

Implications of Targeted Next Generation Sequencing in Forensic Science

Sobiah R, Syeda RH, Zunaira E, Nageen Z, Maria K, Syeda AZ, Shahana SM, Akifa M, Abdul J and Muhammad RK*

Genome Editing and Sequencing Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan

*Corresponding author: Muhammad Ramzan Khan, Genome Editing and Sequencing Lab, National Center for Bioinformatics, Quaid-i-Azam University, Islamabad, Pakistan, Tel: +92-051 9064-4190; E-mail: mrkhan@qau.edu.pk

Received date: December 01, 2017; Accepted date: March 6, 2018; Published date: March 12, 2018

Copyright: © 2018 Sobiah R, et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

The outburst of next generation sequencing has outpaced the traditional capillary electrophoresis (CE) based forensic genomics in terms of throughput, scalability and allelic resolutions. The limitations of CE-based detection system for STRs (Short Tandem Repeats) and SNPs (Single Nucleotide Polymorphisms) markers associated with forensic DNA phenotyping have been greatly overcome through next generation sequencing (NGS). It offers simultaneous analysis of forensically relevant genetic markers including STRs, SNPs, mutations and transcripts to improve efficiency, capacity and resolution through massively parallel sequencing. The detailed sequence information in comparison with ever growing DNA databases across the world may aid mixture interpretation, and will ensure enhancement in statistical weight of the evidence.

Nowadays, a total of 33 NGS machines with low to high throughput are available but two dedicated systems including MiSeq® FGx™ Forensic Genomics System (Illumina) and the HID-Ion Personal Genome Machine (PGM)™ (Thermo Fisher Scientific) offer the sequencing of customized forensically informative SNP panels. Ion torrent launches Ion S5 and Ion S5 XL (with additional feature of local computing which enhances its analysis speed) which include same core instrument. The future of forensic genomics lies with integrated high throughput approaches to solve the increasing number of cases or mass disaster studies where forensic specimens and samples are compromised and degraded. In future new real-time sequencing platforms will revolutionize the field of forensic genomics.

Keywords: Forensic targeted sequencing; Single nucleotide polymorphism (SNP); Capillary electrophoresis (CE); Forensic DNA phenotyping (FDP); Genomics

Introduction

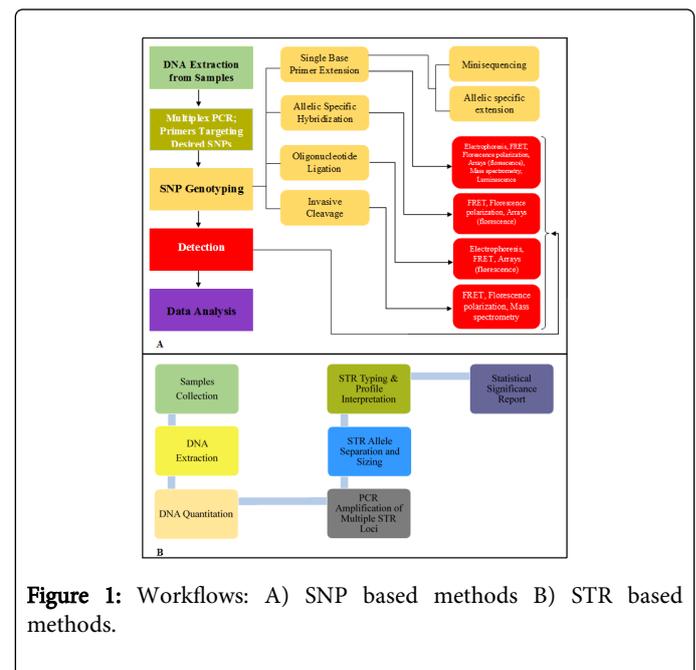
Type of high-throughput sequencing technology which is non-Sanger-based is called next generation sequencing (NGS) technology. Simultaneous sequencing of millions/billions of DNA molecules significantly increases throughput and reduces requirement for fragment-cloning approach which is employed in Sanger sequencing.

NGS involves analysis of a large number of samples simultaneously through loop array sequencing and determination of the base composition of single DNA molecules by second generation and third-generation sequencing technology, respectively [1]. NGS relies on preparation of NGS libraries in a cell-free system rather than bacterial cloning of DNA fragments.

Secondly it has the ability to parallelize thousands to millions rather hundreds of sequencing reaction; and it also detects output of sequencing directly without requirement of electrophoresis. It enables sequencing of entire genomes due to its power to generate vast number of reads at an unmatched speed. These attributes make NGS a need in different areas of life sciences, and therefore it is widely in use [2].

DNA analysis provides basic foundation for contemporary forensic research. Genetic variations exist in human populations in terms of factors such as length and sequence. These forms of variation make possible forensic DNA testing because many different alleles can exist in non-coding regions of the genome. When information from

multiple unlinked genetic markers is combined, higher powers of discrimination are possible. Over the past decade, the primary way for forensic DNA profiling is through length variation; in the terms of markers such as short tandem repeat (STR) [3,4] (Figure 1).



Millions of STR profiles based on a few core STR markers are now contained in large national DNA databases [5-7]. STR analysis is likely to remain the most important and commonly-used genetic technique in forensic science for the foreseeable future. Nevertheless, NGS technology has many potential advantages for STR analysis. These include high throughput, low cost, simultaneous detection of large numbers of STR loci on both autosomes and sex chromosomes, and the ability to distinguish alleles with similar length or digital read count. Over the last more than 20 years, the STRs markers have served as gold standard choice for human identification [8,9]. As SNPs amplifies a small target region hence have ability to work excellent on degraded low quality DNA and compared to STRs, it has lower mutation rates which assist in kinship testing. In future these SNPs have a potential part in helping investigators through predicting ancestry or phenotypic attributes [10-13] (Figure 1).

