Chapter 1
Introduction

Nitin Mahajan and Todd E. Druley

The detection of minimal residual disease (MRD) – or more aptly named measurable residual disease – has evolved substantially over recent decades with the steady improvement of technology. From gross morphology to karyotyping to cytogenetics to flow cytometry, MRD has matured and saved countless lives in the process by identifying those who require augmented therapy in order to overcome refractory or relapsed leukemia. New technologies, particularly with respect to DNA and RNA sequencing, offer such extreme sensitivity that focus has shifted to being certain that the mutation(s) detected is indeed representative of the leukemia population and not an incidental finding. Even healthy individuals harbor a rich profile of clonal hematopoietic

N. Mahajan
Washington University in St. Louis, Pediatric Hematology and Oncology, St. Louis, MO, USA

T. E. Druley (✉)
Washington University in St. Louis, Division of Hematology and Oncology, Department of Pediatrics, Center for Genome Sciences and Systems Biology, Saint Louis, MO, USA
e-mail: druley_t@wustl.edu

T. E. Druley (ed.), *Minimal Residual Disease Testing*,
https://doi.org/10.1007/978-3-319-94827-0_1

mutations [1] that is not fully understood and could lead to false positives without careful calibration. This is a concern for using cell-free or circulating tumor DNA in solid tumors as markers for metastatic or recurrent cancer.

This text is intended to not only review the history of methods utilized for MRD but also summarize the current state of the art as well as predict where MRD will move in the coming years. Clearly, with the rapid decline in sequencing costs coupled with the massive amounts of data generated, it will be sequencing strategies – both in bulk and in single cells – that dominate MRD in the near future. To that end, it seems appropriate to offer a brief history of nucleic acid sequencing and highlight some of the emerging sequencing platforms that are most likely to change the way laboratories and physicians order and view MRD.

History of DNA Sequencing

In year 1910, Albrecht Kossel discovered nucleotide bases adenine, cytosine, guanine, thymine, and uracil as the building block of nucleic acid [2]. Four decades later, Erwin Chargaff recognized the pairing pattern of these nucleotides in DNA and RNA [2]. Robert Holley and colleagues (1965) were accredited for sequencing the first ever full nucleic acid molecule, 77-nucleotide yeast (*Saccharomyces cerevisiae*) alanine tRNA with a proposed cloverleaf structure [3]. It took more than 5 years to extract enough tRNA from the yeast to identify the sequence of nucleotide residues using selective ribonuclease treatment, two-dimensional chromatography, and spectrophotometric procedures [3]. The laborious and expensive nature of the sequencing did not deter the scientists but rather drove the continuous development and refinement of sequencing methods. Initially, scientists focused sequencing efforts on the readily available populations of RNA species because (i) of bulk production in culture, (ii) it is not complicated by a complementary strand, and (iii) it is considerably thought to be shorter than DNA molecules [4, 5].

Fred Sanger and colleagues at Cambridge were one of the groups actively working on methods for sequencing DNA molecules. They developed a technique based on the detection of radiolabeled partial digestion fragments after two-dimensional fractionation [6], allowing addition of nucleotides to the growing pool of ribosomal and transfer RNA sequences. Using a primer extension method, Ray Wu and Dale Kaiser sequenced a short sequence of DNA for the first time [7]. However, the actual determination of bases was still restricted to small sequences of DNA because of the labor and use of radioactive and hazardous chemicals. These continuous efforts resulted in generating the first complete protein-coding gene sequence, coat protein of bacteriophage MS2 in 1972 [8], and the first complete 3569-nucleotide-long genome sequence of the bacteriophage MS2 RNA in 1976 [9].

Two influential techniques in the mid-1970s emerged which later gave a new dimension to the field of molecular biology. The two techniques were Alan Coulson and Sanger's "plus and minus" system and Allan Maxam and Walter Gilbert's chemical cleavage technique [10–12]. Both these techniques used polyacrylamide gel electrophoresis, which provided better resolving power, instead of previously used 2-D fractionation that often consisted of both electrophoresis and chromatography. The plus and minus technique was based on the addition of radiolabeled nucleotides next to the primer using DNA polymerase.

A "plus" reaction is where only a single type of nucleotide is present with an aim that all extensions will end with that particular nucleotide whereas a "minus" reaction three nucleotides are used to produce sequences up to the position before the next missing nucleotide. This led Sanger and colleagues to sequence the first DNA genome, that of bacteriophage [11]. On the other hand, the technique used by Maxam and Gilbert to sequence the DNA was quite different, as they used chemicals to fragment the radiolabeled DNA at particular bases. Fragmented radiolabeled DNA was electrophoresed through a polyacrylamide gel and based on the length of the cleaved fragments the sequence was inferred. Development

of these two methods can be described as the foundation of modern sequencing. However, the major discovery in the field of DNA sequencing came in 1977 with the Sanger's "chain-termination" or "dideoxy technique," and since then it is the most widely used sequencing method.

