Chapter 19
Existing and Emerging Molecular Technologies in Myeloid Neoplasms

Eric Q. Konnick and David Wu

The Era of Genomic Medicine

Genomic medicine has had a long and rich history. In the 1970s, the discovery of restriction endonucleases first allowed scientists to cleave DNA in a reproducible manner, allowing for the probing of specific alterations of DNA sequence at these sites [1]. As additional restriction endonucleases were discovered, simultaneous interrogation of multiple nucleotides at specific genomic positions soon became possible leading to key technological advances such as DNA cloning, DNA sequencing, and in situ hybridization.

The discovery of recurrent translocations in many neoplasms offered the opportunity for identification of these genetic structural rearrangements by molecular methods. Advances in cytogenetic staining made karyotype the initial test of choice for structural genomic alterations, but that method was highly specialized, laborious, and required viable neoplastic cells. Subsequent development of molecular cloning allowed for the introduction of in situ hybridization methods, first with radioactive elements [2] and later with fluorescence labels [3], which then allowed for rapid and specific identification of recurrent chromosomal alterations. Around this time, techniques to determine the sequence of DNA bases were developed by Frederick Sanger and colleagues using radiolabeled nucleotides [4–6]. Although the first use of these techniques was similarly laborious and time consuming, these approaches were critical in establishing foundational knowledge about the sequence and structure of genes. Other advances followed in rapid succession. Polymerase chain reaction (PCR) was invented in the mid-1980s by Kary B. Mullis [7] and was so revolutionary that clinical applications of this technology were adopted almost immediately thereafter [8]. The subsequent discovery and

E.Q. Konnick (✉) • D. Wu
Department of Laboratory Medicine, University of Washington, Seattle, WA, USA
e-mail: Konnick@UW.edu

C.-C. Chang, R.S. Ohgami (eds.), *Precision Molecular Pathology of Myeloid Neoplasms*, Molecular Pathology Library, DOI 10.1007/978-3-319-62146-3_19

application of thermostable enzymes in PCR were further transformational [9, 10] and allowed for novel automation solutions that propelled the field forward, as the combination of these different techniques was powerful. For example, PCR allowed for the rapid amplification of specific nucleotide sequences and when combined with restriction enzyme technology offered a quick and simple way to identify point mutations in neoplastic tissue through the characterization of restriction fragment length polymorphisms (RFLP) [11]. During the 1980s and 1990s, our understanding of the role of genes in disease transformed with improvements in PCR technology, Sanger sequencing methods, and enhanced computing power. During this period, numerous genes were identified and associated with specific neoplastic conditions [12, 13]. These technological advancements led to rapid implementation of PCR-based assays in clinical laboratories for diagnosing diseases with well-known genetic predisposition [14] and infectious diseases associated with cancers [15, 16], and for characterizing alterations of specific neoplasms [17, 18]. PCR-based assays were also described as a possible method of residual disease detection when the genetic abnormalities were characteristic for a given disease [19]. By the end of the twentieth century, a massive expansion of robotic technology and DNA synthesis techniques allowed for the development of DNA microarrays in which the genome could be probed using hundreds of thousands of probes to identify copy number changes, single nucleotide polymorphisms, and quantification of RNA transcripts. At the same time, advancements in fluidics, digital imaging, and computational power allowed for the subsequent development of methods that allowed for genome-wide sequencing of millions of short sequences in a massively parallel manner.

These scientific and technologic advancements in laboratory techniques allowed physicians and scientists to apply insights from molecular biology toward a detailed understanding of hematological malignances due to the relative ease of obtaining viable neoplastic cells and working with such cells under a variety of conditions. As such, the application of the state-of-the-art techniques allowed for hematological disease classification to always reflect the best clinical, scientific and analytical data available. Thus, whereas the French–American–British (FAB) Cooperative Group proposed a system for classifying acute myeloid leukemia (AML) in 1976 using morphologic and cytochemical criteria to characterize the presumed stage of differentiation of myeloid neoplasms, as scientific insight and laboratory techniques improved, disease classification was improved when it was recognized that specific genetic lesions were associated with some entities. Additional knowledge obtained using advanced techniques revealed the importance of specific genetic alterations in not only diagnosis but also prognosis and prediction of therapeutic efficacy. With new data informing the field, the World Health Organization (WHO) developed a new hematologic disease classification approach that included genetic, immunophenotypic, biologic, and clinical features along with morphologic features. This comprehensive approach defined specific entities with diagnostic, prognostic, or therapeutic significance. Since the original WHO classification in 2001 [20], there have been two major revisions due to an evolving understanding of hematologic disease, with the first occurring in 2008 [21], and more recently in 2016 [22].

In this chapter, we attempt to describe some of the key techniques and laboratory approaches that were important in the development of our modern classification of myeloid neoplasia. As we move forward, new laboratory techniques are likely to further refine our understanding of the pathogenesis of neoplastic myeloid diseases, impact our approach for disease classification of myeloid neoplasia, and hopefully offer insight into new treatment opportunities for patients.

Current Laboratory Methods in Common Use in Evaluation of Myeloid Neoplasia

Karyotype

Nobel Prize winning work in the early twentieth century demonstrated that the exchange of genes was related to the exchange of chromosomal material [23, 24]. This work formed the basis for additional studies and experimentation that eventually led to the formalization of cytogenetics and karyotype analysis as the first, "whole-genome" diagnostic tool several decades later. Since that time, cytogenetic techniques and our understanding of its importance in neoplasia have improved, and currently, karyotype analysis is a standard part of the clinical evaluation of many myeloid neoplasms. Although high-resolution, 2000-band karyotypes are available that can identify abnormalities at 1–2 megabases (Mb) resolution, current clinical practice in many institutions is limited to the routine use of a 300-to-500-band karyotype, which is capable of resolution of 7–10 Mb alterations [25]. This resolution nevertheless allows for discernment of large structural changes to be readily identified. Many myeloid neoplasms demonstrate recurrent cytogenetic abnormalities, but a significant proportion, such as acute myeloid leukemia (AML), do not have significant cytogenetic abnormalities [26]. Although a subset of myeloid neoplasms do not have detectable aberrations by karyotype, this method nevertheless provides valuable diagnostic and prognostic information and is commonly employed clinically.

Karyotyping consists of several steps including the growth of cells in culture, arrest of cells in metaphase, treatment of cells with a hypotonic solution, fixation of cells, dropping the cells across a glass slide to disperse the fixed cells, staining of the genetic material, imaging of the stained materials, assembling of a karyogram, and interpretation. Multiple different properties of chromosomes can be described using current karyotyping techniques, including alterations in the absolute number, size, centromeric position, structure, and banding patterns of chromosomes (Fig. 19.1). From these characteristics, cytogeneticists can assess for aneuploidy, structural changes, and presence of unknown genetic material in satellite chromosomes. Abnormal karyotypes in the setting of myeloid neoplasms have been extensively described. As the technique became commonly used in the clinical evaluation of myeloid neoplasms, key observations were made that certain changes were recurrent

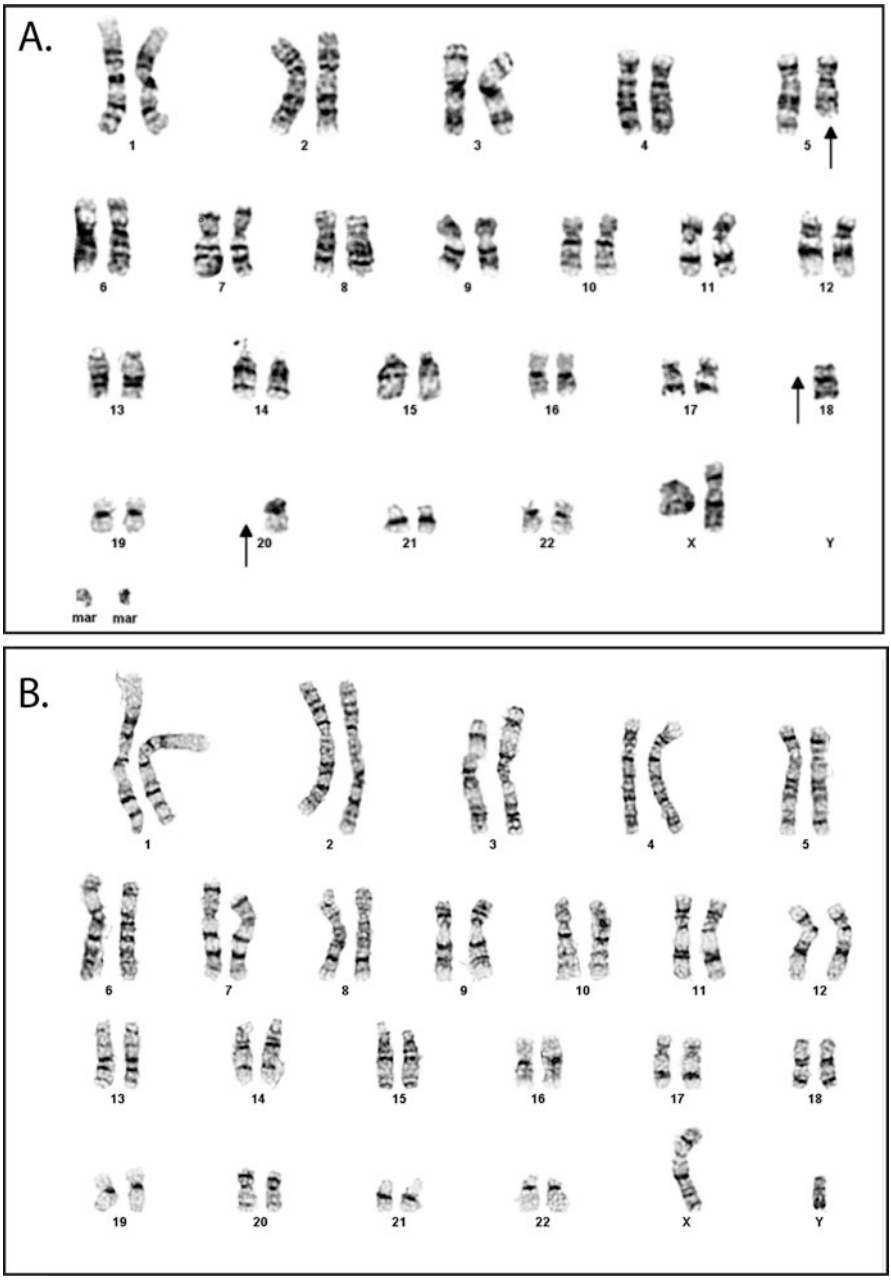

Fig. 19.1 Examples of karyograms representing karyotypes at 300–500 band resolution (resolution of 7–10 mB). (**a**) Abnormal female karyogram from a bone marrow aspirate of a patient with myelodysplastic syndrome. The complex karyotype demonstrates multiple abnormalities (*arrows*), including loss of 5q, monosomy 18, and monosomy 20 in addition to two marker chromosomes of undetermined origin (mar). (**b**) Normal karyogram from a peripheral blood sample from a healthy male

and seemed to be associated with specific neoplastic features [27–29]. Additionally, it became apparent that if serial monitoring were performed clinically in patients during disease progression, there may be concurrent evolution of the karyotype as well [30]. Advantages of karyotypes include the ability to discern whole-genome duplication events, identification of large-scale chromosomal abnormalities, and assessment of changes in the karyotype over time. Currently, the processes for karyotyping have become well established and are generally available in commercial reference laboratories and in most academic medical centers. The karyotype relies on living cells to culture, and yield of viable cells may be limited by chemotherapy exposure or sampling. The presence of subtle genetic rearrangements may not be readily apparent on common preparations, and the sensitivity of the method to subclonal populations is poor. Important genetic changes such as substitution mutations or loss of heterozygosity cannot be resolved using karyotype analysis. Although karyotype has been described as a method for monitoring minimal residual disease, more sensitive methods are generally preferable, when possible [31]. Frequent abnormalities in karyotype are noted in several types of myeloid neoplasia [20, 21], and as such, karyotypic aberrations will likely continue to be considered in description and classification of myeloid neoplasms for many years to come.