Literature Review

Implication of NGS in forensics research

High throughput sequencing has brought fruitful enrichment in research covering many fields of biological and applied sciences. The use of NGS in forensic genetics brought applications mainly in human identification and determination of phenotypic attributes. NGS technology would therefore significantly facilitate in a diverse range of areas including the identification of mixed DNA samples and analysis of complex paternity cases, ultimately greatly increasing the efficiency and cost-effectiveness of legal cases. A sketch has been shown in Figure 2 to describe the diversity of NGS applications in forensic DNA research.

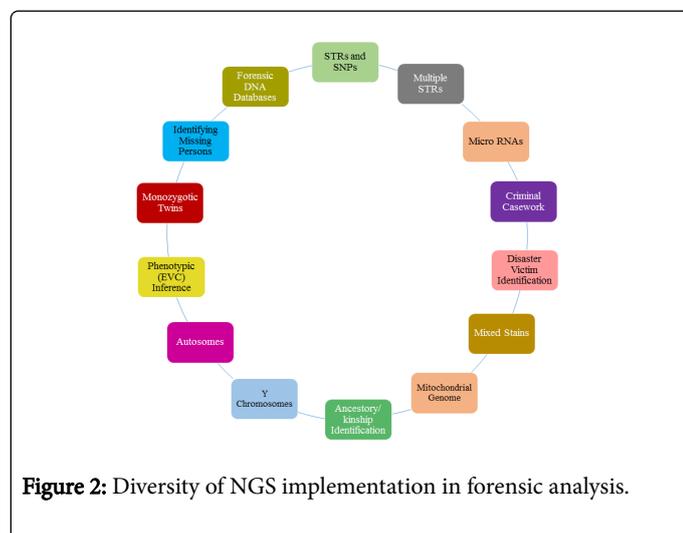


Figure 2: Diversity of NGS implementation in forensic analysis.

Since one decade, forensic DNA phenotyping (FDP) that exploits SNP genetic markers associated with bio-geographical ancestry (BGA) and externally visible characters (EVCs) is the yard stick for criminal investigation when power of STRs is limited [14]. The merits of SNPs over STRs are; lower mutation rates, abundance in genome, short amplicon length, multiplexicity, analysis of degraded DNA, high throughput genotyping, amenability for forensic applications etc. [13,15,16]. The forensic genotyping using SNaPshot[®] assay implicates single base extension (SBE) with capillary electrophoresis (CE)-detection system which is a method of choice for SNPs [16,17-20]. A number of SNP-based SNaPshot[®] assays including SNP for ID 34-plex,

Eurasiaplex, Irisplex and Hirisplex etc. [21-25]. In certain cases of forensic genomics Sanger sequencing has also been used to unveil the specific regions of mitochondrial genome based on CE-detections. However, limitations of STRs in FDP and even CE-based SNPs detection in terms of throughput, scalability and resolution have strained their uses in FDP intelligence [26,27].

Recently, there is dramatic outburst of data in terms of number of samples but the CE-based methods remained static. After the advent of next generation sequencing a decade ago, the forensic genomics has benefitted a lot [2]. The establishment of national DNA databases and their facilitation in investigations and suspect identification have persuaded more and more countries to set up forensic databases. Consequently there is a huge demand of DNA intelligences in solving criminal cases through forensic genomics. The scope of forensic genomics has been broadened by its crucial role in DNA-based investigation of the identification of missing persons, kinship testing, ancestry investigations, and other human identification applications. This increasing demand of DNA-based test cases and broader applications of forensic genomics have strains the limited scale capabilities of CE-based conventional methods. Therefore there is need for low cost high throughput sequencing in forensic genomics.

High throughput sequencing or massively parallel sequencing (MPS) in forensic genomics can genotype hundreds of markers in multiple samples using small quantity of DNA. Different platforms available to date include, Roche, Ion torrent (ThermoFisher), Illumina, Solid, PacBio, Helicose, Oxford Nanopore, Qiagen and BGI etc. [28]. Two of the MPS platforms that are currently available are the MiSeq[®] FGx[™] Forensic Genomics System [29] (Illumina) and the HID-Ion Personal Genome Machine (PGM)[™] (Thermo Fisher Scientific) [30,31] which are coupled with the ForeSeq[™] DNA Signature Prep kit (Illumina) [29] and the HID-Ion AmpliSeq[™] Identity Panel (Thermo Fisher Scientific), respectively [32]. The name of the panel HID-Ion AmpliSeq[™] Identity Panel changed to "Precision ID Identity Panel" in May 2016 [33]. The work flow with little variations includes sample preparation, cluster generation, DNA sequencing and imaging and finally data analysis. These platforms utilize sequencing by synthesis (SBS) technology for parallel sequencing of PCR amplicons and allow forensic scientists worldwide to harness the full power of NGS. With attributes like deep sequencing, maximum coverage, the highest yield of error-free reads, best performance in repetitive sequence regions, and lowest base-by-base price, Illumina SBS sequencing is the most widely adopted chemistry in the industry [34,35]. Using contemporary NGS systems, examiners can generate data that span the entire genome and address a wider range of questions in a single, targeted assay. Moreover, NGS-generated STR allele calls are fully compatible with current database formats, providing integrative seamless link between CE-based and NGS data. With the use of NGS it will be possible to achieve the simultaneous analysis of the standard autosomal DNA (STRs and SNPs), mitochondrial DNA, and X and Y chromosomal markers [36]. Early studies are providing a promising view of the advantages NGS brings to the analysis of even the smallest, most compromised, and highly mixed evidentiary samples. Thus, MPS is important for human DNA genotyping in cases like mass disasters or other events where forensic specimens and samples are compromised and degraded. Additionally, micro-biome typification could be an interesting application to study for crime scene characterization [36]. New NGS platforms are in pipelines with real-time sequencing will revolutionize forensic genomics with speedy identification of donor individual.