The chain-termination technique utilizes labeled (radioactively or fluorescently) chemical analogues of the deoxynucleotides (dNTPs), which are called dideoxynucleotides (ddNTPs). The reaction includes a single-stranded DNA template, DNA primer, DNA polymerase, normal deoxynucleotides (dNTPs), and modified dideoxynucleotides (ddNTPs). Because ddNTPs lack a 3'-OH group, they are unable to make phosphodiester bond which ultimately terminates DNA strand elongation. A total of four DNA sequencing reactions are made, and in each reaction, three normal dNTPs and one labeled ddNTP are added. This results in the synthesis of each possible length of the DNA molecule of interest. The nucleotide sequence is inferred by resolving the product from each reaction in a separate lane of a polyacrylamide gel. Initially, scientists were able to determine the sequence of a molecule up to 300 bp.

The potential Sanger sequencing was realized quickly by the scientific community, and a series of improvements were made in the following years. Major improvements were, first, replacement of radioactive materials with fluorescent-based detection, which allowed the reaction to occur in one vessel instead of four. A second key improvement was the use of capillary-based electrophoresis which provided better resolution, required less equipment space, and decreased the amount of time required for the experiment. Following these improvements, Smith et al. (1986) at Applied Biosystems Instruments (ABI) designed an automated machinery to complete this procedure and later introduced the first commercial automated DNA sequencer [13].

First-generation sequencers incorporated a computer-based data acquisition and analysis and were capable of producing reads >300 bp. However, to analyze longer DNA molecules, "shotgun sequencing" was developed by separately

cloning and sequencing overlapping DNA fragments. After sequencing these molecules are assembled into one long contiguous sequence [14]. The discovery of polymerase chain reaction (PCR) technology during this time period provided a viable solution for generating high concentrations of specific DNA species and aided in the re-sequencing of particular regions. With the addition of newer technologies and increased interest in sequencing, ABI sequencers were significantly improved over the next few years. These improvements included an increase in the number of lanes in gel-based models from 1 (ABI 310) to 16 (ABI 370A) and then to 96 (ABI 377). At the same time, the length of the reads increased from 350 (ABI 370A) to over 900 (ABI 3730xl), while the run times decreased from 18 h to 3 h [15].

The Institute for Genomic Research (TIGR) in Rockville, Maryland, founded by J. Craig Venter in 1992, pioneered the industrialization of an automated sequencer, with a focus on studying various genomes [16, 17]. With the establishment of both the first Affymetrix and GeneChip microarrays in 1996, expression studies involving various genes in prokaryotes and eukaryotes were now possible [18]. By the end of 1999, with continuous effort of various researchers, TIGR generated 83 million nucleotides of cDNA sequence, 87,000 human cDNA sequences, and the complete genome sequences of *Haemophilus influenzae* [19] and *Mycoplasma genitalium* [20].

With the beginning of the new century, though expensive and time-consuming, sequencing centers and international consortiums, such as the TIGR in the USA, the Sanger Centre in the UK, and RIKEN in Japan, using the automated sequencers, produced the complete sequence of the human genome. Additionally, the genomes of *Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (fruit fly), and the plant *Arabidopsis thaliana* were also completed [4, 15, 17, 21]. Despite all these accomplishments, new sequencing methods continued to emerge with the aim to reduce costs, increase multiplexing, decrease time, and increase throughput.

Ultimately these improved methods have been realized over the past few decades and have paved the path forward for next-generation sequencing applications.

Next-Generation Sequencing Application

As advancements were being realized in sequencing applications, often referred to as next-generation sequencing (NGS), key improvements included (i) the parallelization of high number of sequencing reactions, (ii) the preparation of amplified sequencing libraries prior to sequencing, (iii) library amplification on miniature surfaces (solid surfaces, beads, emulsion droplets), (iv) direct monitoring of the nucleotides, (v) reduced cost, and (vi) decreased time.

There are a wide variety of NGS applications that can be used to study the whole genome, coding regions (exomes), transcriptome, DNA methylation, mitochondrial DNA, plus several other novel applications, such as micro-RNA and noncoding RNA sequencing. Sequencing applications for RNA are similar to that of DNA, with an additional step to generate cDNA from RNA using a reverse transcriptase. For targeted sequencing the exomes or regions of interest within the fragmented DNA can be captured and enriched by probe hybridization or by customized PCR amplification. Targeted panel sequencing involves a focused approach on known alleles of gene candidates, associated with the phenotype of interest.

The general workflow for an NGS assay involves (1) the isolation of nucleic acids (DNA or RNA), (2) the capture of DNA molecules of interest, (3) sequencing, and (4) bioinformatics analysis of the massive unstructured dataset [5, 17, 22, 23]. The exact procedures involved in each of these steps vary between sequencing platforms and library preparation protocols. Ultimately, since the genetics of cancer is extremely heterogeneous, it will be essential to use the appropriate technique for the type of variant of interest. Like instrumentation and protocols, the revision of read lengths occurs rapidly and will likely continue to do so as chemistries are

optimized and improved. Determining an appropriate read length for sequencing, short versus long, depends on the goal of the experiment.