Fluorescent In Situ Hybridization

The cloning and restriction enzyme techniques that were developed in the 1970s allowed for a new set of tools that enabled detailed identification of structural and copy number changes in patient material. The first methods describing florescence in situ hybridization (FISH) were published in the early 1980s [32] and rapidly developed into a robust technique with numerous permutations. The technique is conceptually simple, consisting of hybridizing fluorescently labeled oligonucleotide probes complimentary to a sequence of interest to target-specific loci over the course of hours, washing away unbound probes, imaging the slide, and characterizing the binding pattern apparent in an appropriate number of cells. FISH can be used to enumerate specific types of genetic changes and rearrangements that occur in the setting of neoplasia. Assessment of copy gains or losses of whole chromosomes can be inferred using probes targeting the centromere of the chromosome of interest, while using a centromere probe combined with a probe for a gene of interest will allow the operator to determine if a specific gene is gained or lost or if such an event is due to a chromosome-level event. One of the most powerful permutations has been the assessment of chromosomal translocations or other structural rearrangements that are recurrent in neoplasia. Some of the commonly used methods employ fusion probes or break-apart probes. Fusion probes utilize two differently fluorescently-labeled probes that target genes involved in a translocation or other structural rearrangement. When the probes are spatially separated, the probes individually fluoresce at different wavelengths allowing the operator to discern separate signals for each probe. However, when the probes are in close proximity due

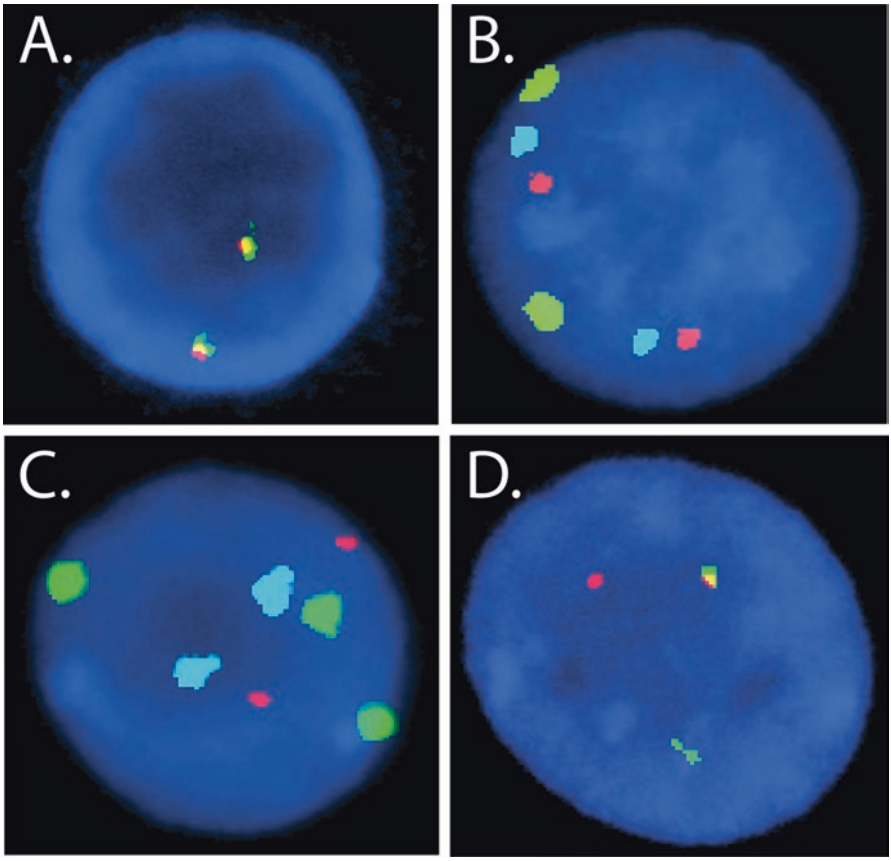

Fig. 19.2 Examples of FISH studies from clinical samples demonstrating different types of FISH design. (**a**) Break-apart probe assay with intact probes. (**b**) No abnormalities – FISH assay using three probes: chromosome 12 centromere (*green* – two signals), 13p (*aqua* – two signals), 13q (*orange* – two signals). (**c**) Trisomy 12 – FISH assay using three probes: chromosome 12 centromere (*green* – three signals), 13p (*aqua* – two signals), 13q (*orange* – two signals). (**d**) Break-apart probe with single *red* and *green* signals representing the separated break-apart probes and a single *yellow* signal representing the intact locus (Images courtesy of Debra Dehoogh-Grigsby, University of Washington Cytogenetics Laboratory)

to a translocation event in which two different genomic loci are fused together, the signals can no longer be separately distinguished and the operator only can identify a composite signal of a color that is distinct from the individual fluorophores. In contrast, break-apart probes use a similar concept but consist of two probes that are on either side of a common translocation breakpoint of a single gene that is involved in the setting of structural rearrangements. In this design, when a given gene is intact, only a single, composite signal per chromosome is observed. When a gene is rearranged across a breakpoint that is between the two probes, the signals separate and either two separate-color signals or a single-color signal will be observed per disrupted chromosome (Fig. 19.2).

Fusion and break-apart probe designs have an inherent specificity built into their design, in which it is unlikely that a false-positive result will occur. When testing clinical formalin-fixed paraffin-embedded (FFPE) samples, it is possible that a subset of the evaluated nuclei will have been cut in such a way that the genomic material being probed will be present on separate planes and, as such, there will generally be a background level of artifactual "abnormal" signals that needs to be considered when the assay is evaluated. Fusion probes are a good design to use when rearrangements are observed between the same two partners in the setting of interest [33], but if a rearrangement occurs between one of the genes and an unknown partner, there is a possibility of a false negative result. One advantage of the break-apart FISH design is that it is agnostic to the partner of the interrogated gene, which is beneficial in diseases in which there are multiple possible rearrangement partners or when the main concern is identification of the presence of gene disruption and not the specific translocation partners. The disadvantage of such a design is that the knowledge of the second gene involved in the rearrangement is not known.

In general, FISH has several advantages, thereby making it useful for assessing large-scale genomic events in neoplasia. The detected signals can be very specific, the testing can be relatively inexpensive, rapid, and the assay can be quickly done on fresh, cultured, or FFPE samples. FISH studies on FFPE samples can often be accomplished in a much more rapid fashion than alternative techniques such as reverse-transcription PCR (RT-PCR) or next-generation sequencing (NGS). A key shortcoming of FISH, however, is that only a few targets can be evaluated per test, requiring selection of the appropriate probes before the test is performed. While such selection is often acceptable, rare cases where incorrect testing is performed can result in false-negative results that have important clinical implications. In cases where multiple targets need to be assessed for diagnostic, predictive, or prognostic uses, panels of simultaneous assays are often employed, and with each additional assay performed, the cost of testing increases linearly as does the labor of evaluating hundreds of cells per study for interpreting the results. In the setting of rearrangements, there are examples of cryptic rearrangements that may not be readily detected by standard FISH methods [34]. Finally, FISH can only identify large-scale genomic changes such as copy number alterations and structural rearrangements, and the technique will not detect small-scale mutations such as single-nucleotide variants and small insertions and deletions (indels). However, with proper understanding of the advantages and limitations of FISH and specific permutations, this technique is a rapid, sensitive, specific, and valuable method for assessing important and recurrent genomic events in the setting of neoplasia, particularly in the clinical laboratory.

Sanger Sequencing

The initial description of sequencing by Frederick Sanger and colleagues used the concept of chain terminating nucleotides labeled with radioactive tracer molecules [4–6]. Separate reactions were used in which each reaction incorporated only a

single terminator nucleotide with a larger proportion of unlabeled nucleotides. After the sequencing reactions were completed, the four reactions were separated in separate lanes on an electrophoretic gel and visualized by autoradiography. As labeled terminator nucleotides were incorporated into a growing DNA strand, the strand would not be able to extend further and the base at that position could be inferred based on the position on the gel. This technology allowed the determination of each subsequent base in a sequence by incorporating nucleotides, specifically dideoxynucleotides that prevented further elongation of the DNA molecule and visualizing on an autoradiographic polyacrylamide gel after size separation.

With the introduction of PCR, Sanger sequencing became technically easier, but it was the introduction of fluorescently labeled nucleotide terminators and automation that greatly expanded the use of this technology, particularly in the clinical environment (Fig. 19.3). By utilizing separate fluorophores for each of the terminator nucleotides, a single reaction for the forward and reverse PCR product decreased the number of reactions necessary to obtain sequence data. Automated instruments,

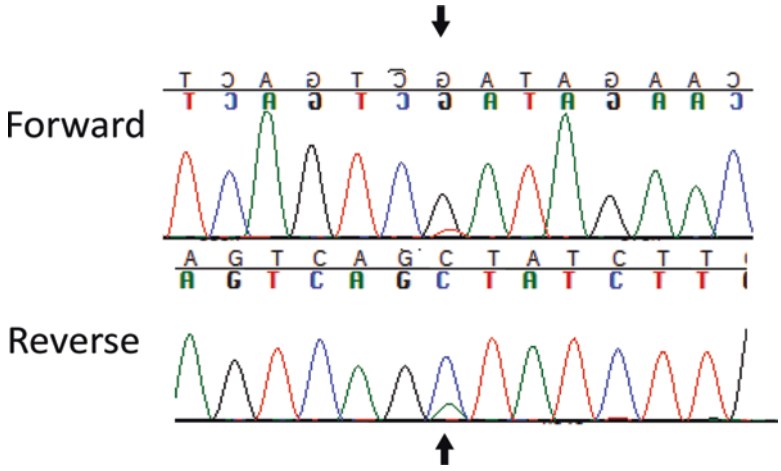

Fig. 19.3 Screen capture of Sanger sequencing trace utilizing fluorescently labeled dideoxynucleotide terminators. In this method, a single reaction tube is required for the forward and reverse sequencing reaction for each sequenced region. Reactions include dideoxynucleotide terminators, each base labeled with fluorophores that emit at different wavelengths. After the sequencing reaction, the sequencing products are denatured and loaded onto an automated instrument which separates the DNA fragments based on size through the application of an electric field applied to a polyacrylamide gel or a sequencing capillary. Smaller-size DNA molecules move more quickly through the matrix and larger products more slowly. A laser or other excitation source and fluorescent detector are positioned at a fixed point along the electrophoresis apparatus. As DNA fragments pass by the detector window, the laser excites the florescent dye, which then returns to its resting state and emits a photon at a longer wavelength. This emission is captured through the detection apparatus and software determines the incorporated bases. Sequence data are represented by an electropherogram, which shows the fluorescence peaks of the labeled dideoxynucleotide terminators. Determination of the base pair composition can be accomplished with software algorithms, which allows for increased throughput. In this example, a G > T nucleotide substitution is identified (*arrows*) in approximately 15–20% of sequencing reads at the indicated position

initially as polyacrylamide slab-gel machines, and later capillary sequencers, allowed for the automation of electrophoresis and data acquisition [35, 36]. Developments during this time resulted in the ability to obtain DNA sequences of 500–1000 base pairs in a rapid and automated fashion.

With the rapid technological developments in the practice of Sanger sequencing, the technology had matured to the point that large-scale sequencing could be considered, and from 1990 to 2003 the Human Genome Project was undertaken to sequence the majority of several reference genomes obtained from volunteers [37]. The demands of this project resulted in numerous innovations in robotics, automation, and informatics that were rapidly adopted beyond the initial research initiative [38, 39]. In the late 1990s, the government-run effort was challenged by a private consortium using a novel approach of so-called, shotgun sequencing, which sheared DNA into random fragments and then used techniques to attach primers and sequence the intervening DNA [40, 41]. This technique relied heavily on the ability of computers to reassemble the DNA sequence by identifying regions that overlapped with one another. This situation led to a competition between the groups, which evolved into a collaborative effort resulting in the release of the first draft of the human genome in 2001 [42, 43].