MiSeq FGx™ (Illumina) system

The MiSeq FGx™ (Illumina) system along with the ForenSeq™ DNA signature panel (Illumina) is commercial MPS product available for operational forensic purposes. This multiplex kit consists of more than 200 markers including 27 autosomal STRs, 24 Y-STRs, 7 X-STRs, 94 identity SNPs, 56 ancestry SNPs, 24 phenotypic SNPs and Amelogenin. The ForenSeq™ workflow (Figure 3) involves template and library preparation, sequencing and the analysis of sequence data using Illumina's Universal Analysis Software (UAS) [37,38]. Several manufacturers have designed MPS kits for forensic testing that include the Combined DNA Index System (CODIS) STRs that are used in the United States for criminal investigations. These kits often include additional autosomal X and Y chromosome STRs, as well as identity, ancestry, and phenotypic SNPs (iSNPs, aSNPs, pSNPs). The latter, in combination with complete MPS mitochondrial sequencing kits, could add important information for the identification of missing persons when only partial or skeletal remains are available. These kits were designed to be utilized with the major MPS sequencing platforms currently in use: Illumina's MiSeq (San Diego, CA) and Thermo Fisher's Ion PGM and S5 (Waltham, MA). A study by Li [39] described an application of MPS to paternity testing on the Ion torrent PGM platform. With more STR loci simultaneously detected by MPS, it is more conducive for distinguishing between mutations and exclusions. Meanwhile, because MPS technology identifies STR loci based on sequence so that repeat motif variation within STRs could be detected, it may help one to infer the origin of the mutation in some cases, and the discrimination power could be increased.

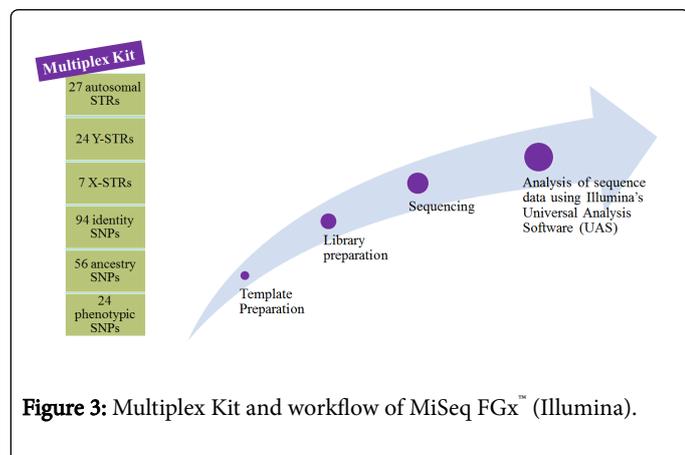


Figure 3: Multiplex Kit and workflow of MiSeq FGx™ (Illumina).

Limitations of STRs and SNP based Identification

Though CE-based assays including STRs and SNP genotyping remained the method of choice in forensic research over the past two decades [9,16], but they also bear limitations in throughput, scalability and allelic resolution just name but a few. One of the most undesirable limitations of the CE-based detection is multiplexing incapacity. A maximum of 24 loci can be PCR amplified and detected resulting in size and overlap constraints of fluorescently tagged amplicons [26,27]. This inability of sample multiplexing leads to individual processing of samples. Consequently the work is not only tedious but also error prone. There is an absolute obscurity in polymorphic allelic capture harbouring substitution variants as CE-based analysis with length-based typing cannot capture sequence differences present in alleles of the same length. Hence even polymorphic samples cannot be detected out of mixture which limits

the resolution of technology. In case of compromised and degraded samples the sometime it may not be feasible to generate the required DNA data for forensic casework samples. Even iterative sampling will result in complete depletion of samples and desirable results may not be achieved to solve a particular case. Another important limitation of the STRs and SNP methods is the genetic dead end of the investigation in case there is no database hit while doing genome wide association studies in case of exclusion mapping.

Advantages of high throughput targeted sequencing in forensic genomics

Whole-genome sequencing (WGS) manifests complete blue print of allelic variations in genome structure of the individuals but analysis and management of huge genomic data is a cumbersome. Therefore, targeted sequencing of relevant loci seems to be the best strategy for forensic genomics research. It is a high frequency HTS approach for screening variations within target regions in populations through extraction and sequencing these genomic targets from a sample library. Targeted sequencing has emancipated the genotyping procedure from CE-based fragment length detection [40,41]. As already mentioned NGS-based forensic typing [37,42] based on SBS chemistry implicates SBS cyclic reversible termination chemistry [43], a major advantage associated with this synthesis is low incorporation bias, removal of error and missed base calls in the homo-polymers and repetitive regions [44].

The detection of Intra-STR SNPs is a direct benefit of NGS application in forensic genomics. It has the resolving power to differentiate or sort out the alleles that are identical in length but different in sequence. This hold true for intra-STR SNPs, SNPs for mt-DNA, and the whole sequence of signals that are unexpected as well as evaluation of data artifacts at the level of nucleotide facilitating a robust and precise strategy for casework and applications in human identification. The MPS is supplemented with software that has a user friendly interface, thus facilitating easy and simple visualization of intra-STR SNPs data. With MPS, multiple polymorphisms can be analyzed simultaneously now, which was a prominent constraint of CE-based genotyping [37,45-47]. Analysis of degraded samples has improved due to short read length of the amplicons. The phenotypic study of externally visible characteristics such as biogeographic ancestry, hair and eye color is also getting better accompanying with genotyping [38]. The compatibility of MPS with STRs and databases not only maintain standard allele nomenclature but also enable consolidation of overlapping marker set in use across the world. A prominent merit of MPS is the higher sensitivity with digital data instead of analog metrics for example color of peak, shape, size, and height resulted in CE-based system. The data generated by NGS platforms in quantitative in the form of discrete read counts. Further sensitivity can be increased by coverage level and in depth sequencing. This helps in revealing the minor contribution of donor individual in a complex mixture. This is not possible with CE-based methods. A simple example is that in case of mt-DNA heteroplasmy examination, NGS deep sequencing has the ability to recognize minor alteration frequencies of approximately 1% of the major in comparison to >10-20% minor mutation frequencies resulted from CE-based sequencing. NGS allows library multiplexing to assist scalable throughput at an altitude which is not possible with alternative CE methods. This is because; NGS performs this multiplexing by attachment of distinctive barcodes (index sequences). Forensic libraries can be merged in a controlled way to sequence up to 96

libraries at a time. This can be done using ForenSeq DNA Signature Prep Kit (384 potential) or 384 libraries with the Nextera XT DNA Library Prep Kit.