Long-read sequencing (LRS) techniques have been key for phasing studies and alternative splicing. However, as short-read sequencing (SRS) technologies advance in the single-cell sequencing field, these types of analysis will be more easily attended with SRS technology. Short-read sequencing (SRS) typically produces reads that are 50–600 bp in length and often results in sequences with scaffolding gaps, bias due to high GC content, repeat sequences, and missing insertions. LRS techniques produce reads between 10Kb and 40Kb [24–27].

Illumina is the dominant SRS platform by supporting paired-end sequencing (although other platforms exist including Thermo Fisher Scientific, Ion Torrent, and Complete Genomics) [28], whereas Pacific Biosystems and Oxford Nanopore Technologies sequencers are dominant in generating long reads. There are several advantages SRS have which include high throughput, low cost per base, and a low raw read rate [28]. However, the short-read length complicates genome alignment leading to false-positive and false-negative variant calling [29, 30]. Furthermore, *de novo* assembly of short sequencing reads can be challenging due to minimal overlap between raw reads, which therefore require enhanced algorithms for successful assembly, such as SOAPdenovo [31], in order to assemble a large genome of interest, although genome assemblies, especially for non-model organisms, generated from SRS are limited as long-range linking information is limited [32].

There are several variant algorithm detection methods, including FreeBayes [33] that are specific for SRS data. The advantages for SRS for MRD include low error rate and the ability to generate deep coverage for a specific region of the genome. Therefore, SRS has dominated the field for cancer genomics as variant detection is more accurate with SRS over LRS techniques that have a higher error rate and less sensitive limit of detection.

More than 70% of genetic variations seen in humans are non-SNP variations and can be missed easily with short-read sequencing [34]. Long-read sequencing enables reads longer than 10 kb, which improves alignment to the reference genome, high consensus accuracy, uniform coverage, and detection of epigenetic modifications. In addition, long-read sequencing is beneficial in transcriptomic analyses as it allows detection of splice isoforms with a high level of confidence without requiring assembly. High costs of long-read sequencing and high error rates are the major hurdle for adopting these platforms as a global sequencing platform.

Roche 454 Pyrosequencing

The first commercially available second-generation sequencer was developed by 454 Life Sciences in 2005 and was based on pyrosequencing. In 2007, 454 Life Sciences was acquired by Roche [24]. Pyrosequencing is based on the detection of light signal generated by the release of pyrophosphate (PPi) upon incorporation of dNTP in presence of ATP sulfurylase, luciferase, DNA polymerase, and adenosine 5′-phosphosulfate (APS). Luciferase ATP mediates conversion of luciferin to oxyluciferin, which generates a light signal during repeated nucleotide incorporation into the newly synthesized DNA chain. The ability to run massive sequencing reactions in parallel per run is the obvious advantage of this machine. DNA libraries were tagged to beads using adaptor sequences and using emulsion PCR in a pico-liter plate where ideally each well gets one DNA bead [23]. This miniaturized system used massively parallelized sequencing to produce more than 200,000 reads at 100–150 bp per read with an output of 20 Mb per run in 2005 [35].

In 2008, Roche released the new version 454 GS FLX titanium system with improved average read length up to 700 bp with an accuracy of 99.997% and an output of 0.7 Gb of data per run within 24 h. Roche combined the 454 sequencing system in 2009 with the GS junior, a benchtop system. This new application simplified library preparation protocol, improved data processing steps, and improved the time requirements per run to 14 h. The use of these systems was

limited by the high cost of reagents and high error rates in homopolymer repeats [26, 36–38]. However, with the commercial availability of the sequencing, various other companies have launched new sequencers as discussed below.

Illumina (Solexa) Sequencers

In 2006, Solexa released the Genome Analyzer (GA), and in 2007 the company was purchased by Illumina. The Illumina sequencer is different from the Roche 454 sequencer as it uses bridge amplification for colony generation and is based on the sequencing by synthesis (SBS) approach. The library with fixed adaptors is denatured to single strands and grafted to a flow cell, followed by bridge amplification to form clusters (miniature colonies or polonies) that contain clonal DNA fragments. Sequencing by synthesis approach uses removable fluorescently labeled chain-terminating nucleotides, which produce a larger output at lower reagent cost. All the steps during sequencing in the Illumina technology are carried out in a flow cell. Flow cells can have single or multiple lanes depending upon the Illumina instrument.

