Kislyuk, K.E. Spittle, M. Boitano, A. Fomenkov, R.J. Roberts, J. Korlach, Characterization of DNA methyltransferase specificities using single-molecule, real-time DNA sequencing, Nucleic Acids Res. 40 (2012) e29.
[49] J. Schreiber, Z.L. Wescoe, R. Abu-Shumays, J.T. Vivian, B. Baatar, K. Karplus, M. Akeson, Error rates for nanopore discrimination among cytosine, methylcytosine, and hydroxymethylcytosine along individual DNA strands, Proc. Natl. Acad. Sci. U. S. A. 110 (2013) 18910–18915.
[50] M. Nothnagel, A. Herrmann, A. Wolf, S. Schreiber, M. Platzer, R. Siebert, M. Krawczak, J. Hampe, Technology-specific error signatures in the 1000 Genomes Project data, Hum. Genet. 130 (2011) 505–516.
[51] H.Y. Lam, M.J. Clark, R. Chen, R. Chen, G. Natsoulis, M. O'Huallachain, F.E. Dewey, L. Habegger, E.A. Ashley, M.B. Gerstein, A.J. Butte, H.P. Ji, M. Snyder, Performance comparison of whole-genome sequencing platforms, Nat. Biotechnol. 30 (2012) 78–82.
[52] A. Ratan, W. Miller, J. Guillory, J. Stinson, S. Seshagiri, S.C. Schuster, Comparison of sequencing platforms for single nucleotide variant calls in a human sample, PLoS One 8 (2013) e55089.
[53] N. Rieber, M. Zapatka, B. Lasitschka, D. Jones, P. Northcott, B. Hutter, N. Jager, M. Kool, M. Taylor, P. Lichter, S. Pfister, S. Wolf, B. Brors, R. Eils, Coverage bias and sensitivity of variant calling for four whole-genome sequencing technologies, PLoS One 8 (2013) e66621.
[54] H.L. Rehm, Disease-targeted sequencing: a cornerstone in the clinic, Nat. Rev. Genet. 14 (2013) 295–300.
[55] C. Beadling, T.L. Neff, M.C. Heinrich, K. Rhodes, M. Thornton, J. Leamon, M. Andersen, C.L. Corless, Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping, J. Mol. Diagn. 15 (2013) 171–176.
[56] S.Q. Wong, J. Li, A.Y. Tan, R. Vedururu, J.M. Pang, H. Do, J. Ellul, K. Doig, A. Bell, G. A. MacArthur, S.B. Fox, D.M. Thomas, A. Fellowes, J.P. Parisot, A. Dobrovic, C. Cohort, Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing, BMC Med. Genomics 7 (2014) 23.
[57] L. Castera, S. Krieger, A. Rousselin, A. Legros, J.J. Baumann, O. Bruet, B. Brault, R. Fouillet, N. Goardon, O. Letac, S. Baert-Desurmont, J. Tinat, O. Bera, C. Dugast, P. Berthet, F. Polycarpe, V. Layet, A. Hardouin, T. Frebourg, D. Vaur, Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes, Eur. J. Hum. Genet. 22 (2014) 1305–1311.
[58] G. Millat, V. Chanavat, R. Rousson, Evaluation of a new high-throughput next-generation sequencing method based on a custom AmpliSeq Library and Ion Torrent PGM Sequencing for the rapid detection of genetic variations in long QT syndrome, Mol. Diagn. Ther. 18 (2014) 533–539.
[59] J. Tarabeux, B. Zeitouni, V. Moncoutier, H. Tenreiro, K. Abidallah, S. Lair, P. Legoix-Ne, Q. Leroy, E. Rouleau, L. Golmard, E. Barillot, M.H. Stern, T. Rio-Frio, D. Stoppa-Lyonnet, C. Houdayer, Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model, Eur. J. Hum. Genet. 22 (2014) 535–541.
[60] G.J. Tsongalis, J.D. Peterson, F.B. de Abreu, C.D. Tunkey, T.L. Gallagher, L.D. Strausbaugh, W.A. Wells, C.I. Amos, Routine use of the Ion Torrent AmpliSeq cancer hotspot panel for identification of clinically actionable somatic mutations, Clin. Chem. Lab. Med. 52 (2014) 707–714.
[61] T.J. Treangen, S.L. Salzberg, Repetitive DNA and next-generation sequencing: computational challenges and solutions, Nat. Rev. Genet. 13 (2012) 36–46.
[62] S.L. Fordyce, M.C. Avila-Arcos, E. Rockenbauer, C. Børsting, R. Frank-Hansen, F.T. Petersen, E. Willerslev, A.J. Hansen, N. Morling, M.T. Gilbert, High-throughput sequencing of core STR loci for forensic genetic investigations using the Roche Genome Sequencer FLX platform, Biotechniques 51 (2011) 127–133.
[63] C. Van Neste, F. Van Nieuwerburgh, D. Van Hoofstat, D. Deforce, Forensic STR analysis using massive parallel sequencing, Forensic Sci. Int. Genet. 6 (2012) 810–818.