MPS has the ability to investigate familial relationships and personal identification using X and Y STRs, without iterative testing. This reduces consumption of DNA samples and the need for deciding which assay to implement, as all classes of forensically significant loci is amplified in one forensic NGS multiplex. Forensic scientists over the globe are actively investigating and implementing NGS systems for forensic genomics due to their ability to improve DNA typing [42,44,48]. By delivering targeted data on forensically significant loci, a broad span of queries can be responded in a single assay by forensic scientists.

NGS platforms for targeted sequencing

Presently, a number of companies have offered NGS systems. There are around 33 different machines in the market with low to high throughput capacity [28]. These include Roche, Ion torrent (ThermoFisher), Illumina, Solid, PacBio, Helicose, Oxford nanopore, Qiagen, BGI etc. At present available MPS platforms includes MiSeq[®] FGx[™] Forensic genomics system (Illumina) and HID-Ion personal genome machine (PGM)[™] (Thermo fisher scientific) which are integrated with the ForenSeq[™] DNA signature prep kit (Illumina) and the HID-Ion ampliSeq[™] Identity panel (Thermo fisher scientific) respectively [32].

Illumina is the widely spread accepted platform in the world. It uses sequencing by synthesis chemistry with cyclic reversible termination method [43]. MiSeq FGx[™] Forensic genomics system, a dedicated system for forensic genomics, is a completely integrated, DNA-to-data program, which offers library preparation, resource for DNA sequencing, and data analysis software. The MiSeq FGx System is the present available fully integrated workflow. Over 200 forensically relevant genetic markers are included in a single, efficient workflow ForenSeq[™] DNA Signature Prep Kit hence removing multiple STR kits requirement [49]. The kit not only integrates the autosomal STR markers which are in use around the world for casework and criminal DNA data basing but it also merges marker sets of Y-STRs, X-STRs

and SNPs which are not usually available with conventional CE-methods. These include a heavy set of iiSNPs, piSNPs as well as aiSNPs. iiSNPs is identity-informative single nucleotide polymorphisms, useful for determination of source attribution, specifically with samples that are degraded, PCR-inhibited or mixed [45,50]. Phenotypic-informative SNPs (piSNPs) which give approximation of externally visible characteristics such as eye and hair color (blue, brown, black, blond, red, and intermediate etc.) [25,51,52]. In research, MiSeq FGx that uses only (RUO) mode, can be employed for analysis of mtDNA as well as a wide span of applications. For mt-DNA, after DNA extraction step, mtDNA is amplified. Amplification is done using methods such as D-loop amplification or with a whole mitochondrial genome [53]. Nextera[®] XT DNA Library prep kit is used for preparation of sequencing library. It can generate data about a maximum data od 3.5 to 3.8Gb in a single run with less than 0.1% substitution error rate. ForenSeq Universal Analysis software is available for data analysis.

Ion torrent's PGM system has also been used for forensic genomics [32,54]. It uses sequencing by synthesis chemistry with single nucleotide addition method. Using life technologies' Ion torrent PGM system, 136 unique SNPs contained in grouped PCR amplicons, were sequenced. Sequence (~72 Mb) was generated from two chips (10 Mb Ion 314TM v1) [55]. Raw sequencing data was processed in the Ion torrent Suite[™] Software for the HID-Ion PGM[™]. This software is with the HID SNP Genotyper plugin which is transformed for data analysis. Matlab v.2 (Mathworks, Natick, MA, USA) and R software v.3.3.1 using the ggplot2 package can be employed for data analysis and presentation. In comparison with Illumina, Ion torrent induces 1% indels errors which are undesirable.

NGS workflow

All the NGS platforms have different chemistry but the most commonly used platforms for targeted sequencing such as Ion torrent and Illumina have chemistry of sequencing by synthesis. Few characteristics of both have been compared in Table 1. Illumina uses cyclic reversible termination while Ion torrent makes use of single nucleotide addition method. The basic difference between CE-sequencing and NGS is scalability.

Characteristics	Illumina	ION-torrent PGM	References
Amplification Strategy	Solid phase bridge PCR	Bead based Emulsion PCR	56
Sequencing Chemistry	Cyclic reversible termination	Single nucleotide addition; semiconductor sequencing	43, 56
Read Length	36 (SE); 25-300 (PE)	200-400 (SE)	28
Pair End Sequencing	Yes	No	28
Run Time	27 h	2 h	57
Throughput	Up to 15 Gb	Ion 314TM v2 Up to 100 Mb	28
		Ion 316TM v2 Up to 1 Gb	
		Ion 318TM v2 Up to 2 Gb	
Instrument Cost	\$128 K	\$80 K including PGM, server, onetouch and onetouch ES	57
Cost Per Gb	\$502	\$1000 (318 chip)	57
Errors	0.1% Substitution	1% Indels	28

Forensically Significant	MiSeq® FGxTM Forensic	HID-Ion Personal Genome Machine	
Platform	Genomics System	(PGM)TM	32
Kit	ForenSeq™ DNA Signature Preparation kit	Ion PGM™ Template OT2 200 Kit and the Ion onetouch™ ES	32
SNP Markers sequenced in one reaction	173	124	32
Data Processing and Analysis Tool	ForenSeq™ Universal Analysis software (UAS)	Ion torrent Suite™ software with HID SNP Genotyper Plugin	32
Compatibility with Capillary Electrophoresis	Yes	Yes	49,54

Table 1: Comparison of Illumina Mi-Seq and Ion torrent PGM sequencing platforms.

A single DNA fragment is sequenced by CE. In comparison, NGS in a massively parallel fashion expands this process covering millions of fragments. Forensic genomic applications use targeted sequencing approach rather whole genome sequencing. In this case bar coded sequences (forensically relevant) are targeted via an amplicon-based workflow. Both the Ion torrent and Illumina workflows consist of four basic steps given below as well as illustrated in Figure 4.