Sanger sequencing is widely distributed and many laboratories are capable of generating high-quality sequence for research or clinical use with a high likelihood of success. This technique has some limitations, primarily that the technique can generally detect only minor sequence populations that are greater than 20% of the mixture, although recent developments suggest that more sensitive approaches may be available. While the output of these techniques has improved dramatically over the decades, the technology has limited throughput, often requiring dozens of reactions to fully analyze the exons of a single gene. This limitation makes large-scale analysis of genes difficult. Finally, while the informatics tools have improved dramatically, the commonly used methods still require a large amount of manual effort to review sequence data.

Polymerase Chain Reaction

Polymerase chain reaction (PCR) was invented by Kary Mullis in the mid-1980s (Fig. 19.4) [7] and clinical uses of this powerful technique were implemented almost immediately [8]. PCR is accomplished through the combination of extracted sample DNA with oligonucleotide primers flanking the sequence of interest, deoxynucleotide triphosphates (dNTPs), a polymerase enzyme, buffers to allow enzyme function, and cations required for polymerase function. The reaction mixture is heated to denature the double-stranded DNA (denaturation) and then cooled to a temperature that allows binding of the primer oligonucleotides (primers) to the template strands. Because primers are much shorter than the template DNA, reaction kinetics favor the binding of these primers to the template molecules. In addition, primers are added in molar excess, which further favors the binding of primers to the

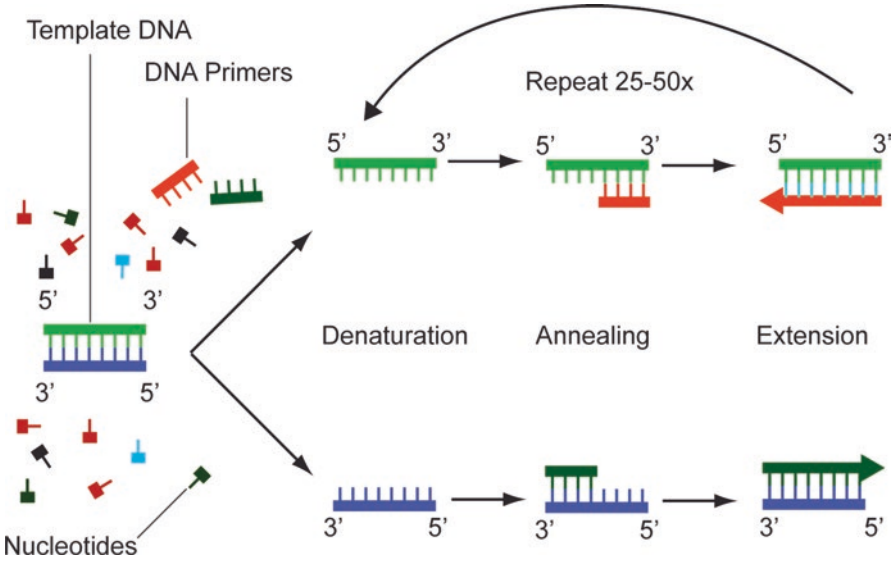

Fig. 19.4 Schematic showing the key steps of the polymerase chain reaction. Template DNA is incubated with synthetic oligonucleotide primers specific for the sequence of interest, with dNTPs, and a thermostable *Thermus aquaticus (taq)* polymerase. Template DNA is denatured using high temperature and the reaction mixture is cooled to allow annealing of oligonucleotide primers. Primers are extended by *taq* polymerase. The process is repeated, with the amount of template DNA effectively doubling with each reaction cycle

template DNA molecules. A key to proper PCR assay design is the incorporation of one primer on each strand of the target sequence. After annealing, the reaction temperature is raised to the optimum temperature for the polymerase enzyme, which then incorporates deoxynucleotide triphosphates (dNTPs) into a complimentary DNA molecule. The process is then repeated for multiple cycles, with a theoretical doubling of PCR product with each cycle. As the cycles progress, DNA molecules created in the PCR process (amplicons) become the predominant species, with the size of the amplicon being determined by the positioning of PCR primers. The cycling process is repeated 25–50 times, creating billions of copies of the target sequence, which can be detected or further analyzed by additional techniques.

Initial implementations utilized polymerase enzymes that were not heat stable, requiring addition of fresh polymerase enzyme after each denaturation step, requiring large amounts of enzyme and constant attention by operators. However, the introduction of thermostable polymerases [9, 10] allowed for automation and rapid dissemination of the technique. PCR allowed for the rapid amplification of specific nucleotide sequences and, when combined with restriction enzyme technology (RFLP), offered a relatively quick and simple way to identify point mutations in neoplastic tissue [11]. During the 1980s and 1990s, PCR-based techniques had broad applications in the setting of myeloid neoplasia testing and they were widely used. Numerous adaptations of the core PCR method have been made to facilitate

increased flexibility, automation, and specificity. Some examples of these methods include the use of fluorescently labeled, sequence-specific probes that allow for identification of specific mutations or fluorescently labeled primers that allow for accurate sizing of PCR products that are amenable to analysis on capillary sequencers [44]. A key advantage of PCR-based techniques is the exquisite sensitivity of the method, which enables detection of rare transcripts with excellent specificity in the appropriate context, making such methods excellent approaches for minimal residual disease monitoring [45]. A weakness of PCR, however, is that it is best suited to situations in which targets have a limited spectrum of possible mutation sequences and in which primer-binding sequences are likely to be constant across all targets, neither of which is guaranteed in the setting of neoplasia.

Reverse-Transcription Polymerase Chain Reaction

Reverse-transcription PCR methods have been developed to interrogate RNA molecules, first by using a reverse transcriptase enzyme to convert RNA to DNA and then by using conventional PCR to amplify the target of interest. This general approach has been extensively adapted, similar to PCR, to allow for numerous variations and automation to improve accuracy and throughput. Common uses for these techniques include fluorescently labeled, sequence-specific probes that allow for monitoring and quantitation of RT-PCR products that allow for rapid reporting of results [46] and interrogation of fusion transcripts to aid in diagnosis or monitoring of minimal residual disease [47–51]. Indeed, RT-PCR serves as important clinical tool for the evaluation of many chromosomal aberrations in routine clinical practice.

Microarray Testing

At the turn of the century, there was an explosion of technical innovation that allowed for extensive probing of the genome at a much higher resolution than was previously available using karyotype or FISH. The concept of a DNA microarray originated as part of dot–blot methods in which one or more nucleic acid probes, specific for known nucleotide sequences, were adhered to membranes and sample DNA was allowed to hybridize against these probes. If there was a sequence complementarity of the sample to the target, the sample nucleic acid would remain bound to the probe and thus the membrane [52]. Using a variety of detection methods, including first radioactive and subsequently nonradioactive methods, the sequence of interest in a given sample could be inferred. The development of large-scale methods for cloning and oligonucleotide syntheses, combined with advances in robotics, allowed for a technological shift from membranes dotted with probe sequences to increasingly dense arrays of sequence-specific oligonucleotides

arrayed on solid substrates [25, 53, 54]. Current iterations of this technology allow for hybridization of DNA, RNA, or more complicated substrates, such as protein–nucleic acid complexes, to the dense arrays that can then generate signals, indicating if a given probe region has increased or decreased target binding compared to a reference sample. Data acquisition and translation are accomplished through the use of automated imaging and computational analysis.

Initial iterations of microarray technologies used differentially labeled DNA from the test sample and a well-characterized reference sample to compare the differences in signal between the two samples that were interpreted as the relative copy number of a specific genomic region on the test sample. After labeling and normalization, the products are hybridized to a solid substrate, allowed to equilibrate, and then residual unbound material is removed prior to imaging. Imaging allows the assessment of the relative signals of the test sample and the control sample, and in the situation in which equal amounts of DNA are present from each sample, the signal is interpreted as the two samples having equivalent genomic material. By contrast, when there is either a copy gain or loss of genetic material in the sample relative to the reference, the ratio between signals is greater than or lower than 1, respectively (Fig. 19.5a). With refinement of the technology, arrays were developed with a higher density of probes, allowing assessment of increasing numbers genomic loci, such that comparative samples were no longer necessary. Further, with knowledge of the human genome provided by the Human Genome Project, increased probe density allowed microarray platforms to identify copy-number variation with greater resolution, enabling the routine identification of genomic sites of microdeletion and gains.

Current copy-number arrays (CNAs) are capable of detecting gains and losses as small as 1 kilobase and as large as up to megabases in size and were critical for identifying normal copy number alterations that occur in the genome [55]. Recent advances have allowed the inclusion of probes capable of resolving single nucleotide variants at specified positions, which allows the assessment of single-nucleotide

Fig. 19.5 (continued) amplification. (*2*) The prepared sample is hybridized onto a microarray "chip," allowing the sample DNA to bind specific spots with a prelabeled and known DNA sequence. (*3*) A higher magnification view of the individual spots on the microarray chip, when DNA is bound it results in a signal of varying intensity depending on the amount of bound DNA. (*4*) A simplified diagram of how the specific probes may be arranged on a given microarray chip. In this schematic, the CNA probes are placed in the lower three rows, while the SNP probes are placed in the other rows. The corresponding signals are seen with differential amounts of DNA binding. (**b**) Example of data from a high-resolution genomic microarray assay with both CNA and SNP probes. This view allows for visualization of the entire genome from chromosome 1 on the left to the gender chromosomes on the far right. There are two rows of data, the top row being the data generated from the spots with copy number probes, and the lower row generated from data with the SNP probes. (**c**) Example of virtual karyotype generated from high-resolution genomic microarray assay with both CNA and SNP probes. In this example, copy gains are highlighted in blue and shown to the right of the respective chromosome (1q), copy losses are shown in red and shown to the left of the respective chromosome (17p, 18q), and regions of copy neutral loss of heterozygosity (cnLOH) are highlighted in orange on top of the chromosome (4q)