[64] E. Rockenbauer, S. Hansen, M. Mikkelsen, C. Børsting, N. Morling, Characterization of mutations and sequence variants in the D21S11 locus by next generation sequencing, Forensic Sci. Int. Genet. 8 (2014) 68–72.
[65] S. Dalsgaard, E. Rockenbauer, A. Buchard, H.S. Mogensen, R. Frank-Hansen, C. Børsting, N. Morling, Non-uniform phenotyping of D12S391 resolved by second generation sequencing, Forensic Sci. Int. Genet. 8 (2014) 195–199.
[66] C. Gelardi, E. Rockenbauer, S. Dalsgaard, C. Børsting, N. Morling, Second generation sequencing of three STRs D3S1358, D12S391 and D21S11 in Danes and a new nomenclature for sequenced STR alleles, Forensic Sci. Int. Genet. 12 (2014) 38–41.
[67] C. Tomas, H.S. Mogensen, S.L. Friis, C. Hallenberg, M.C. Stene, N. Morling, Concordance study and population frequencies for 16 autosomal STRs analyzed with PowerPlex(R) ESI 17 and AmpFlSTR(R) NGM SElect in Somalis, Danes and Greenlanders, Forensic Sci. Int. Genet. 11 (2014) e18–e21.
[68] A.A. Westen, T. Kraaijenbrink, E.A. Robles de Medina, J. Harteveld, P. Willemse, S.B. Zuniga, K.J. van der Gaag, N.E. Weiler, J. Warnaar, M. Kayser, T. Sijen, P. de Knijff, Comparing six commercial autosomal STR kits in a large Dutch population sample, Forensic Sci. Int. Genet. 10 (2014) 55–63.
[69] M. Scheible, O. Loreille, R. Just, J. Irwin, Short tandem repeat typing on the 454 platform: strategies and considerations for targeted sequencing of common forensic markers, Forensic Sci. Int. Genet. 12 (2014) 107–119.
[70] S.Y. Anvar, K.J. van der Gaag, J.W. van der Heijden, M.H. Veltrop, R.H. Vossen, R. H. de Leeuw, C. Breukel, H.P. Buermans, J.S. Verbeek, P. de Knijff, J.T. den Dunnen, J.F. Laros, TSSV: a tool for characterization of complex allelic variants in pure and mixed genomes, Bioinformatics 30 (2014) 1651–1659.
[71] D.M. Bornman, M.E. Hester, J.M. Schuetter, M.D. Kasoji, A. Minard-Smith, C.A. Barden, S.C. Nelson, G.D. Godbold, C.H. Baker, B. Yang, J.E. Walther, I.E. Tornes, P. S. Yan, B. Rodriguez, R. Bundschuh, M.L. Dickens, B.A. Young, S.A. Faith, Short-read, high-throughput sequencing technology for STR genotyping, Biotechniques 1–6 (2012) , doi:http://dx.doi.org/10.2144/000113857.
[72] D.H. Warshauer, D. Lin, K. Hari, R. Jain, C. Davis, B. Larue, J.L. King, B. Budowle, STRait Razor: a length-based forensic STR allele-calling tool for use with second generation sequencing data, Forensic Sci. Int. Genet. 7 (2013) 409–417.
[73] C. Van Neste, M. Vandewoestyne, W. Van Criekinge, D. Deforce, F. Van Nieuwerburgh, My-Forensic-Loci-queries (MyFLq) framework for analysis of forensic STR data generated by massive parallel sequencing, Forensic Sci. Int. Genet. 9 (2014) 1–8.
[74] S.L. Fordyce, H.S. Mogensen, C. Borsting, R.E. Lagace, C.W. Chang, N. Rajagopalan, N. Morling, Second-generation sequencing of forensic STRs using the Ion Torrent HID STR 10-plex and the Ion PGM, Forensic Sci. Int. Genet. 14 (2015) 132–140.
[75] S. Dalsgaard, E. Rockenbauer, C. Gelardi, C. Børsting, S.L. Fordyce, N. Morling, Characterization of mutations and sequence variations in complex STR loci by second generation sequencing, Forensic Sci. Int. Genet. Suppl. 4 (2013) e218–e219.
[76] C. Børsting, S.L. Fordyce, J. Olofsson, H.S. Mogensen, N. Morling, Evaluation of the Ion Torrent HID SNP 169-plex: a SNP typing assay developed for human identification by second generation sequencing, Forensic Sci. Int. Genet. 12 (2014) 144–154.
[77] S.B. Seo, J.L. King, D.H. Warshauer, C.P. Davis, J. Ge, B. Budowle, Single nucleotide polymorphism typing with massively parallel sequencing for human identification, Int. J. Legal Med. 127 (2013) 1079–1086.