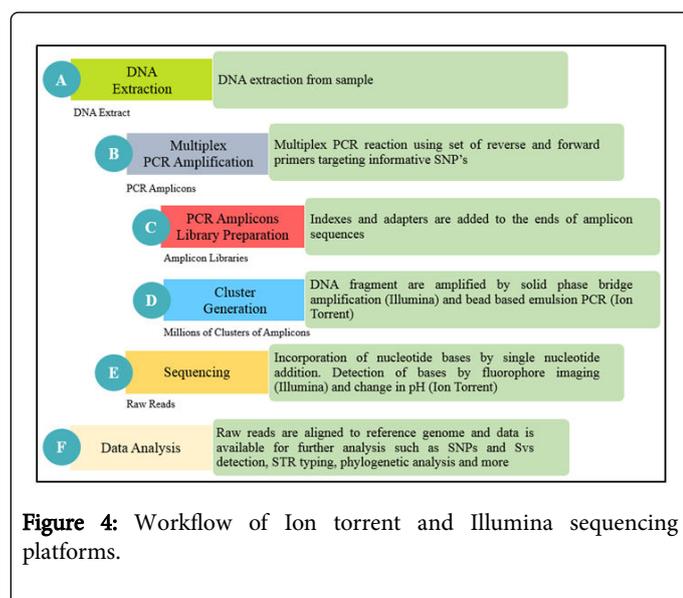


Figure 4: Workflow of Ion torrent and Illumina sequencing platforms.

Sample preparation

Instead of genomic DNA fragmentation, in forensic, relevant amplicon are sequenced. For this purpose a library is prepared by PCR amplification using sequence-specific, tagged primer pairs. Indexes and adapters are attached with the amplicons during the 2nd PCR step. The amplicon libraries following purification, are merged into a single tube, and finally linearized.

Generation of clusters

Clonal templates are prepared during this step. Illumina uses solid-phase bridge amplification method. Free templates are allowed to hybridize with solid-bound adapters. Hybridized templates outer ends interact with nearby primers. This is the point where amplification can

takes place. Clonal clusters are formed after many (100-200 million) amplification rounds. On flow cells, Microwells guide cluster generation, hence expanding cluster density. Ion torrent uses bead based emulsion PCR for cluster generation. Along with primers, templates, dNTPs and polymerase, Micelle droplets are also loaded. Templates are amplified after hybridization to bead bound primers. The complement strand dissociates following amplification, leaving bead bound single stranded DNA templates. 100 to 200 million beads with 1000's of bound templates come as a final product. After cluster generation, sequencing of the templates can be done.

Sequencing and imaging

Illumina employs sequencing by synthesis technology with reversible termination method of cyclic fashion [43]. Fluorophore-labeled nucleotides which are blocked from ends are hybridized to a complementary base. On the slide every single cluster can add a different base. Two or four laser channels are used to obtain slides image. Emission of a color by each cluster corresponds to addition of a specific base during cycle. From the flow cell following removal and cleavage of Fluorophore, regeneration of 3-OH group is done. The incorporation of a new nucleotide initiates a new cycle. Natural competition of nucleotides during their incorporation minimizes bias and greatly reduces raw error rates even within repetitive sequence regions and photopolymers [34,35,56,57]. Ion torrent is also called a semiconductor sequencer in which addition of a base results in release of a single H⁺ ion which further leads to 0.02 unit change in pH. To detect this released H⁺ ion, integrated device such as CMOS (complementary metal-oxide semiconductor) and ISFET (ion-sensitive field-effect transistor) is employed. During each cycle only one dNTP species is there for incorporation of a single nucleotide. Addition of several identical dNTP's during a cycle, leads to an increase in the emitted ions which are detected by pH change.

Data analysis and application

After trimming the read for Phred quality control, alignment of the sequence reads to a reference genome is done. Specific analysis can be done depending on nature of the experiments for example investigation of single nucleotide polymorphism (SNP), mitochondrial DNA study, STR typing, phylogenetic or metagenomic studies etc. ForenSeq Universal Analysis Software is available for analysis Illumina sequence files. For the HID-Ion PGM™, processing was done for raw data of sequencing in the Ion torrent suite™ software with the HID SNP

Genotyper plugin designed for data analysis. In Matlab v.2 (Mathworks, Natick, MA, USA) and R software v.3.3.1 using the ggplot2 package, Data can be presented and analyzed.

Advancements in NGS

Ion torrent is launching the third iteration of its technology in the Ion S5 and Ion S5 XL. These systems include the same core instrument, but the Ion S5 XL adds local computing to enhance analysis speed. Taking a step back from the whole-genome market, these Ion torrent instruments are aimed at targeted sequencing workflows, with the goal of greatly simplifying the hands-on time. The first released applications will be amplicon based via Ion AmpliSeq technology, a well-tested and robust multiplex amplification method for creating sequencing templates [58,59] and will require a reported 45 minutes of hands-on time when coupled with the supporting Ion Chef system. The output of both instruments will use one of three chips, with outputs per chip ranging from 600 Mb to 15 GB and read lengths of 200 or 400 nucleotides (nt), although current specifications list the highest-output mode (10-15 Gb) as being limited to 200-nt reads [60]. Other NGS technologies have been developed by companies like Qiagen, Helicos biosciences, and Pacific biosciences (now part of Roche). These platforms sequence directly from template DNA, while the previously discussed NGS technologies amplify the DNA template during the library preparation steps. Thus, these technologies are called third-generation sequencing technologies (to distinguish from NGS technologies). The leader of the third-generation sequencing field is Roche. Roche has released PacBio RS II from Pacific biosciences technology [61], and the PacBio RS II generates several thousands of long reads with up to 20,000bp. Recently, Oxford nanopore developed a pocket-size MPS device called MinION. This study demonstrates proof-of-concept forensic SNP genotyping using the Oxford nanopore MinION sequencing platform and shows the current capabilities of the system [62].