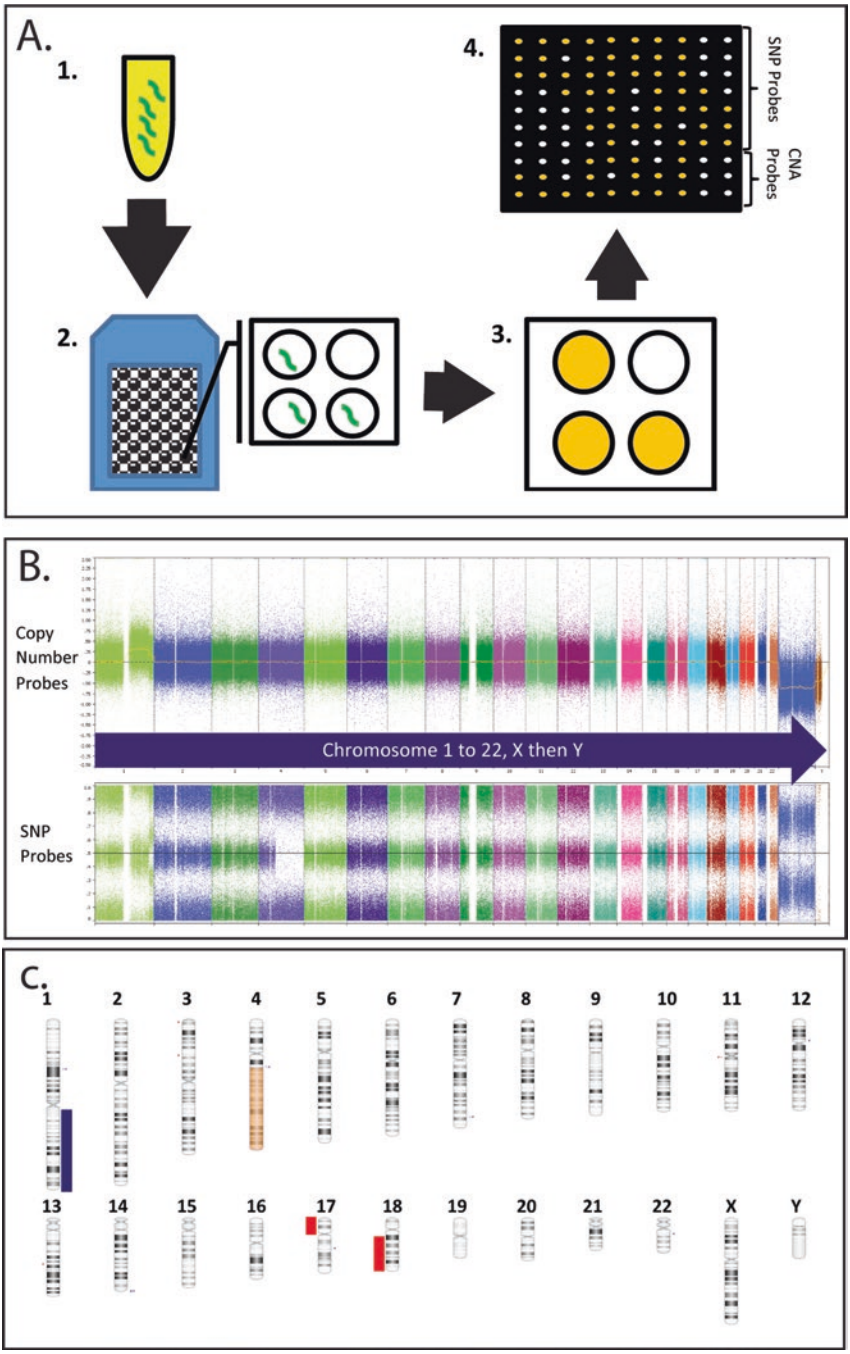

Fig. 19.5 (**a**) A simplified schematic of newer generation high-resolution DNA microarray chips, many of which feature copy number alterations (CNA) and single nucleotide polymorphism (SNP) probes. (*1*) Sample DNA is processed through a number of preparative steps usually including

polymorphisms (SNPs) in samples. Such SNP arrays are useful in determining if copy neutral loss of heterozygosity (cnLOH) is present in samples, which is previously not detectable by other cytogenetic techniques. Copy-number and SNP arrays can be combined in a single platform to allow for assessment of copy number variants (CNVs) and LOH (Fig. 19.5b). Microarrays combining CNA and SNP probes are, therefore, able to provide high-resolution "virtual karyotype" with added cnLOH data that enable surveillance of the entire human genome for chromosomal aberrations (Fig. 19.5c). Various techniques and technologies have been applied to microarrays to allow the platform to interrogate different starting materials, such as DNA, messenger RNA (mRNA) expression, microRNA, protein–DNA complexes, and epigenetic modifications. DNA microarrays are highly customizable and custom arrays can be developed easily, allowing for customized, targeted platforms for research or clinical uses.

Microarrays are powerful tools that can be employed in the assessment of chromosomal losses and gains in the setting of myeloid neoplasia. The technique is capable of defining the chromosomal positions of gains and losses much more precisely than karyotype or FISH. Such information can leverage information on gene position to identify specific genes that are gained or lost in a given sample. SNP arrays are able to identify cnLOH in samples that appear karyotypically normal [56, 57], suggesting genes and genomic regions that may be important to disease pathogenesis. Automated instrumentation and software tools allow users to process the incredible amounts of data that can be generated by a microarray, but often the knowledge of specific regions that appear to have CNVs or LOH remains to be characterized. The technique is also not able to identify structural rearrangements that can commonly be identified using karyotype or FISH techniques, although some modifications may allow detection of balanced translocations [58]. The platform is sensitive to the proportion of sampled cells that possess the abnormal genomic complement, typically exceeding 20% of the sampled cells, limiting the use of the technique in evaluating subclonal populations or assessment of minimal-residual disease (MRD). While the technology is capable of identifying SNPs, it is not feasible to resolve all possible specific single-nucleotide mutations possible in a neoplastic genome. DNA microarrays, as currently available, complement other genomic techniques and offer a more granular interrogation of the genome than karyotype or FISH studies.

Massively Parallel Sequencing

At the turn of the century, the first descriptions of massively parallel sequencing were published [59, 60]. Although these were different in approach and chemistry, their commonality was the ability to sequence millions of short DNA sequence reads, in parallel, without requiring a priori knowledge of the sequence. This technology has since been termed, massively parallel or "next-generation sequencing"

(NGS). These techniques rely on fragmentation of the DNA to be sequenced and ligation of common DNA sequences, which then allows amplification and enrichment of the sequences after a single DNA molecule isolation step, which are analogous to techniques used in shotgun sequencing approaches. Simultaneous sequencing of millions of small fragments of DNA is then accomplished in a parallel fashion, generating billions of bases of sequence data. The massive amount of short-read data require advanced computational approaches to assemble the data and align the sequences against the reference human genome, which was one product of the Human Genome Project. From the initial descriptions of the techniques in a research setting, the technology has evolved and become commercialized by multiple entities. NGS techniques have also proliferated, with numerous adaptations of the methods to allow for diverse applications, and with the maturation of the systems, there has been a rapid adoption of the techniques in research and clinical laboratories. As costs continue to decrease and analysis tools become more sophisticated, these techniques continue to find new uses and increasingly compete to displace some established techniques in clinical and research applications.

Of the published methods that have been employed in both research and clinical laboratories, there are several that have been widely used that warrant description. These approaches include sequencing by synthesis, pyrosequencing, sequencing by ligation, and semiconductor sequencing [61, 62]. Improvements in these techniques continue to occur and most certainly, new technologies will augment these approaches, such as demonstrated by recent advancements of third-generation technologies utilizing nanopore technologies. Each platform has strengths and limitations that must be considered. A brief description of these approaches is included below. However, this represents only a survey of the technologies and should not be considered a comprehensive review (Table 19.1).

In general, techniques that are widely used have the key common steps: isolation of genetic material, fragmentation of the genetic material into specific size ranges, ligation of sequence adapter molecules with known sequence (“barcodes”), enrichment for sequences of interest, massively parallel sequencing, bioinformatics pipeline data analysis, variant calling, and variant annotation. The use of bar-coded sequences, simply strings of nucleotides, allows for multiple samples to be sequenced together, with individual sequences attributed to specific samples through assessment of the barcode sequences through a demultiplex algorithm.

An innovation that allowed the use of NGS technologies was the implementation of bioinformatics pipelines that made the analysis of billions of bases of sequencing data a manageable task. These data analyses pipelines generally consist of multiple, separate, computer programs, which are linked together using additional programs such that sequence data can flow from one program to another or be analyzed for different features by different programs, either in a serial or parallel manner. After demultiplexing to separate and assign sequencing reads to respective samples through the use of sample-specific barcodes, individual sequences can be aligned against a reference genome that allows for the identification of variants

Table 19.1. Comparison of key features of commonly encountered sequencing platforms

Method	Manufacturer	Read length (base pairs)	Accuracy, %	Time per run	Cost per 1 million bases (US$, approx.)	Advantages	Disadvantages
Sequencing by synthesis	Illumina/Solexa	100–300	98	Varies (1–10 d)	$0.07	Potential for high sequence yield, depending upon sequencer model and desired application. Multiple scales of sequencing available.	Expensive equipment. Short read lengths.
Ion semiconductor	Ion Torrent	≈400	98	2 h	$1	Less expensive equipment. Fast.	Homopolymer errors.
Chain termination (Sanger)	Thermo Fisher (ABI)	Up to 1000	99.9	20 min −3 h	$2400	Long individual reads. Useful for many applications.	High cost, low throughput. Low sensitivity to minor variant populations.
Nanopore	Oxford Nanopore	10,000+	92–95	Minutes to hours	Unknown	Compact instrumentation, long read lengths.	Lower accuracy, unknown cost, limited distribution. Frequent changes in hardware and chemistry. Requires high-quality DNA.
Single-molecule real-time sequencing	Pacific Biosciences	10,000–15,000+	85	30 min – 4 h	$0.13–$0.60	Longest read length. Fast.	Low accuracy. Moderate throughput. Expensive equipment. Requires high-quality DNA.
Pyrosequencing	454/Roche	700	99.9	24 h	$10	Long read size. Fast.	Homopolymer errors, cost. Discontinued.
Sequencing by ligation	SOLiD	50 + 35 or 50 + 50	99.9	1–2 week	$0.13	Accuracy, low cost per base.	Short read assembly, difficulties with palindromes, slow.

using bioinformatics programs referred to as variant callers (Fig. 19.6). In addition to programs that can identify single nucleotide changes from the reference sequence, additional computational approaches have been developed to detect insertion/deletions (indels), copy number variants (CNVs), and other structural alterations such as translocations [63]. Comparison studies have demonstrated that CNVs identified using some NGS assays show equivalency with results obtained using karyotype [64] or microarray [65], suggesting the possibility that NGS may be able to generate results similar to these well-established platforms.

Research groups and software developers are constantly producing new software packages and modifying existing programs to allow for improved performance and addition of new algorithms when possible. The cost of computational software, hardware, and the necessary technical expertise to implement such data analysis approaches is a critical consideration when developing and supporting NGS-based assays that should be considered from the outset [66]. While the hardware and software infrastructure required to support bioinformatics pipeline data analyses is a significant expense, the costs associated with maintaining the raw data acquired from NGS assays can quickly eclipse the initial sequencing expense, as the raw data can exceed hundreds of gigabytes per batch. Thus, it is not only a challenge to deal with the initial bolus of data but special consideration must be made for long-term storage of data because if such data are to be maintained over a long period of time as the resources for maintenance may exceed the initial cost of generating the data [67]. Researchers and clinical laboratories must therefore thoroughly evaluate and optimize their data analysis and storage methods. As this technology transitions from research laboratories to clinical laboratories, the pathology and laboratory medicine professionals who will most likely be responsible for clinical implementation and interpretation of such methods will require new skills and training programs in order to provide safe and effective tests for patient care [68, 69]. Although new training programs will enhance the abilities of new professionals, the wide range of experience in professionals who have previously completed training will likely benefit from decision support systems that can aid in the selection, interpretation, and reporting of genomic tests, as well as treatment decisions [70–72].

Commonly Used Platforms for Next-Generation Sequencing

Multiple platforms have been developed to take advantage of the concept of massively parallel sequencing. Although multiple different approaches have been commercialized, the field has consolidated into a few dominant platforms. New sequencing technologies are always in development that promise to offer advantages over current technology, but are still in their infancy in comparison. In this next section, the main sequencing platforms used in current clinical testing and research are described, including several platforms that were previously widely employed and may be encountered in the literature.