[78] J.J. Sanchez, C. Phillips, C. Børsting, K. Balogh, M. Bogus, M. Fondevila, C.D. Harrison, E. Musgrave-Brown, A. Salas, D. Syndercombe-Court, P.M. Schneider, A. Carracedo, N. Morling, A multiplex assay with 52 single nucleotide polymorphisms for human identification, Electrophoresis 27 (2006) 1713–1724.
[79] A.J. Pakstis, W.C. Speed, R. Fang, F.C. Hyland, M.R. Furtado, J.R. Kidd, K.K. Kidd, SNPs for a universal individual identification panel, Hum. Genet. 127 (2010) 315–324.
[80] R. Nassir, R. Kosoy, C. Tian, P.A. White, L.M. Butler, G. Silva, R. Kittles, M.E. Alarcon-Riquelme, P.K. Gregersen, J.W. Belmont, F.M. De La Vega, M.F. Seldin, An ancestry informative marker set for determining continental origin: validation and extension using human genome diversity panels, BMC Genet. 10 (2009) 39.
[81] C.M. Nievergelt, A.X. Maihofer, T. Shekhtman, O. Libiger, X. Wang, K.K. Kidd, J.R. Kidd, Inference of human continental origin and admixture proportions using a highly discriminative ancestry informative 41-SNP panel, Investig. Genet. 4 (2013) 13.
[82] S. Walsh, F. Liu, A. Wollstein, L. Kovatsi, A. Ralf, A. Kosiniak-Kamysz, W. Branicki, M. Kayser, The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA, Forensic Sci. Int. Genet. 7 (2013) 98–115.
[83] L. Poulsen, S.L. Friis, C. Hallenberg, B.T. Simonsen, N. Morling, A report of the 2009–2011 paternity and relationship testing workshops of the English Speaking Working Group of the International Society for Forensic Genetics, Forensic Sci. Int. Genet. 9 (2014) e1–e-2.
[84] A.R. Thomsen, C. Hallenberg, B.T. Simonsen, R.B. Langkjaer, N. Morling, A report of the 2002–2008 paternity testing workshops of the English Speaking Working Group of the International Society for Forensic Genetics, Forensic Sci. Int. Genet. 3 (2009) 214–221.
[85] D.W. Gjertson, C.H. Brenner, M.P. Baur, A. Carracedo, F. Guidet, J.A. Luque, R. Lessig, W.R. Mayr, V.L. Pascali, M. Prinz, P.M. Schneider, N. Morling, ISFG: recommendations on biostatistics in paternity testing, Forensic Sci. Int. Genet. 1 (2007) 223–231.
[86] C.L. Hertz, L. Ferrero-Miliani, R. Frank-Hansen, N. Morling, H. Bundgaard, A comparison of genetic findings in sudden cardiac death victims and cardiac patients: the importance of phenotypic classification, Europace (2014) , doi: http://dx.doi.org/10.1093/europace/euu210.
[87] L.E. Visser, R.H. van Schaik, M. van Vliet, P.H. Trienekens, P.A. De Smet, A.G. Vulto, A. Hofman, C.M. van Duijn, B.H. Stricker, Allelic variants of cytochrome P450 2C9 modify the interaction between nonsteroidal anti-inflammatory drugs and coumarin anticoagulants, Clin. Pharmacol. Ther. 77 (2005) 479–485.
[88] D. Sullivan, J.K. Pinsonneault, A.C. Papp, H. Zhu, S. Lemeshow, D.C. Mash, W. Sadee, Dopamine transporter DAT and receptor DRD2 variants affect risk of lethal cocaine abuse: a gene-gene-environment interaction, Transl. Psychiatry 3 (2013) e222.
[89] M.J. Cox, W.O. Cookson, M.F. Moffatt, Sequencing the human microbiome in health and disease, Hum. Mol. Genet. 22 (2013) R88–R94.

[90] M.E. Quinones-Mateu, S. Avila, G. Reyes-Teran, M.A. Martinez, Deep sequencing: becoming a critical tool in clinical virology, J. Clin. Virol. 61 (2014) 9–19.
[91] K.G. Frey, J.E. Herrera-Galeano, C.L. Redden, T.V. Luu, S.L. Servetas, A.J. Mateczun, V.P. Mokashi, K.A. Bishop-Lilly, Comparison of three next-generation sequencing platforms for metagenomic sequencing and identification of pathogens in blood, BMC Genomics 15 (2014) 96.
[92] S.G. Tringe, E.M. Rubin, Metagenomics: DNA sequencing of environmental samples, Nat. Rev. Genet. 6 (2005) 805–814.
[93] S. Giampaoli, A. Berti, R.M. Di Maggio, E. Pilli, A. Valentini, F. Valeriani, G. Gianfranceschi, F. Barni, L. Ripani, V. Romano Spica, The environmental biological signature: NGS profiling for forensic comparison of soils, Forensic Sci. Int. 240 (2014) 41–47.
[94] J.M. Young, L.S. Weyrich, A. Cooper, Forensic soil DNA analysis using high-throughput sequencing: a comparison of four molecular markers, Forensic Sci. Int. Genet. 13 (2014) 176–184.