Illumina provides two styles of sequencing machines – benchtop sequencers (MiniSeq System, MiSeq Series, and NextSeq Series) and production scale sequencers (NextSeq Series, HiSeq Series, HiSeq X Series, and NovaSeq 600 system). These sequencers range from low (0.3Gb) to mid (120Gb) and to high output (1500 Gb). In the present day, Illumina is the dominant sequencing platform for clinical research efforts.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)

Supported oligonucleotide ligation and detection (SOLiD) is another next-generation application that was first released in 2008 by Applied Biosystems Instruments (ABI) and marketed by Life Technologies. This platform is based on two-nucleotide sequencing by ligation (SBL) strategy where sequential annealing of probes is followed by ligation. This process generates

hundreds of millions to billions of short reads with simultaneous two-base encoding for each nucleotide. Sequencer 5500 W series with either one (5500 W system) or two (5500xl system) flow chips are present in the market which are suitable for small- and large-scale projects involving whole genomes, exomes, small RNA, and transcriptomes [39]. As each base is interrogated twice, this platform provides a high accuracy of 99.85% after filtering in addition to the low cost per base; however, short-read lengths (35–85 bp), long run times (7–14 days), and requirement of huge computational infrastructure are major shortcomings [17, 40].

Ion Torrent

The Ion Torrent technology is another platform produced by the inventors of 454 sequencing [41]. This technology is fundamentally different from other platforms as it does not use either fluorescence or luminescence (post-light sequencing technology) but instead uses microchip amalgamated flow cells coupled with electronic sensors. The incorporation of a single nucleotide releases a proton which results in a change of pH and can be measured electronically as a voltage change; if there are two nucleotides added, double voltage is detected [17]. Two sequencing platforms, Proton Sequencer (with more than 165 million sensors) and the Ion Personal Genome Machine (PGM) (a benchtop sequencer with 11.1 million sensors), adapted this technology. Of interest, this technique does not require fluorescence. and the utilization of camera scanning improves the speed, cost, and size of the instrument. The major disadvantages include short-read length and problem in reading homopolymer stretches and repeats [4, 38].

DNA Nanoball Sequencing (DNBS)

DNA nanoball sequencing (DNBS) was developed by the inventor of SBH as a hybrid sequencing application that uses hybridization and ligation. Using four adapter sequences, small fragments of genomic DNA or cDNA (400–500 bp) are

amplified into microscopic DNA nanoballs (roughly 300 nm in size) by rolling circle amplification, generating ssDNA concatemer. The DNA nanoballs are sequenced at a high density as one nanoball per well onto an arrayed flow cell. Up to ten bases of the template are read in the 5′ and 3′ direction from each adapter. Since only short sequences, adjacent to adapters, are read, this sequencing format resembles a multiplexed form of mate-pair sequencing. The short length of reads and sequencing time are the major disadvantages of DNBS, whereas the high density of arrays and therefore the high number of DNBS (~350 million) that can be sequenced are the major advantages [42].

Third-Generation Sequencing

So-called "third-generation" sequencing differs from next- (or second) generation sequencing as (i) PCR is not needed before sequencing which results in shortened time and reduced bias and error caused by PCR; (ii) the signal is captured in real time, which means that the signal, no matter whether it is fluorescent or electric current, is monitored during the enzymatic reaction of adding nucleotide in the complementary strand; (iii) it is capable of sequencing single molecule; (iv) it has low price of sequencing; and (v) it is simpler (the preparatory procedures and sequencing methods are simpler compared to second-generation sequencing).

Single-Molecule Real Time (SMRT)

Single-molecule real-time (SMRT) technology is developed by Pacific Biosciences (Menlo Park, CA, USA) and uses modified enzyme and direct observation of the enzymatic reaction in real time. SMRT cells contain 150,000 ultra-microwells where reaction takes place at a zeptoliter scale [10–21, 43]. Each well is coated with a molecule of DNA polymerase using the biotin-streptavidin system in nanostructures known as zero-mode waveguides (ZMWs) and DNA template that can be detected during the whole process.

During the reaction, fluorescently labeled dNTPs are added to the growing strand and monitored by CCD cameras.

PacBio machines produce long reads (up to and exceeding 10 kb in length), which are useful for de novo genome assemblies, and accuracy reported is >99%. Compared to other technologies, PacBio (i) adapters used in SMRT have a hairpin structure (SMRT loop adapters) which allows circularization of dsDNA after ligation and (ii) does not rely on interrupted cycles of extension and imaging to read the template strand as signals of newly added nucleotides are recorded in real time [24, 40, 43].

Helicos Sequencing

The Helicos sequencing system was the first implementation of single-molecule fluorescent sequencing. Sheared DNA is tailed with polyA tail and hybridized in a flow cell surface coated with oligo-dT for sequencing by synthesis of billions of molecules in parallel. The fluorescent signals were used to detect labeled nucleotide triphosphates incorporated onto DNA templates bound to a quartz slide. This technology sequences the DNA by both the hybridization and sequencing by synthesis using a DNA polymerase. The HeliScope sequencing read lengths range from 25 to over 60 bases, with 35 bases being the average. The Seqll (http://seqll.com) markets this technology to sequences genomic DNA and RNA using the Helicos sequencing system and HeliScope single-molecule sequencers. This method has successfully sequenced the human genome and provided disease signatures in a clinical evaluation and was implemented for sequencing RNA molecules for quantitative transcriptomic analysis of tissues and cells [44–46].