Sequencing in Nanopore

The MinION is a 90 g portable device. At its core is a flow cell bearing up to 2048 individually addressable nanopores that can be controlled in groups of 512 by an application-specific integrated circuit (ASIC). Prior to sequencing, adapters are ligated to both ends of genomic DNA or c-DNA fragments. These adapters facilitate strand capture and loading of a processive enzyme at the 5'-end of one strand. The enzyme is required to ensure unidirectional single-nucleotide displacement along the strand at a millisecond time scale. The adapters also concentrate DNA substrates at the membrane surface proximal to the nanopore, boosting the DNA capture rate by several thousand-fold. In addition, the hairpin adapter permits contiguous sequencing of both strands of a duplex molecule by covalently attaching one strand to the other. Upon capture of a DNA molecule in the nanopore, the enzyme processes along one strand (the 'template read'). After the enzyme passes through the hairpin, this process repeats for the complementary strand (the 'complement read'). As the DNA passes through the pore, the sensor detects changes in ionic current caused by differences in the shifting nucleotide sequences occupying the pore. These ionic current changes are segmented as discrete events that have an associated duration, mean amplitude, and variance. The information from template and complement reads is combined to produce a high-quality '2D read', using a pairwise alignment of the event sequences [63]. Sequence scalability, by using chips of various capacity ranging from 50 Mb to 1 Gb combined with short run times are the main advantages of the platform, although the 2-hour run time is only theoretical, as the

run time is 5 h for a 318 chip and the whole procedure takes 1 working day, taking into account reagent preparation and PGM initialization [64].

Limitations of Next-Generation Sequencing

Although NGS beneficial in various aspects but comes with several limitations too. As it provides information on a number of molecular abnormalities, but for many of these, clinical significance is currently unknown. With evolution of technology, new challenges emerge, covering entire NGS workflow from sample preparation to steps of data analysis. Coping up with these challenges is not only difficult but also do not ensure good impact on human health [1]. Although much cost and time effective compared to 1st-generation sequencing, NGS is still an expensive tech for many labs. NGS platforms as a start-up can cost \$100,000, and \$1,000 per genome is cost for single sequencing reactions. Wrong sequencing of spans of repeating nucleotides i.e. homopolymer regions, on platforms such as Ion torrent PGM, and short read sequencing of 200 to 500 nucleotides can lead to sequence errors. Platforms such as Illumina and Ion torrent as discussed in detail in current review also have limitations. Possibility of substitution error rates and lagging strand dephasing are Illumina's limitations. Moreover it can have relatively few reads and more cost per Mb. Ion torrent also needs more time and yields fewer reads at higher cost per Mb. A persistent up-gradation of overall sequencing technology, vigorous data infrastructure, analytical tools development, vast all-time knowledge of genome and aberrations in it are still required for further development of this field. NGS requires costly resources including fast processing of data, advanced efficient bioinformatics systems and large data storage potential. Another limitation is lack and shortage of trained manpower, computational aid, and special knowledge of bioinformatics to garner accurate information to analyse and interpret data [65].

Conclusion

Next generation sequencing launched a broad brand new field of study and research with the abilities to transform forensic sciences. Forensic workflows available at present which employs those old approaches existing before Human Genome Project and needs several analysis rounds to generate thorough genetic profiles, prune the potential of genomics. NGS is facilitating forensic scientists by giving access to a large number of important loci, better study of devalued samples as well as providing higher sequencing resolution and significant overall throughput with the help of library multiplexing. Such advancements assist in resolving large number of cases in a time effective way and will also yield productive results for those cases that would have reached blind alley. Specifically for forensic genomic applications, the pioneer NGS system which was designed is MiSeq FGx System [66] with Illumina SBS as its fundamental technology. For an enhanced well-planned worldwide collaborative work with respective regional law enforcement, DNA database organizations are increasing number of their marker sets. It is a never ending journey of research teams world-wide as with every passing day, they persistently bringing novel potential using NGS for applications in human identity testing as well as forensic. At present research efforts and attempts are analyzing piSNPs that gives link with physical characteristics including overall facial morphology for example shape of features like nose, lips, ears etc. as well as with nature of hair (straight or curly). Whether present or forthcoming, all of these NGS tools have the capability to make revolutionary science possible and also has potential in forensic

genomics applications. Although having an important part in future forensic studies, NGS still needs more efforts to fully attain goals including: overcoming problems of error rate, issues with low-template library preparation, type estimations and issues related to mining and processing of NGS data. There is also need to lay-out a proper guideline for the application of NGS in forensics. NGS will soon become an approachable routine in forensic practice due to the technical advances in it and persistent translational attempts of forensic scientists.

Acknowledgements

Prof. Dr. Muhammad Ansar at QAU; Dr. Allah Rakha Lahore; Dr. Kulsoom Sughra at UoG for valuable suggestions.