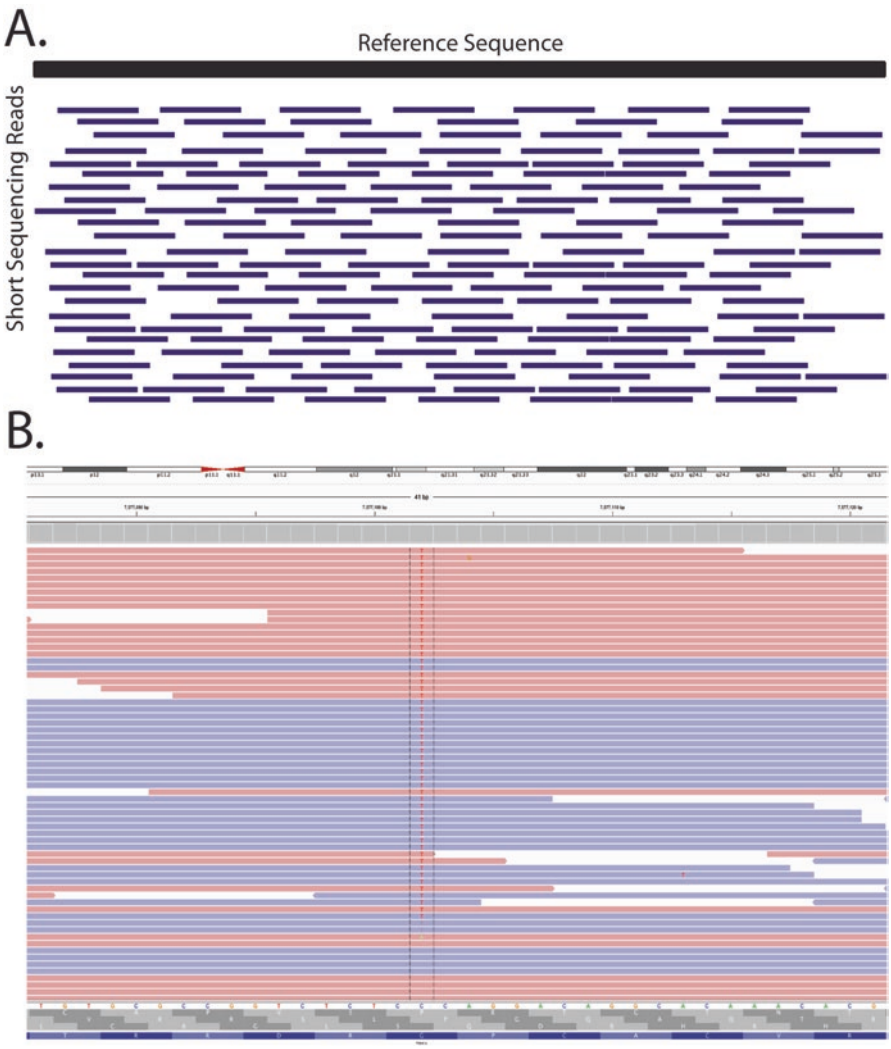

Fig. 19.6 (**a**) Schematic showing the alignment of numerous short DNA sequencing reads to a template. Note the overlap in sequence from multiple short-read length fragments with often unique start and stop sites. When numerous fragments overlap at a given position, low-frequency variants may be identified in a small subset of the reads. Conversely, when there is low coverage of a given region of DNA, rare variants may not be detected because they represent the minor population of the sample. (**b**) Screen capture of clinical next generation sequence data as displayed using the Integrated Genomics Viewer (Broad Institute, Cambridge, MA, USA). The highlighted base is a variant that is identified in the *TP53* gene (p.G279E, NM_000546.5:c.836G > A)

Sequencing by Synthesis (Solexa/Illumina)

The sequencing-by-synthesis (SBS) method, pioneered by Solexa, is a hybrid method combining fluorescent dideoxy Sanger sequencing and solid surface sequencing by synthesis [73, 74]. The use of fluorescently labeled, reversible dideoxy terminators [75] allows for the sequence-data acquisition to be decoupled from the sequencing chemical reaction. In this method, denatured DNA fragments with ligated adapters are flowed across a glass substrate ("flow cell") with oligonucleotides complimentary to the adapter molecules added to the DNA of interest during the library preparation stage. If the concentration of the sequencing library is correct, individual DNA molecules become spatially distributed on the flow cell, which will allow accurate interrogation at the sequencing stage. The attached sequences are then amplified in place using PCR such that products of the reaction are also attached to the solid substrate via the capture oligonucleotides, thus forming cluster of identical DNA fragments generated from the same initial library DNA fragment. Postamplification, fluorescently labeled, reversible terminators are added in bulk to the flow cell, a single base is incorporated into each bound molecule, and the residual reagent is washed away. The fluorescently tagged bound DNA molecules are excited via a laser, and digital imaging is used to scan the flow cell and identify incorporated bases in each DNA cluster based on fluorescent signal. After imaging, the reversible terminator is then cleaved from the captured DNA molecules, allowing for the incorporation of another fluorescently labeled nucleotide on the next cycle. Depending on the instrument, reagents, and DNA quality, the cycle can be repeated up to 150–300 times and then the unique indices are sequenced to establish the identity of each imaged cluster. After removing the sequencing products from the first round, a second round of amplification allows the sequencing of the captured molecule from the reverse direction and the confirmation of the identity of the cluster by sequencing a second index. A critical innovation of this technology was the decoupling of the enzymatic sequencing synthesis reaction from the acquisition of base identity, which allows for huge arrays of DNA library molecules to be sequenced simultaneously. This method has been shown to be susceptible to decreased read depth with increasing GC content [76–78] and may be susceptible to bias introduced via the multiple PCR amplifications utilized at different steps in the method [79]. Comparisons of available technologies have suggested that the SBS method has the highest sequencing throughput per batch coupled to the lowest error rates [80]. The need for specialized hardware and optics to identify the sequencing products, stability of the reagents, and the complicated fluidics has been cited as potentially limiting the read length and accuracy of the method [81]. Improved engineering, hardware upgrades, and innovation of software are likely to continue to improve the performance of this approach in the future. Currently, this technology dominates many of the research and clinical methods that have been used to evaluate the spectrum of variation that occurs in the neoplastic disease.

Semiconductor-Based (Ion Torrent)

Semiconductor-based sequencing also relies on the combination of a solid substrate combined with numerous individual reactions targeting individual DNA molecules. In this method, microscopic reaction wells have been created with a semiconductor at the base of the reaction chamber. After DNA library preparation, the reaction device is flooded with droplets containing single strands of DNA [82]. Subsequently, the reaction device is sequentially treated with a single nucleotide, and if a given nucleotide is incorporated into a nascent strand of DNA, a hydrogen ion is liberated as part of the reaction. The released hydrogen molecule is then detected using a sensitive ion sensor located in the individual reaction well. A unique aspect of this design is that base incorporation is not limited to a single nucleotide, such that in the case of a homopolymer repeat region, multiple nucleotides will be incorporated in a single cycle. Such incorporation of multiple nucleotides results in a corresponding increase in the number of released hydrogen atoms, which then result in a proportionally greater electronic signal detected by the sensor. This aspect of the method can cause difficulty in assessing the true number of nucleotides within a homopolymer region and can result in a false-positive determination of insertion–deletion (indel) events. Nevertheless, numerous approaches have been developed to decrease the incidence of false-positive indel calls in commonly used bioinformatics pipelines, which may be useful in clinical laboratory settings [83–85]. A potential advantage that is often cited is the possibility of semiconductor sequencing to decrease in cost and improve in performance because the core technology is able to leverage the scale and infrastructure of the semiconductor industry [86]. Comparisons of semiconductor NGS methods to alternative technologies have demonstrated higher throughput than other methods [85].

"Third-Generation Sequencing" Sequencing Methods

The majority of currently available NGS approaches that are in wide research and clinical use require one or more amplification steps where the nucleic acid sequences of interest are enriched before the actual sequencing reaction. This reliance on amplification prior to sequencing can have consequences such as bias and sequencing artifacts, which may limit the applicability in some situations. Additionally, the current generation of technologies in common use relies on interrogation of relatively short DNA sequences, which can hinder the evaluation of classes of mutation that are important in the setting of neoplasia, such as structural rearrangements or assignment of variants to a pseudogene. These factors have led to the development of computational methods to aid in the detection of such alterations [87–89]. Currently, several methods are available that sequence individual nucleic acids using approaches that allow contiguous sequence read lengths of several thousand kilobases or more. Although these methods have lower throughput compared to the commonly used platforms, it appears that fewer long-read sequences may result in

improved assemblies compared to short-read sequencing at higher depths [90, 91]. Additionally, these so-called third-generation sequencing technologies have been suggested to potentially further decrease sequencing costs [92], reduce bias in sequenced regions due to elimination of amplification steps, and improve sequencing of high GC-content DNA sequences [93]. In order to maximize the length of individual DNA molecules derived from a sample, high-quality samples and high-quality DNA extraction techniques must be employed in order to maximize the strengths of long-read nanopore sequencing approaches.

A major drawback of common NGS methods is the necessity for amplification of the source DNA prior to DNA sequencing. This approach can lead to amplification artifacts and biased coverage of the genome related to the G–C content or local structural properties of specific DNA sequences. Current NGS technologies produce relatively short reads, with median sequence lengths ranging from 100 to 300 base pairs (bp) for some methods (such as those by the manufacturer, Illumina) and up to 700 bp for the approach championed by the technology by 454. Short sequence reads are in general more difficult to align to a reference genome, especially in regions of high homology or in the context of pseudogenes [94, 95]. While short-length DNA sequencing may work approximately equally well with fresh or formalin-fixed, paraffin-embedded (FFPE) tissues, long-read sequencing of thousands of bases is unlikely to be successfully accomplished using the fragmented and lower-quality DNA found in clinical tissues due to formalin fixation. Use of abundant, high-quality DNA as is commonly found in hematological samples may make this sequencing approach an attractive research and diagnostic sequencing method in the near future.

Single-Molecule, Real-Time Sequencing

Pacific Biosciences (PacBio) has developed a single-molecule sequencing method that utilizes a sequencing-by-synthesis approach combined with a zero-mode waveguide that allows for real-time sequencing of individual DNA molecules, which has been termed “single molecule, real time (SMRT)” sequencing [96]. The method uses small wells that have a diameter less than the wavelength of light chosen for interrogation (zero-mode waveguide) [97], where a DNA polymerase is bound to the bottom of each well. Nucleotides, each labeled with a different fluorophore, are added to the wells, and individual nucleotides are incorporated into a complementary DNA strand. As the fluorophore is released with each nucleotide addition, the zero-mode waveguide is used to detect the fluorophore released from the incorporated base, and thus the sequence within each waveguide is measured in real time.

SMRT sequencing does not require amplification prior to sequencing and has been reported to reduce compositional bias compared to other sequencing technologies [98], with median DNA sequence length reported in thousands of nucleotides, with substantial fractions of the DNA sequences greater than 10,000 base pairs [99]. Additionally, the lack of a DNA amplification step offers the potential for a shorter time between DNA extraction and sequence generation [100]. Such long-read

sequencing technology offers the potential of resolving pseudogenized genetic sequences [101], resolution of variant phasing, haplotype resolution, structural rearrangement determination, and indel identification [102, 103], all of which are problematic using short-read techniques.

The major drawback cited when describing SMRT sequencing is the relatively low accuracy of approximately 85% [99, 100], with indel errors predominating. Methods combining data from multiple sequencing modalities have been demonstrated to improve accuracy, albeit with increased costs and algorithm complexity [96, 104]. SMRT sequencing accuracy has also been demonstrated to improve when the same DNA molecule is sequenced multiple times within the same SMRT cell using a technique called "circular consensus sequencing" (CCS) [105, 106]. The principle of CCS is that by decreasing the length of the sequencing insert, the same DNA molecule will be sequenced multiple times, improving the overall accuracy of the consensus sequence because the sequencing errors occur randomly. Platform improvement and error-reduction optimization are ongoing.

Nanopore Technologies

Nanopore sequencing of DNA uses techniques in which DNA molecules are passed through artificial nanoscale pore composed of organic or nonorganic molecules, and the DNA sequence order is determined based on electrical or other signals that are generated during passage of the DNA molecule through the pore [107–110]. In most nanopore methods, DNA passing through the nanopore results in changes in ion current, with multiple parameters impacting the measured current, with each nucleotide impacting the ion flow in such a way that the sequence can be gleaned from the measurements [111]. Similar to SMRT technology, these techniques offer the possibility of rapid progression from sample preparation to sequence generation, long sequencing reads, low cost and high speed with compact instrumentation [112].

Nanopore sequencing is based on the concept that single stranded nucleic acid (DNA or RNA) molecules can be forced through a biological pore by electrophoresis or other mechanisms in a linear fashion, with the determination of individual base composition at a specific position detected by a change in ionic current while the molecule is moving through the pore. Pores are created so that a detection mechanism is able recognize signals generated by different bases and a unique signal is generated corresponding to the sequence. Multiple different materials and designs for the nanopore have been developed, with differences in construction of the pore influencing the speed and accuracy of the sequencing process. Currently, the two classes of pore materials are solid state and protein based, each with different strengths and weaknesses. Protein nanopores are created from membrane protein complexes composed of alpha-hemolysin or *Mycobacterium smegmatis* Porin A (MspA) [107, 108, 111], while solid-state nanopores are created using synthetic materials. One theoretical advantage of synthetic materials is the possibility of customizing pore configurations specific to given applications, allowing for potentially

more stable chemical, mechanical, and thermal properties. Nanopore technologies are currently undergoing extensive ongoing development and testing and offer different performance characteristics that are still undergoing evaluation. In addition to the important nature of the pore materials and construction, precise control of the transport of the single-stranded nucleic acid through the pore is critical for accurate sequence determination, and multiple enzymes such as polymerases and exonucleases have been evaluated for this use [108].