Next-Generation Sequencing by Electron Microscopy

Though detection is not easy, electron microscopy allows the direct visualization of the sequence of DNA molecules. Samples were sequenced by the enzymatic incorporation of modified bases with atoms of increased atomic number.

These high atomic number atoms allow the direct visualization and identification of individually labeled bases under the electron microscope as dark dots. Direct visualization and identification of individually labeled bases within a synthetic 3272 base-pair DNA molecule and a 7249 base-pair viral genome have been demonstrated [47, 48]; however, the technology has not yet been commercially developed.

Two companies, ZS Genetics (http://www.zsgenetics.com) and Electron Optica [49], are continuously working on the DNA sequencing technique by electron microscopy, with different approaches. ZS Genetics demonstrated the labeling and identification of the four bases of DNA with an electron microscopy for the first time in the year 2012. The sequence read lengths range from 5 to 50 kb and are useful for de novo genome assembly and for analysis of full haplotypes and copy number variants (http://www.zsgenetics.com).

However, the heavy atom labeling has few disadvantages which include: (i) chances of incomplete labeling reaction which might result in missing few base pairs, (ii) four different reactions have to run as the labels are difficult to distinguish and interfere with each other when they get too close, and (iii) high-energy electrons damage the DNA and therefore cause errors to precisely locate the bases [47, 50]. Electron Optica uses low-energy electron microscopy (LEEM) which does not need labeling with heavy metals for DNA sequencing. Though LEEM causes less damage to DNA and thus reduces sequencing errors [48, 51, 52], there is no update on this technology since 2014 [49].

Fourth-Generation Sequencing

The fourth-generation sequencing platforms sequence without amplification, real-time sequencing without repeated cycles, and elimination of synthesis. These technologies preserve the spatial coordinates of DNA and RNA sequences with up to subcellular resolution, thus enabling back mapping of sequencing reads to the original histological context [22, 27].

Nanopore Sequencing

Nanopore-based sequencers open a new door to molecular biology investigation at the single-molecule scale. In the 1990s, Church et al. and Deamer and Akeson separately proposed sequencing of DNA using nanopores [53, 54]. The sequencing of DNA is done by passing the single-stranded DNA molecule through a nanopore chamber, which can be found in protein channel which facilitates ion exchange. With the application of an external voltage, particles with sizes smaller than the pore size are passed through the pore which are either embedded in a biological membrane or formed in solid-state film. Major advantages include simple experimental procedure, required no labeling and less input, and generate real-time data with ultra-long reads and high throughput [4, 53, 55]. The nanopore sequencer has wider applications in many areas, such as analysis of DNA, RNA, proteins, peptides, drugs, polymers, etc. [4, 55, 56].

The nanopores can either be from the biological system or solid-state synthetic nanopores. In combination with the other devices and electronic circuits, these pores are integrated in a form of portable sequencing chips. Biological nanopores, also termed transmembrane protein channels, are usually embedded in liposomes or polymer films. There are three main biological experiments involving this technology and the study of *Staphylococcus aureus* α hemolysin (α HL), *Mycobacterium smegmatis* porin A (MspA), and *Bacteriophage* phi29. On the other hand, solid-state synthetic systems fabricate nanopores in silicon nitride (Si3N4), silicon dioxide (SiO2), aluminum oxide (Al2O3), boron nitride (BN), graphene, polymer membranes, and hybrid materials [4, 17, 53, 55].

Solid-state nanopore is reliably more stable than biological and could be multiplexed to work in parallel on a single device and achieve higher readout within a short time. Oxford Nanopore Technologies (ONT) (https://nanoporetech.com), founded by Hagan Bayley and Gordon Sanghera, provide the commercially available nanopore

sequencing instruments that provide real-time data. Major devices from ONT include MinION (pocket-sized, portable device), GridIONX5 (multiplex sequencing device), PromethION (high-throughput, high-sample number benchtop system), and SmidgION (smallest device designed for use with a smartphone in any location) (https://nanoporetech.com). One of the major disadvantages of this technology is the rapid DNA translocation velocity (1–3 ls/base) which limits the identification of single nucleotide bases and increases error rate (up to 90%) [57].

Success with this platform has been noted by combining results with short-read sequencing to improve error rates at the single-base resolution while also producing reads long enough for de novo assembly efforts [58]. Using this technique Goodwin et al. were able to create an open-source hybrid error correction algorithm, Nanocorr, that combines Illumina SRS (MiSeq) data and Nanopore, which enhanced genome assembly and variant detection [59].