References

1. Yang Y, Xie B, Yan J (2014) Application of next-generation sequencing technology in forensic science. *Genomics proteomics bioinformatics* 12: 190-197.
2. Van Dijk EL, Auger H, Jaszczyszyn Y, Thermes C (2014) Ten years of next-generation sequencing technology. *Trends Genet* 30: 418-426.
3. Tang H, Kirkness EF, Lippert C, Biggs WH, Fabani M, et al. (2017) Profiling of short-tandem-repeat disease alleles in 12,632 human whole genomes. *Am J Hum Genet* 101: 700-715.
4. Butler JM (2005) *Forensic DNA typing: Biology, technology, and genetics of STR markers*, 2nd edn, Academic Press, New York, USA.
5. Parson W (2018) Age estimation with DNA: From forensic DNA fingerprinting to forensic (Epi) genomics: A mini-review. *Gerontology* 1 : 1-7.
6. Gill P (2002) Role of short tandem repeat DNA in forensic casework in the UK-past, present, and future perspectives. *Biotechniques* 32: 366-372.
7. Butler JM (2006) Genetics and genomics of core short tandem repeat loci used in human identity testing. *J Forensic Sci* 51: 253-265.
8. Amer SAM, Al-Harthi HS, Refaat AM, Bakdash A, Kassab AC, et al. (2017) Identification of human bone remains by autosomal STRs and mitochondrial DNA SNPs. *J Hard Tis Bio* 26: 111-118.
9. Jobling MA, Gill P (2004) Encoded evidence: DNA in forensic analysis. *Nat Rev Genet* 5: 739-751.
10. Lee H, Lee JW, Jeong SJ, Park M (2017) How many single nucleotide polymorphisms (SNPs) are needed to replace short tandem repeats (STRs) in forensic applications? *Int J Legal Med* 131: 1203-1210.
11. Cho S, Shin ES, Yu HJ, Lee JH, Seo HJ, et al. (2017) Set up of cut-off thresholds for kinship determination using SNP loci. *Forensic Sci Int Genet* 29: 1-8.
12. Visscher PM, Brown MA, McCarthy MI, Yang J (2012) Five years of GWAS discovery. *Am J Hum Genet* 90: 7-24.
13. Butler JM, Coble MD, Vallone PM (2007) STRs vs. SNPs: Thoughts on the future of forensic DNA testing. *Forensic Sci Med Pathol* 3: 200-205.
14. Kayser M, de Knijff P (2011) Improving human forensics through advances in genetics, genomics and molecular biology. *Nat Rev Genet* 12: 179-192.
15. Gill P, Koumi P, Allen H (2001) Sizing short tandem repeat alleles in capillary array gel electrophoresis instruments. *Electrophoresis* 22: 2670-2678.
16. Prinz M (2003) Advantages and disadvantages of Y-Short tandem repeat testing in forensic case work. *Forensic Sci Rev* 15: 191-198.
17. Lazaruk K, Walsh PS, Oaks F, Gilbert D, Rosenblum BB, et al. (1998) Genotyping of forensic short tandem repeat (STR) systems based on sizing precision in a capillary electrophoresis instrument. *Electrophoresis* 19: 86-93.
18. Butler JM, Buel E, Crivellente F, McCord BR (2004) Forensic DNA typing by capillary electrophoresis using the ABI prism 310 and 3100 genetic analysers for STR analysis. *Electrophoresis* 25: 1397-1412.
19. Sobrino B, Brion M, Carracedo (2005) A SNPs in forensic genetics: A review on SNP typing methodologies. *Forensic Sci Int* 154: 181-194.
20. Pati N, Schowinsky V, Kokanovic O, Magnuson V, Ghosh S, et al. (2004) A comparison between SNaPshot, pyrosequencing, and biplex invader SNP genotyping methods: Accuracy, cost, and throughput. *J Biochem Biophys Methods* 60: 1-12.
21. Phillips C, Salas A, Sanchez JJ, Fondevila M, Gomez-Tato A, et al. (2007) Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs. *Forensic Sci Int Genet* 1: 273-280.
22. Fondevila M, Phillips C, Santos C, Freire Aradas A, Vallone PM, et al. (2013) Revision of the SNPforID 34-plex forensic ancestry test: Assay enhancements, standard reference sample genotypes and extended population studies. *Forensic Sci Int Genet* 7: 63-74.
23. Phillips C, Aradas AF, Kriegel AK, Fondevila M, Bulbul O, et al. (2013) Eurasiaplex: A forensic SNP assay for differentiating european and south asian ancestries. *Forensic Sci Int Genet* 7: 359-366.
24. Walsh S, Lindenbergh A, Zuniga SB, Sijen T, Knijff PD, et al. (2011) Developmental validation of the IrisPlex system: Determination of blue and brown iris colour for forensic intelligence. *Forensic Sci Int Genet* 5: 464-471.
25. Walsh S, Liu F, Wollstein A, Kovatsi L, Ralf A, et al. (2013) The HIrisPlex system for simultaneous prediction of hair and eye colour from DNA. *Forensic Sci Int Genet* 7: 98-115.
26. Wang L, Hea W, Mao J, Wang H, Jina B, et al. (2015) Development of a SNP-STRs multiplex for forensic identification. *Forensic Sci Int Genet* 5 :e598-e600.
27. Oostdik K, Lenz K, Nye J, Schelling K, Yet D, et al. (2014) Developmental validation of the PowerPlex[®] fusion system for analysis of casework and reference samples: A 24-locus multiplex for new database standards. *Forensic Sci Int Genet* 12: 69-76.
28. Goodwin S, McPherson JD, McCombie WR (2016) Coming of age: Ten years of next-generation sequencing technologies. *Nat Rev Genet* 17: 333-351.
29. Illumina (2014) ForenSeq DNA signature prep kit data sheet, California, USA.
30. Churchill JD, Chang J, Ge J, Rajagopalan N, Wootton SC, et al. (2015) Blind study evaluation illustrates utility of the Ion PGM[™] system for use in human identity DNA typing. *Croat Med J* 56: 218-229.
31. Zhang J, Kobert K, Flouri T, Stamatakis A (2014) PEAR: A fast and accurate Illumina paired-end reAd mergeR. *Bioinformatics* 30: 614-620.
32. Apaga DLT, Dennis SE, Salvador JM, Calacal GC, Ungria MCAD, et al. (2017) Comparison of two massively parallel sequencing platforms using 83 single nucleotide polymorphisms for human identification. *Sci Rep* 7: 398.
33. Buchard A, Kampmann ML, Poulsen L, Børsting C, Morling N, et al. (2016) ISO 17025 validation of a next-generation sequencing assay for relationship testing. *Electrophoresis* 37: 2822-2831.
34. Sebastian J, Fritz JS, Karola P, Albersmeier A, John U, et al. (2013) Updating benchtop sequencing performance comparison. *Nat Biotechnol* 31: 294-296.
35. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, et al. (2012) Performance comparison of benchtop high-throughput sequencing platforms. *Nat Biotechnol* 30: 434-439.
36. Alvarez-Cubero MJ, Saiz M, Martínez-García B, Sayalero SM, Entrala C, et al. (2017) Next generation sequencing: An application in forensic sciences? *Ann Hum Biol* 44: 581-592.
37. Churchill JD, Schmedes SE, King JL, Budowle B (2016) Evaluation of the Illumina[®] Beta version ForenSeq[™] DNA signature prep kit for use in genetic profiling. *Forensic Sci Int Genet* 20: 20-29.
38. Åger AC, Alvarez ML, Davis CP, Guzmán E, Han Y, et al. (2017) Developmental validation of the MiSeq FGx forensic genomics system for targeted next generation sequencing in forensic DNA casework and database laboratories. *Forensic Sci Int Genet* 28: 52-70.