Read lengths for some applications have been shown to be ~10 kb [113], but error rates for the technology have been shown to range between 5% and 8% for insertions, deletions, and substitutions for some methods [114]. Some evaluations of early instruments and chemistries have indicated that only a fraction of the sequencing reads are able to be mapped to the reference sequence of a known sequencing target [115]. These factors limit the current standalone, direct utility of the technologies in the diagnostic setting, but can be used as a method to generate a scaffold sequence that can be combined with other sequencing methods. Some analyses have suggested that accuracy can be improved through increasing the read coverage of individual sequenced molecules [116]. The methods may also require bioinformatics tools specific to nanopore sequencing [85].

Summary

Although SMRT and nanopore sequencing technology have been available in select laboratories for several years, there is limited data on the use of such techniques in the setting of hematological malignancies and particularly in the clinical laboratory [117]. The currently described technologies capable of sequencing single-molecule nucleic acids with long-read techniques offer attractive features that would be useful in both research and diagnostic applications in myeloid neoplasia. Contiguous sequencing reads capable of routinely resolving tens of kilobases of sequence would be useful in the setting of indels and structural rearrangements, which are encountered in myeloid neoplasia. Additionally, the ability to resolve the phase of mutations, haplotypes, and discriminate between real and pseudogenes could potentially impact our understanding of myeloid neoplasia and diagnosis. Although there are methods that promise improved accuracy, the current techniques are hampered by high error rates. The possibility of using combinations of sequencing technologies to use strengths of one technique to offset the weaknesses of another is appealing but will require new software solutions to merge data from different platforms and produce a unified output. Currently, the main limitation of these techniques that limits their widespread use is the high error rates in sequenced DNA. Technological improvements may yield improvements in accuracy and strategies have been developed to increase the sequencing accuracy of individual molecules but such accuracy often comes at the expense of read length. Developments in these technologies and their successors warrant special attention and consideration for the potential scientific and diagnostic uses for these techniques in the setting of myeloid neoplasia should be considered in the future.

Other Technologies Present in the Literature

Although there are several varieties of next-generation sequencing technologies that are in common use, multiple other technologies were commercialized and extensively used in the literature. While some technologies have faded from use or even been discontinued by the manufacturer, large amounts of data were generated using these techniques and studies using them may be encountered not only in older work [118, 119] but in recently published studies as well [120].

Pyrosequencing Sequencing by Synthesis (454/Roche)

Key aspects of the 454 sequencing method include emulsion PCR and pyrosequencing. As part of the sequencing library preparation, target DNA is ligated to primers where one primer is linked to a biotin molecule. During library preparation, individual molecules of DNA are captured on streptavidin beads within picoliter emulsions of oil and reagents for PCR (emulsion PCR) [121]. After clonal amplification of the target DNA bound within a given droplet, the resulting clonal products are deposited on custom microtiter plate ("picotiterplate") with wells ~29 μm in diameter, with millions of clonal sequences deposited across a single plate. Sequence determination is done through the use of pyrosequencing technology, where individual nucleotides are sequentially added, and if a given nucleotide is incorporated into the nascent DNA strand, pyrophosphate (PPi) is released into the milieu of the microwell. Also present within the reaction mixture is the enzyme luciferase, which uses the PPi as a substrate to generate light for detection allowing inference of the incorporated nucleotide at a given position [122]. Similar to the semiconductor NGS methods used on the Ion Torrent platform, an increased signal is generated when homopolymer tracts are encountered because a chain-termination strategy is not employed. As such, uncertainty in homopolymer length is a common occurrence in this method, and computational methods have been developed to address this issue [80, 83, 84]. The pyrosequencing method is capable of producing large amounts of long, high-quality sequencing reads, and this technology has been described to produce read lengths that are longer than many other NGS technologies but less than the maximum sequence lengths generated using traditional Sanger techniques [61]. Further, this method had lower throughput than other systems [80, 123] and a higher cost-per-base [123]. The platform was purchased by Roche Diagnostics in 2007 and manufacturer support was scheduled to be discontinued in 2016.

Sequencing by Oligonucleotide Ligation and Detection (SOLiD)

Similar to other NGS library preparation methods, SOLiD relies on single nucleic acid targets isolated for subsequent clonal enrichment. In SOLiD, magnetic beads are used to isolate a single target sequence per bead, and emulsion PCR is used to

amplify DNA to increase the number of copies of the unique target DNA bound to each bead as well as incorporate adapter sequences into the amplified DNA. Postamplification, PCR products are covalently bound to a glass slide for subsequent DNA sequencing. Using primers specific to the adapter sequences incorporated in PCR, the captured DNA is sequenced by using a set of four fluorescently labeled, two-base probes, which compete to be ligated to the sequencing primer. The fluorescence is measured to determine the incorporated probe, and the fluorescent molecule is released and the process is repeated. Following a series of ligation cycles, the extension product is removed and the template is reset with a primer complementary to the $n-1$ position for a second round of ligation cycles. This process is repeated multiple times for each adapter, allowing for each base to be interrogated in multiple independent ligation reactions by different primers [124]. This approach obtains specificity by interrogating every first and second base in each ligation reaction [125] and redundancy through interrogation by multiple ligation probes. The SOLiD method has been reported to have difficulty with palindromic sequences [126] and has been claimed to have decreased susceptibility to homopolymer tracts. This technology has been described to have intermediate throughput and costs compared to other NGS methods [123].

Scale of Sequencing

Each of these sequencing-based methods is capable of being used on multiple scales. Briefly, the scales of next-generation sequencing are whole genome, exome, and targeted panels (Fig. 19.7). Applications used extensively by research groups include sequencing an entire genome for identification of alterations with potential clinical importance [127]. Whole-genome sequencing (WGS) has the advantage of relatively unbiased sequencing, the capability of detecting CNVs across the genome, and the ability to identify genetic changes that would be missed by more targeted methods. However, WGS generally has lower overall coverage of individual bases, requires additional instrumentation and bioinformatics resources compared to other

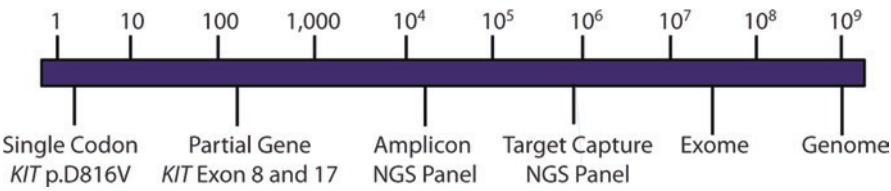

Fig. 19.7 Linear scale approximating the number of base pairs analyzed in different scales of molecular diagnostic testing using single-site assays, limited exon targeting assays, amplicon-based NGS panels, comprehensive NGS panels using hybrid capture designs, whole-exome and whole-genome approaches. Note that typical target-capture NGS designs capture several orders of magnitude more base pairs than amplicon panels

methods, and thus, is generally more expensive to produce data. Multiple groups have used WGS to understand the evolution of individual neoplasms over time [128, 129]. While such studies may be useful to help identify individual mutations associated treatment response or failure, the breadth of data generated from such studies is unlikely to prove cost effective in routine clinical use at the present time.

Exome-level sequencing utilizes multiple techniques to isolate the coding regions of genes, most frequently utilizing hybrid-capture techniques where antisense oligonucleotides are used to enrich for the sequences of interest [130, 131]. Whole-exome sequencing (WES) offers the potential advantage of overall high-sequencing depth compared to WGS because the targeted DNA represents a much smaller fraction of the total genomic DNA, with decreased reagent and bioinformatics costs; however, the cost of the reagents needed to isolate the exonic DNA can be expensive and adds complexity to the sample library preparation. As a result of the relative decreased cost of sequencing and increased depth, many studies have been undertaken to evaluate exomes of hematopoietic neoplasms. Studies using whole-exome sequencing, like studies employing WGS, often identify known driver mutations, show evidence of clonal evolution through therapy, and identify new mutations that arise over time [132, 133]. Although the cost of exome sequencing and the ability to utilize the necessary bioinformatics pipelines are within the reach of many groups, routine exome sequencing may be in limited clinical use for the evaluation of neoplastic disease due to the more manageable data that are generated using targeted gene panels.

Utilizing techniques similar to WES, targeted gene sequencing panels can utilize hybrid capture techniques [65, 134] or a more limited coverage can be obtained using ligation-probe or PCR-based amplification techniques [135–138]. Such targeted techniques allow for greater sequencing depth in specific genes of interest, potentially allowing for greater sensitivity in detecting minor clonal components. Total assay costs may be less than WES and WGS because of the limited coverage allowing for the use of smaller scale sequencers, multiplexing of multiple samples, and less resource-intensive bioinformatics pipelines than those used in WES and WGS. One potential advantage of using a limited set of genes is the understanding that the significance of individual mutations may be more achievable than variants detected in more expansive platforms. When comparing hybrid capture methods against amplicon-based enrichment techniques, the ability to detect CNVs, large indels, and translocations may be greater in target-capture techniques due to the less-specific genomic enrichment using such techniques and the availability of numerous software packages that can be incorporated into the data analysis pipeline after sequence acquisition [87, 89, 139–141]. Numerous examples of custom-developed methods targeting specific genes related to myeloid neoplasia are described in the literature. Most methods describe the ability to detect mutations that are commonly encountered in myeloid neoplasia [142, 143], with hybridization-capture enrichment techniques often noting the ability to detect large structural rearrangements and translocations [144]. Although amplicon-based NGS panels typically are not designed to detect larger indels or CNVs, several bioinformatics

approaches have been developed that appear to glean such information from such assay designs [145, 146].

Assessment of the mutational spectrum of myeloid neoplasia has become a common tool in clinical care of patients and in the research setting. While techniques such as WGS and WES are of interest to researchers, their use in routine clinical care is not currently clear and limited. In contrast, the use of targeted gene panels offers the benefits of sequencing a more focused set of genomic targets, resulting in greater sequencing depth, lower cost, and potentially easier path to demonstrate clinical utility.

Emerging Techniques and Applications

The previously described technologies and techniques have been applied to numerous scientific and clinical problems and have found acceptance in the diagnostic workup, prognostication, and therapy prediction of myeloid malignancies. Additional technologies and applications have been described and are currently undergoing evaluation for their use in clinical applications. It is unknown if any of the techniques or applications described will ultimately be found in routine use in the evaluation of myeloid neoplasia, but general understanding of the conceptual basis and early scientific data may be useful to practitioners as they encounter such techniques in the literature.

Minimal Residual Disease Detection

It is increasingly being appreciated that minimal residual disease (MRD) detection by conventional means, such as flow cytometry, is useful for following the response to treatment and for identifying high-risk patients who are beginning to relapse. In some situations, treatment decisions that impact patient outcome may be made based on the MRD test results [147], making accurate and timely results very important in clinical care. Current clinical approaches for detecting MRD and supplementing morphologic examination and cytogenetic studies in myeloid neoplasms include flow cytometry [148], RT-PCR [48–51], and PCR assays [149]. While understanding of flow cytometry and PCR are common in many clinical laboratories, application of these methods to MRD detection requires some expertise to properly design these assays and interpret the results in order to ensure standardization for comparison of results. As such, such assays are often limited to larger commercial reference laboratory and academic institutions. The development of new laboratory approaches for minimal residual disease detection may result in improved techniques that may be less reliant on pathologist expertise to determine the MRD status.

Several promising approaches exist that may prove useful in adapting new techniques to MRD analysis, including deep mutation scanning of many different possible genetic mutations [150, 151] and deep sequencing of specific recurrent mutations that are observed in a subset of patients [152–154].