BioNano Genomics

BioNano's next-generation mapping uses nano-channel arrays with optical mapping to image extremely long, high-molecular-weight DNA in its most native state instead of classical DNA sequencing devices. This technology provides a detailed genome map, which helps to finish sequencing and to remove sequencing errors caused by repetitive regions. Genome map in addition to the sequencing provides a better resolution of the whole genome, showing its features in context and functional relationships, across kilobases to megabases. BioNano Saphyr provides rapid, high-throughput, long-range genome mapping with the ability to detect the large-scale structural variations (ranging from 1 kb to megabases) missed by next-generation sequencing (NGS) systems (https://bionanogenomics.com/products/saphyr/). The high-resolution Irys System from BioNano Genomics offers whole-genome maps at a single-molecule resolution.

Emerging Platforms: Single-Cell Sequencing (SCS)

Advancements in microfluidics have enabled the robust isolation of single cells, which then facilitates individual cellular analysis of DNA or RNA. Bulk sequencing analyses are based on an averaged signal obtained from a heterogeneous cell or nucleic acid population and therefore obviate resolution at a cellular level – which is overcome via single-cell analysis [60, 61]. Single-cell sequencing is rapidly becoming established as an important tool in a diverse series of disciplines ranging from characterization of cellular diversity to identification of new cell types [62, 63]. Over the past decade, there has been extraordinary progress in the development and application of single-cell DNA and RNA sequencing methods. Studies by Tang et al. (2009) [64] and Navin et al. (2011) [60] describing single-cell RNA and DNA analysis, respectively, are pioneering discoveries in the field. As we are moving toward the era of precision medicine, the data generated from multiple "omics" strategies provides complex insight on biological events. Single-cell analyses of DNA, RNA, and protein generate data demonstrating the heterogeneity of a given cellular network, which ultimately leads to improved understanding of the underlying mechanisms of any physiologic or disease-related process. Every improvement in these assays increases sensitivity and throughput. Additional applications in which single-cell sequencing could offer new insights are with respect to circulating tumor cells prior to a diagnosis, residual circulating cancer cells post-therapy, and stem cell identification and specification. Single-cell arrays provide tools to measure different aspects of various substrates (like DNA for chromatin structure, histone modification, or sequence variability; RNA for gene expression changes, allele-specific expression, fusion events; protein for correlating surface immunophenotypes or ligand-receptor interactions, and many others) [60, 61, 65]. These platforms are expanding rapidly (Table 1.1), and single-cell RNA sequencing platforms have been reviewed elsewhere [66].

Each platform possesses its own unique profile of advantages and disadvantages depending upon the intended application.

In conclusion, the ultimate goal of true precision (personalized) therapy for cancer or several other diseases rests on the accurate characterization of the systemic heterogeneity in genetic and epigenetic variability combined with understanding their respective and additive impact on cellular function as it relates to risk prognostication and therapeutic selection. MRD is a direct reflection of how biomedical scientists and clinicians have applied exciting new technologies to understand the minutiae of human physiology in order to improve human health.

Table 1.1 Overview of currently available single-cell sequencing platforms

Method or company	Applications and references
Drop-Seq	RNA-Seq [67]
InDrops	RNA-Seq [68]
CEL-Seq2	RNA-Seq [69]
Quartz-Seq2	RNA-Seq [70]
Cyto-Seq	RNA-Seq [71]
MARS-Seq	RNA-Seq [72]
Hi-SCL	RNA-Seq [73]
Chromium System (10x Genomics)	RNA-Seq, DNA-Seq, immune repertoire profiling
Nadia (Dolomite Bio)	RNA-Seq, DNA-Seq
ddSEQ Single-Cell Isolator (Bio-Rad)	RNA-Seq
Tapestri Platform (MissionBio)	Targeted DNA-Seq
BD Rhapsody Single-Cell Analysis System (Becton Dickinson)	Targeted RNA-Seq
C1 System and Polaris (Fluidigm)	RNA-Seq, DNA-Seq, miRNA-Seq, epigenomics, RT-qPCR