39. Li H, Zhao X, Ma K, Cao Y, Zhou H, et al. (2017) Applying massively parallel sequencing to paternity testing on the ion torrent personal genome machine. *Forensic Sci Int Genet* 31: 155-159.
40. Zascavage RR, Shewale SJ, Planz JV (2013) Deep-sequencing technologies and potential applications in forensic DNA testing. *Forensic Sci Rev* 25: 79-105.
41. Børsting C, Morling N (2015) Next generation sequencing and its applications in forensic genetics. *Forensic Sci Int Genet* 18: 78-89.
42. Caratti S, Turrina S, Ferriani M, Cosentino E (2015) MiSeq FGx sequencing system: A new platform for forensic genetics. *Forensic Sci Int Genet* 5: e98-e100.
43. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, et al. (2008) Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* 456: 53-59.
44. Bragg LM, Stone G, Butler MK, Hugenholtz P, Tyson GW, et al. (2013) Shining a light on dark sequencing: Characterising errors in ion torrent PGM data. *PLoS Comput Biol* 9: e1003031.
45. Sanchez JJ, Phillips C, Børsting C, Balogh K, Bogus M, et al. (2006) A multiplex assay with 52 single nucleotide polymorphisms for human identification. *Electrophoresis* 27: 1713-1724.
46. Pakstis AJ, Speed WC, Fang R, Hyland FCL, Furtado MR, et al. (2010) SNPs for a universal individual identification panel. *Hum Genet* 127: 315-324.
47. Kidd KK, Kidd JR, Speed WC, Fang R, Furtado MR, et al. (2012) Expanding data and resources for forensic use of SNPs in individual identification. *Forensic Sci Int Genet* 6: 646-652.
48. Hussing C, Kampmann ML, Mogensen HS, Børsting C, Morling N, et al. (2015) Comparison of techniques for quantification of next-generation sequencing libraries. *Forensic Sci Int Genet* 5: e276-e278.
49. Mehta B, Daniel R, Phillips C, Doyle S, Elvidge G, et al. (2016) Massively parallel sequencing of customized forensically informative SNP panels on the MiSeq. *Electrophoresis* 37: 2832-2840.
50. Kidd KK, Pakstis AJ, Speed WC, Grigorenko EL, Sylvester LBK, et al. (2006) Developing a SNP panel for forensic identification of individuals. *Forensic Sci Int* 164: 20-32.
51. Kidd KK, Speed WC, Pakstis AJ, Furtado MR, Fang R, et al. (2013) Progress toward an efficient panel of SNPs for ancestry inference. *Forensic Sci Int Genet* 10: 23-32.
52. Phillips C, Prieto L, Fondevila M, Salas A, Gómez-Tato A, et al. (2009) Ancestry analysis in the 11-M madrid bomb attack investigation. *PLoS One* 4: e6583.
53. Illumina (2017) Innovative technologies, California, USA.
54. Børsting C, Fordyce SL, Olofsson J, Mogensen HS, Morling N, et al. (2014) Evaluation of the Ion torrent™ HID SNP 169-plex: A SNP typing assay developed for human identification by second generation sequencing. *Forensic Sci Int Genet* 12: 144-154.
55. Daniel R, Santos C, Phillips C, Fondevila M, van Oorschot RAH, et al. (2015) A SNaPshot of next generation sequencing for forensic SNP analysis. *Forensic Sci Int Genet* 14: 50-60.
56. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 475: 348-352.
57. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, et al. (2012) A tale of three next generation sequencing platforms: Comparison of ion torrent, pacific biosciences, and Illumina MiSeq sequencers. *BMC Genomics* 13: 341.
58. Burghel GJ, Hurst CD, Watson CM, Chambers PA, Dickinson H, et al. (2015) Towards a next-generation sequencing diagnostic service for tumour genotyping: A comparison of panels and platforms. *Biomed Res Int* 2015: 478017.
59. Trujillano D, Weiss ME, Koster J, Papachristos EB, Werber M, et al. (2015) Validation of a semiconductor next-generation sequencing assay for the clinical genetic screening of CFTR. *Mol Genet Genom Med* 3: 396-403.
60. Shawn EL, Myers RM (2017) Advancements in next-generation sequencing. *Annu Rev Genomics Hum Genet* 17: 95-115.
61. Menlo P (2015-2016) Pacific Biosciences; The original long-read sequencer, California, USA.
62. Cornelis S, Gansemans Y, Deleze L, Deforce D, Nieuwerburgh FV, et al. (2017) Forensic SNP genotyping using nanopore MinION sequencing. *Sci Rep* 7.
63. Jain M, Olsen HE, Paten B, Akeson M (2016) The oxford nanopore MinION: Delivery of nanopore sequencing to the genomics community. *Genome Biol* 17: 239.
64. Tarabeux J, Zeitouni B, Moncoutier V, Tenreiro H, Abidallah K, et al. (2014) Streamlined ion torrent PGM-based diagnostics: BRCA1 and BRCA2 genes as a model. *Eur J Hum Genet* 22: 535-541.
65. Basho RK, Agda KE, Funda MB (2015) Clinical applications and limitations of next-generation sequencing. *Am J Hematol Oncol* 11: 17-22.
66. Sharma V, Chow HY, Siegel D, Wurmbach E (2017) Qualitative and quantitative assessment of Illumina's forensic STR and SNP kits on MiSeq FGx™. *PLoS One* 12: e0187932.