Many myeloid neoplasms have recurrent abnormalities and mutational profiles that show remarkable consistency between individuals [155]. Some studies have evaluated diagnostic methods that are specific to these unique genomic signatures of specific neoplasms, potentially allowing specific laboratory methods to be used to assess MRD in patients with neoplasms harboring these genetic lesions [155]. An example of using a recurrent mutation to monitor MRD with next-generation sequencing has been described in the setting of *NPM1*-mutated AML [152], which represents ~60% of AML cases with a normal karyotype. In this example, amplification and sequencing of exon 12 of *NPM1* allowed detection of low-burden AML disease that was shown to be capable of detecting mutant cells down to approximately 0.001% of a mixture, using next-generation sequencing. Compared to allele-specific PCR approaches, a priori knowledge of the patient's *NPM1* mutation allele is not needed. Next-generation sequencing of *NPM1* mutations also showed the potential of identifying subclonal heterogeneity within a given sample, offering the possibility of tracking clonal evolution of the neoplastic populations over time. Such an approach to MRD detection offers the potential for a more simplified approach to MRD monitoring that does not require the extensive expertise needed for flow cytometry assessment of MRD or the complexity of performing multiple allele-specific quantitative PCR approaches. It is, however, critical to recognize that with highly sensitive NGS methods, there is a possibility of sample-to-sample or amplicon contamination that must be addressed, and careful precautions, such as those used in viral molecular diagnostics, must be taken to limit the possibility of such events [156]. Clinical laboratories will also have to determine if maintaining and performing complex assays that are specific to only a subset of their patient population is a feasible operation from the standpoint of cost and turnaround time. Finally, the clinical significance of such low-level persistent or recurrent disease will need to be studied to determine if such methods are appropriate.

An alternative approach to neoplasm subtype specific assays is to identify the most common genes mutated across many myeloid malignancies and then design a method that is capable of identifying the majority of mutations in those genes. By expanding the pool of possible genetic targets, there is the possibility that the presence of multiple mutations will increase the specificity of the analysis, as well as potentially increase the sensitivity through the elimination of false-negative results. Such methods propose a more flexible approach, which could allow the MRD analysis to include novel mutations or to identify populations evolving from the original leukemic clone. Approaches have been described that are approximately tenfold more sensitive than Sanger-sequencing-based methods for detection of genetic variants, and was highly concordant multiple laboratories [150, 151]. While these results indicate that sensitive assays can be implemented across multiple laboratories and achieve comparable results, such methods are currently less sensitive than typical state-of-the-art multidimensional flow cytometric techniques used to mea-

sure MRD. Identifying patient-specific mutations early in disease and then using deep-sequencing of the mutated regions has been suggested to approach the sensitivity of flow cytometric methods [154], but such a strategy requires multistep, patient-specific evaluation methods and may lack the flexibility of the more generalized methods or may be relatively costly with limited sensitivity. Additionally, deep-sequencing methods using amplicon approaches run the risk of introducing PCR artifacts in early amplification cycles that could mimic, by chance, previously identified mutations, resulting in false positive results.

With all proposed methods for detecting MRD, the hope is that such approaches could shorten the window period between the recurrence of a neoplasm and the ability of laboratory techniques to detect disease relapse. Additional studies need to evaluate not only the analytical sensitivity and specificity of these methods, but whether clinical benefit is enhanced. While initial studies evaluating MRD in multiple laboratories have shown good concordance of results [150], additional work will need to demonstrate that such concordance continues, if such methods are brought into common practice, especially if other methods of MRD analysis are not available for confirmatory testing. With all approaches, it must be remembered that persistent, nonleukemic hematopoietic clones may present in any given patient and that such clones will share many mutations common with truly neoplastic leukemic populations, and that such populations may persist over time [157, 158], and even after initial chemotherapy treatment [159].

Methods to Increase Accuracy of Deep Sequencing

Application of conventional NGS techniques to the ultrasensitive deep-sequencing techniques, such as identification of MRD, is hampered by the inability to resolve very small populations of mutant genetic changes from errors introduced by amplification or sequencing. One approach that has been suggested to remedy this problem is duplex sequencing, which proposes a method to identify the source of each individual amplicon in a sequencing reaction, allowing for the identification of artifacts that are present in only one strand of an amplification reaction [160]. A second comparable approach uses single molecular tagging to permit detection of rare variants beyond the traditional limits due to error-correction [160–162]. Through incorporation of additional random sequence "barcodes" to the primers for specific sequence targets, the providence of individual amplicons can be ascertained, such that artifacts can be excluded through the comparison of specific sequencing products. This method works by grouping individual sequences into sequence families that have the same barcodes identifier in common, which are then compared against other reads within the family to remove individual errors through creation of an error-corrected consensus sequence. Descriptions of this technique claim to allow a single mutation to be identified among 1×10^7 wild-type sequences, and a study using a similar approach in the setting of preleukemic clonal hematopoiesis in patients who later developed treatment-related myeloid neoplasms demonstrated the ability to identify 1 abnormal sequencing read in 10,000 [161]. These methods

offer the potential to detect mutations related to residual or incipient diseases with much greater sensitivity than currently available flow cytometric or nonerror-corrected NGS MRD methods.

Error-correction NGS methods will require adjustments in bioinformatics pipelines between the sequence demultiplexing and alignment steps, in addition to synthesis of primer sequences incorporating the random sequence barcodes. The depth of sequence at a given position will need to be high, so that there is adequate sampling of the available primers, which will require either more sequencing throughput or fewer targets so that sufficient sequencing depth can be achieved. Ultimately, the necessity for such error-correction methods in NGS MRD applications will be highly dependent on the clinical need and utility for such ultrasensitive monitoring of mutant populations and the lack of alternative methods to increase the fidelity of the sequence calls. Although there are ample data supporting the utility of clinical MRD assessment in the setting of myeloid neoplasia, there are many factors that may limit wider clinical adoption of current methods [163, 164]. The necessity for such error-correction methods in NGS MRD applications will be dependent on the clinical context and will be incumbent on the laboratory to demonstrate the enhanced analytic sensitivity will translate to enhanced clinical care. Nevertheless, the ability to probe and evaluate for ultrarare mutations beyond standard limits of detection will likely be important in the future [161, 165].

Single Cell Sequencing

As NGS has become a common research tool, many groups have applied this toward the analysis of the clonal evolution of hematopoietic neoplasms. Some studies have demonstrated that the acquisition of mutations is likely to occur over the evolution of the neoplasm [166] and in response to treatment [159]. In many of these studies, the presence of unique or evolving mutations acquired by neoplastic clones within the larger leukemic population is inferred from the variant allele fraction (VAF), the measure of the proportion of sequencing reads with a mutation versus the total number of reads (mutation and wild-type) at a particular loci as determined by bulk sequencing of diseased tissue. However, this assumption may not be correct. Development of new approaches, including analysis of flow-cytometric cell sorted cells, has allowed a refinement of this understanding through the ability to sequence individual neoplastic cells in order to compare the genetic aberrations present in those cells compared to the bulk neoplasm.

Current single-cell sequencing techniques potentially require special preservation and handling of the sample material and currently require whole-genome amplification of the individual genomes [167]. Such an approach offers an unprecedented, granular assessment of the mutations within individual leukemic cells, but requires specialized expertise and is susceptible to incorporation of artifacts from whole-genome amplification [168, 169].

The evaluation of individual neoplastic cells has revealed that the mutation spectrum within leukemic subpopulations is more nuanced than previously appreciated [170]. In one study, individual AML cells were shown to possess *FLT3* and *NPM1* mutations in both heterozygous and homozygous states and were distributed across multiple, distinct clonal populations [170]. Although these observations give insight into the disease process, it is currently unknown if there is a clinical benefit to monitoring of this mutational heterogeneity in clinical samples, although it is readily conceivable that such evaluation at the single-cell level could identify clones that may be resistant to therapy and thereby could provide information to guide treatment decisions. This emerging technique, however, will require additional study and evaluation of clinical usefulness in order to determine the potential clinical benefits and establish feasibility of implementation in the clinical environment.

Expression Profiling

Expression profiling of AML has provided scientific insight into the implications of the altered genetic makeup of leukemic cells [171]. Such profiling has revealed that some neoplasms may have different subtypes that are distinguished by expression profiling [172]. Transcriptional profiling has even been done at the single-cell level for myeloid cells, demonstrating that individual cells have expression profiles indicating a distinct lineage without overlap [173]. Some recent studies have suggested that alterations in expression may have prognostic significance [174–177] but these analyses have not yet been widely translated into routine clinical practice.

RNA-SEQ

RNA-seq is a modification of NGS in which the starting template is RNA from the patient sample and not DNA. RNA-seq has been proposed as a useful complementary technology to conventional diagnostic techniques because of the ability to detect complex genomic events and splice-site alterations that may not be as readily identifiable by other techniques [178]. RNA-Seq may be useful at identifying novel or unexpected translocations that may be clinically or diagnostically relevant but not readily identifiable by standard techniques [179, 180]. RNA-seq has demonstrated that differential expression of genes can be detected in different cell types related to disease entities [181] or discerning between closely related clones that have different characteristics [182]. RNA-seq may be useful in identifying splice-site changes or variants that may not be detectable by other means [183–185]. Research uses for RNA-seq include identification of mechanisms of drug resistance in the setting of myeloid neoplasia [186], but its utility in a routine clinical environment is yet unproven. While there are potential advantages of using RNA-seq, such as more efficient detection of the impact of epigenetic changes, splice-site changes

and structural rearrangements, the ability to analyze these parameters requires specialized informatics tools [187, 188]. Some limitations of RNA-seq include the starting substrate is more labile than DNA and thus is technically more difficult with which to work, and, at the present time, offers few advantages over complimentary methods. As such, it currently is uncertain if RNA-seq will prove a useful additional to the clinical armamentarium but will most certainly be useful in research. New approaches for RNA-seq include the ability to profile transcriptomes of tens of thousands of single cells [189, 190]. This technology provides unprecedented granularity and insight into the different cell populations that may be present in normal and disease [173, 191] and is likely to provide insight into the complexity of clonal competition and evolution in hematologic neoplasms.

Epigenetics

Epigenetic modifications are reversible alterations to DNA or histones of a cell that impact gene expression without altering the underlying genetic sequence of the cell. Many genes involved in the epigenetic modification of the genome have been identified to be mutated in myeloid disorders and have been the subject of intense study [192–195]. Further work has suggested that genes involved in epigenetic modification may be useful as prognostic markers in the setting of myeloid neoplasia [196, 197] and studies investigating the use of therapies targeting aberrant epigenetic modification have been published [198, 199]. The relationship between individual disease entities, genomic mutations in genes related to epigenetics, and the consequences of the mutations are under active investigation [200]. Tools for the evaluation of epigenetic alterations in myeloid neoplasia in clinical settings are available [201] but it is unknown if the information provided by such methods will add information to what is already captured through the extensive analysis that is currently applied to many individuals with myeloid malignancies [202]. As with many other laboratory techniques, the clinical utility of the approach will need to be evaluated to determine if there is meaningful added clinical benefit.

Confounding

As next-generation sequencing was applied in different clinical arenas, it became unexpectedly apparent that many of the mutations commonly associated with myeloid neoplasia can be detected in individuals without evidence or involvement of a myeloid malignancy. The detection of these mutations confounds routine clinical laboratory testing as the presence of such mutations does not necessarily indicate the presence of disease as currently defined or understood. In such situations, additional clinical testing could be unnecessary and may result in misdiagnosis. Studies examining the prevalence of hematopoietic mutations in otherwise healthy

populations have shown that the incidence of mutations in genes commonly associated with myeloid neoplasms appears to increase with age [157, 158, 203]. In these individuals, a small percentage of the individuals with age-related hematopoietic clones appear to develop hematological neoplasia, with individuals with a larger proportion of the cells in the peripheral circulation harboring such mutations having an increased risk. Other studies have demonstrated that some mutations in hematopoietic cells arise in patients with other types of cancer, often after the selective pressure of chemotherapy [204]. While the emergence of hematopoietic clones is not uncommon, the incidence of hematopoietic neoplasia in such settings appears to be uncommon based on retrospective studies [205, 206]. Some authors have termed this phenomena, "clonal hematopoiesis of indeterminate potential," or as abbreviated, CHIP [206]. Additional studies evaluating the future risk for neoplastic disease and the recommendations for reporting, screening, and monitoring of such clonal populations will need to be developed.