References

1. Young AL, Challen GA, Birmann BM, et al. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun. 2016;7:12484.
2. The History of DNA Timeline, https://www.dna-worldwide.com/resource/160/history-dna-timeline.
3. Holley RW, Apgar J, Everett GA, et al. Structure of a ribonucleic acid. Science. 1965;147:1462–5.
4. Buermans HPJ, den Dunnen JT. Next generation sequencing technology: advances and applications. Biochim Biophys Acta BBA Mol Basis Dis. 2014;1842:1932–41.
5. Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. Genomics. 2016;107:1–8.
6. Sanger F, Brownlee GG, Barrell BG. A two-dimensional fractionation procedure for radioactive nucleotides. J Mol Biol. 1965;13:373–98.
7. Wu R, Kaiser AD. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. J Mol Biol. 1968;35:523–37.
8. Jou WM, Haegeman G, Ysebaert M, et al. Nucleotide sequence of the gene coding for the bacteriophage MS2 coat protein. Nature. 1972;237:82.
9. Fiers W, Contreras R, Duerinck F, et al. Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature. 1976;260:500.
10. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94:441–8.
11. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage φX174 DNA. Nature. 1977;265:687.
12. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977;74:560–4.
13. Smith LM, Sanders JZ, Kaiser RJ, et al. Fluorescence detection in automated DNA sequence analysis. Nature. 1986;321:674.
14. Anderson S. Shotgun DNA sequencing using cloned DNase I-generated fragments. Nucleic Acids Res. 1981;9:3015–27.
15. Ansorge WJ. Next-generation DNA sequencing techniques. New Biotechnol. 2009;25:195–203.
16. Adams MD, Kelley JM, Gocayne JD, et al. Complementary DNA sequencing: expressed sequence tags and human genome project. Science. 1991;252:1651–6.

17. Kulski JK. Next-generation sequencing — an overview of the history, tools, and "Omic" applications: InTech; 2016. http://creativecommons.org/licenses/by/3.0.
18. Bumgarner R. DNA microarrays: types, applications and their future. Curr Protoc Mol Biol Ed Frederick M Ausubel Al. 2013;0 22:Unit-22.1.
19. Fleischmann RD, Adams MD, White O, et al. Whole-genome random sequencing and assembly of haemophilus influenzae Rd. Science. 1995;269:496–512.
20. Fraser CM, Gocayne JD, White O, et al. The minimal gene complement of mycoplasma genitalium. Science. 1995;270:397–404.
21. Stein L. Genome annotation: from sequence to biology. Nat Rev Genet. 2001;2:493.
22. Ke R, Mignardi M, Hauling T, et al. Fourth generation of next-generation sequencing technologies: promise and consequences. Hum Mutat. 2016;37:1363–7.
23. Koboldt DC, Steinberg KM, Larson DE, et al. The next-generation sequencing revolution and its impact on genomics. Cell. 2013;155:27–38.
24. Liu L, Li Y, Li S, et al. Comparison of next-generation sequencing systems, https://www.hindawi.com/journals/bmri/2012/251364/.
25. Illumina | Sequencing and array-based solutions for genetic research, https://www.illumina.com/index-d.html.
26. SE L, Myers RM. Advancements in next-generation sequencing. Annu Rev Genomics Hum Genet. 2016;17:95–115.
27. Srinivasan S, Batra J. Four generations of sequencing- is it ready for the clinic yet? J Gener Seq Appl. 2014;1:107.
28. Caspar SM, Dubacher N, Kopps AM, et al. Clinical sequencing: from raw data to diagnosis with lifetime value. Clin Genet:n/a–a.
29. Derrien T, Estellé J, Sola SM, et al. Fast computation and applications of genome Mappability. PLoS One. 2012;7:e30377.
30. Mandelker D, Schmidt RJ, Ankala A, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genet Med. 2016;18:1282.
31. Li R, Zhu H, Ruan J, et al. De novo assembly of human genomes with massively parallel short read sequencing. Genome Res. 2010;20:265–72.
32. The Long and the Short of DNA Sequencing, https://www.genengnews.com/gen-exclusives/the-long-and-the-short-of-dna-sequencing/77899725.
33. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. arXiv preprint arXiv. 2012;1207.3907 [q-bio.GN].

34. The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. Nature. 2010;467:1061–73.
35. van Dijk EL, Auger H, Jaszczyszyn Y, et al. Ten years of next-generation sequencing technology. Trends Genet. 2014;30:418–26.
36. Bio-IT World, http://www.bio-itworld.com.
37. Yuzuki D. Next-generation sequencing - its historical context, 2012, http://www.yuzuki.org/next-generation-sequencing-its-historical-context/
38. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. Arch Pathol Lab Med. 2017;141:1544–57.
39. SOLiD® Next-Generation Sequencing Systems & Accessories, https://www.thermofisher.com/us/en/home/life-science/sequencing/next-generation-sequencing/solid-next-generation-sequencing/solid-next-generation-sequencing-systems-reagents-accessories.html.
40. Kruglyak KM, Lin E, Ong FS. Next-generation sequencing and applications to the diagnosis and treatment of lung Cancer. Adv Exp Med Biol. 2016;890:123–36.
41. Rothberg JM, Hinz W, Rearick TM, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348.
42. Human Genome Sequencing Using Unchained Base Reads on Self-Assembling DNA Nanoarrays | Science, http://science.sciencemag.org/content/327/5961/78.long.
43. Koren S, Harhay GP, Smith TP, et al. Reducing assembly complexity of microbial genomes with single-molecule sequencing. Genome Biol. 2013;14:R101.
44. Thompson JF, Steinmann KE. Single molecule sequencing with a HeliScope genetic analysis system. In: Current protocols in molecular biology. Hoboken: Wiley; 2001.
45. Ashley EA, Butte AJ, Wheeler MT, et al. Clinical assessment incorporating a personal genome. Lancet. 2010;375:1525–35.
46. Hickman SE, Kingery ND, Ohsumi TK, et al. The microglial sensome revealed by direct RNA sequencing. Nat Neurosci. 2013;16:1896.
47. Bell DC, Thomas WK, Murtagh KM, et al. DNA Base identification by Electron microscopy. Microsc Microanal. 2012;18:1049–53.
48. Mankos M, Shadman K, Persson HHJ, et al. A novel low energy electron microscope for DNA sequencing and surface analysis. Ultramicroscopy. 2014;145:36–49.

49. NEWS, http://www.electronoptica.com/Electron_optica/NEWS.html.
50. Ari Ş, Arikan M. Next-generation sequencing: advantages, disadvantages, and future, in: plant Omics: trends and applications. Cham: Springer; 2016. p. 109–35.
51. Mankos M, Shadman K, N'Diaye AT, et al. Progress toward an aberration-corrected low energy electron microscope for DNA sequencing and surface analysis. J Vac Sci Technol B Nanotechnol Microelectron. 2012;30:6F402.
52. Mankos M, Persson HHJ, N'Diaye AT, et al. Nucleotide-specific contrast for DNA sequencing by Electron spectroscopy. PLoS One. 2016;11:e0154707.
53. Deamer DW, Akeson M. Nanopores and nucleic acids: prospects for ultrarapid sequencing. Trends Biotechnol. 2000;18:147–51.
54. Feng Y, Zhang Y, Ying C, et al. Nanopore-based fourth-generation DNA sequencing technology. Genomics Proteomics Bioinformatics. 2015;13:4–16.
55. Ansorge WJ. Next generation DNA sequencing (II): techniques, applications. Next Generat Sequenc & Applic. 2016;S1:005. https://doi.org/10.4172/2469-9853.S1-005.
56. Jain M, Olsen HE, Paten B, et al. The Oxford Nanopore MinION: delivery of nanopore sequencing to the genomics community. Genome Biol. 2016;17:239.
57. Mikheyev AS, Tin MMY. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour. 2014;14:1097–102.
58. Laver T, Harrison J, O'Neill PA, et al. Assessing the performance of the Oxford Nanopore technologies MinION. Biomol Detect Quantif. 2015;3:1–8.
59. Goodwin S, Gurtowski J, Ethe-Sayers S, et al. Oxford Nanopore sequencing, hybrid error correction, and de novo assembly of a eukaryotic genome. Genome Res. 2015;25:1750–6.
60. Navin N, Kendall J, Troge J, et al. Tumour evolution inferred by single-cell sequencing. Nature. 2011;472:90–4.
61. Gawad C, Koh W, Quake SR. Single-cell genome sequencing: current state of the science. Nat Rev Genet. 2016;17:175–88.
62. Kester L, van Oudenaarden A. Single-cell Transcriptomics meets lineage tracing. Cell Stem Cell. 2018;23(2):166–79.
63. Ortega MA, Poirion O, Zhu X, et al. Using single-cell multiple omics approaches to resolve tumor heterogeneity. Clin Transl Med. 2017;6(46):46.

64. Tang F, Barbacioru C, Wang Y, et al. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods. 2009;6:377–82.
65. Wang Y, Navin NE. Advances and applications of single-cell sequencing technologies. Mol Cell. 2015;58:598–609.
66. Picelli S. Single-cell RNA-sequencing: the future of genome biology is now. RNA Biol. 2017;14:637–50.
67. Macosko EZ, Basu A, Satija R, et al. Highly parallel genome-wide expression profiling of individual cells using Nanoliter droplets. Cell. 2015;161:1202–14.
68. Klein AM, Mazutis L, Akartuna I, et al. Droplet barcoding for single-cell Transcriptomics applied to embryonic stem cells. Cell. 2015;161:1187–201.
69. Hashimshony T, Senderovich N, Avital G, et al. CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. Genome Biol. 2016;17(77):77.
70. Sasagawa Y, Danno H, Takada H, et al. Quartz-Seq2: a high-throughput single-cell RNA-sequencing method that effectively uses limited sequence reads. Genome Biol. 2018;19:29.
71. Fan HC, Fu GK, Fodor SPA. Combinatorial labeling of single cells for gene expression cytometry. Science. 2015;347:1258367.
72. Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Massively parallel single-cell RNA-Seq for marker-free decomposition of tissues into cell types. Science. 2014;343:776–9.
73. Rotem A, Ram O, Shoresh N, et al. High-throughput single-cell labeling (hi-SCL) for RNA-Seq using drop-based microfluidics. PLoS One. 2015;10:e0116328.