As another example, new approaches to prenatal diagnosis evaluating genomic CNVs in cell-free DNA (cfDNA) have also reported several instances of inadvertent detection of maternal neoplasia in pregnant women undergoing routine screening [207]. It is conceivable that assays evaluating either fetal or neoplastic cfDNA may be capable of detecting CNVs and mutations that are originating in hematopoietic cells. Awareness of the possibility may be sufficient to identify ancillary methods that may be useful in resolving the observed data, but additional tools may be useful in the future to help resolve these situations. For example, recent work has demonstrated that the cell-of-origin of cfDNA fragments can be ascertained based upon nucleosome signatures of the detected DNA molecules [208], potentially allowing for the discernment not only of the variants associated with neoplasia, but also the cell of origin (i.e., hematopoietic or nonhematopoietic in nature). Incorporation of such methods into cfDNA screening and monitoring methods will help to ensure that detected mutations are appropriately classified and understood.

These examples together highlight the importance of open communication between clinicians and laboratory pathologists. Indeed, awareness of these potential confounders of sequencing data obtained from advanced diagnostic tests will help physicians and scientists appropriate counsel providers and patients.

Conclusions

The determination of the diagnosis and prognosis of myeloid neoplasia is greatly facilitated by advanced laboratory techniques that have been employed to dissect the underlying molecular lesions related to the malignancy. The classification of myeloid neoplasia has become more granular, specific, and detailed as the available laboratory techniques and biological understanding of the diseases have advanced. Laboratory techniques such as karyotype, FISH, PCR, Sanger sequencing, and microarray remain the mainstay of current clinical diagnostic laboratories, but are now often complemented by massively parallel sequencing. Due to the

comprehensive nature of NGS, it is conceivable that this technology may replace some of the more traditional methods for molecular cytogenetic analysis. As NGS has proliferated in scientific and clinical applications, new permutations have been proposed that may aid in improvements in clinical care, if analyses of clinical utility bear out. As these new technologies and tools continue to mature, new opportunities will arise to allow greater scientific understanding and potential for improved clinical care. Through knowledge of the laboratory methods that are currently in clinical use, the strengths and limitations of the different approaches will be important to consider.

References

1. Smith HO, Wilcox KW. A restriction enzyme from Hemophilus influenzae. I Purification and general properties. J Mol Biol. 1970;51(2):379–91.
2. Price PM, Hirschhorn K. In situ hybridization of chromosome loci. Fed Proc. 1975;34(13):2227–32.
3. Pinkel D, Straume T, Gray JW. Cytogenetic analysis using quantitative, high-sensitivity, fluorescence hybridization. Proc Natl Acad Sci U S A. 1986;83(9):2934–8.
4. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94(3):441–8.
5. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.
6. Maxam AM, Gilbert W. A new method for sequencing DNA. Proc Natl Acad Sci U S A. 1977;74(2):560–4.
7. Bartlett JM, Stirling D. A short history of the polymerase chain reaction. Methods Mol Biol. 2003;226:3–6. doi:10.1385/1-59259-384-4:3.
8. Saiki RK, Bugawan TL, Horn GT, Mullis KB, Erlich HA. Analysis of enzymatically amplified beta-globin and HLA-DQ alpha DNA with allele-specific oligonucleotide probes. Nature. 1986;324(6093):163–6. doi:10.1038/324163a0.
9. Lawyer FC, Stoffel S, Saiki RK, Myambo K, Drummond R, Gelfand DH. Isolation, characterization, and expression in Escherichia coli of the DNA polymerase gene from Thermus aquaticus. J Biol Chem. 1989;264(11):6427–37.
10. Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, Erlich HA. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science. 1988;239(4839):487–91.
11. Dryja TP, Rapaport JM, Joyce JM, Petersen RA. Molecular detection of deletions involving band q14 of chromosome 13 in retinoblastomas. Proc Natl Acad Sci U S A. 1986;83(19):7391–4.
12. Lübbert M, Mirro J, Miller CW, Kahan J, Isaac G, Kitchingman G, Mertelsmann R, Herrmann F, McCormick F, Koeffler HP. N-ras gene point mutations in childhood acute lymphocytic leukemia correlate with a poor prognosis. Blood. 1990;75(5):1163–9.
13. Iggo R, Gatter K, Bartek J, Lane D, Harris AL. Increased expression of mutant forms of p53 oncogene in primary lung cancer. Lancet. 1990;335(8691):675–9.
14. Komminoth P, Kunz EK, Matias-Guiu X, Hiort O, Christiansen G, Colomer A, Roth J, Heitz PU. Analysis of RET protooncogene point mutations distinguishes heritable from nonheritable medullary thyroid carcinomas. Cancer. 1995;76(3):479–89.
15. Claas EC, Melchers WJ, van der Linden HC, Lindeman J, Quint WG. Human papillomavirus detection in paraffin-embedded cervical carcinomas and metastases of the carcinomas by the polymerase chain reaction. Am J Pathol. 1989;135(4):703–9.

16. Feinmesser R, Miyazaki I, Cheung R, Freeman JL, Noyek AM, Dosch HM. Diagnosis of nasopharyngeal carcinoma by DNA amplification of tissue obtained by fine-needle aspiration. N Engl J Med. 1992;326(1):17–21.
17. Melo JV, Gordon DE, Cross NC, Goldman JM. The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. Blood. 1993;81(1):158–65.
18. Dobrovic A, Trainor KJ, Morley AA. Detection of the molecular abnormality in chronic myeloid leukemia by use of the polymerase chain reaction. Blood. 1988;72(6):2063–5.
19. Thompson JD, Brodsky I, Yunis JJ. Molecular quantification of residual disease in chronic myelogenous leukemia after bone marrow transplantation. Blood. 1992;79(6):1629–35.
20. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. Blood. 2002;100(7):2292–302.
21. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. Blood. 2009;114(5):937–51.
22. Swerdlow SH, Campo E, Pileri SA, Harris NL, Stein H, Siebert R, et al. The 2016 revision of the World Health Organization classification of lymphoid neoplasms. Blood. 2016;127(20): 2375–90.
23. Creighton HB, McClintock B. A correlation of cytological and Genetical crossing-over in Zea Mays. Proc Natl Acad Sci U S A. 1931;17(8):492–7.
24. Coe E, Kass LB. Proof of physical exchange of genes on the chromosomes. Proc Natl Acad Sci U S A. 2005;102(19):6641–6.
25. Martin CL, Warburton D. Detection of chromosomal aberrations in clinical practice: from Karyotype to genome sequence. Annu Rev Genomics Hum Genet. 2015;16:309–26.
26. Marcucci G, Mrozek K, Bloomfield CD. Molecular heterogeneity and prognostic biomarkers in adults with acute myeloid leukemia and normal cytogenetics. Curr Opin Hematol. 2005;12(1):68–75.
27. Van Den Berghe H, Louwagie A, Broeckaert-Van Orshoven A, David G, Verwilghen R, Michaux JL, Sokal G. Chromosome abnormalities in acute promyelocytic leukemia (APL). Cancer. 1979;43(2):558–62.
28. Berger R, Bernheim A, Daniel MT, Valensi F, Flandrin G. Karyotype and cell phenotypes in primary acute leukemias. Blood Cells. 1981;7(2):287–92.
29. Sakurai M, Hayata I, Sandberg AA. Prognostic value of chromosomal findings in Ph1-positive chronic myelocytic leukemia. Cancer Res. 1976;36(2 Pt 1):313–8.
30. Meisner L, Inhorn SL, Nielsen P. Karyotype evolution of cells with the Philadelphia chromosome. Acta Cytol. 1970;14(4):192–9.
31. Grimwade D, Walker H, Oliver F, Wheatley K, Clack R, Burnett A, Goldstone A. What happens subsequently in AML when cytogenetic abnormalities persist at bone marrow harvest? Results of the 10th UK MRC AML trial. Medical Research Council Leukaemia Working Parties. Bone Marrow Transpl. 1997;19(11):1117–23.
32. Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on drosophila polytene chromosomes. Proc Natl Acad Sci U S A. 1982;79(14):4381–5.
33. Landstrom AP, Tefferi A. Fluorescent in situ hybridization in the diagnosis, prognosis, and treatment monitoring of chronic myeloid leukemia. Leuk Lymphoma. 2006;47(3):397–402.
34. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, et al. Use of whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA. 2011;305(15):1577–84.
35. Huang XC, Quesada MA, Mathies RA. DNA sequencing using capillary array electrophoresis. Anal Chem. 1992;64(18):2149–54.
36. Heather JM, Chain B. The sequence of sequencers: the history of sequencing DNA. Genomics. 2016;107(1):1–8.
37. Hood LE, Hunkapiller MW, Smith LM. Automated DNA sequencing and analysis of the human genome. Genomics. 1987;1(3):201–12.
38. Zimmermann J, Voss H, Schwager C, Stegemann J, Ansorge W. Automated Sanger dideoxy sequencing reaction protocol. FEBS Lett. 1988;233(2):432–6.

39. Hood L, Rowen L. The human genome project: big science transforms biology and medicine. Genome Med. 2013;5(9):79.
40. Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. Whole-genome random sequencing and assembly of Haemophilus influenzae rd. Science. 1995;269(5223):496–512.
41. Venter JC, Adams MD, Sutton GG, Kerlavage AR, Smith HO, Hunkapiller M. Shotgun sequencing of the human genome. Science. 1998;280(5369):1540–2.
42. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. Science. 2001;291(5507):1304–51.
43. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, et al.; Consortium IHGS (2001) Initial sequencing and analysis of the human genome. Nature 409(6822):860–921.
44. Schnittger S, Bacher U, Kern W, Alpermann T, Haferlach C, Haferlach T. Prognostic impact of FLT3-ITD load in NPM1 mutated acute myeloid leukemia. Leukemia. 2011;25(8):1297–304. doi:10.1038/leu.2011.97.
45. Kern W, Haferlach C, Haferlach T, Schnittger S. Monitoring of minimal residual disease in acute myeloid leukemia. Cancer. 2008;112(1):4–16.
46. Jobbagy Z, van Atta R, Murphy KM, Eshleman JR, Gocke CD. Evaluation of the Cepheid GeneXpert BCR-ABL assay. J Mol Diagn. 2007;9(2):220–7.
47. Yin JA, O'Brien MA, Hills RK, Daly SB, Wheatley K, Burnett AK. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. Blood. 2012;120(14):2826–35.
48. van Dongen JJ, Macintyre EA, Gabert JA, Delabesse E, Rossi V, Saglio G, et al. Standardized RT-PCR analysis of fusion gene transcripts from chromosome aberrations in acute leukemia for detection of minimal residual disease. Report of the BIOMED-1 concerted action: investigation of minimal residual disease in acute leukemia. Leukemia. 1999;13(12):1901–28.
49. van der Velden VH, Hochhaus A, Cazzaniga G, Szczepanski T, Gabert J, van Dongen JJ. Detection of minimal residual disease in hematologic malignancies by real-time quantitative PCR: principles, approaches, and laboratory aspects. Leukemia. 2003;17(6):1013–34.
50. Santamaria C, Chillon MC, Fernandez C, Martin-Jimenez P, Balanzategui A, Garcia Sanz R, San Miguel JF, Gonzalez MG. Using quantification of the PML-RARalpha transcript to stratify the risk of relapse in patients with acute promyelocytic leukemia. Haematologica. 2007;92(3):315–22.
51. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe against cancer program. Leukemia. 2003;17(12):2318–57.
52. Kafatos FC, Jones CW, Efstratiadis A. Determination of nucleic acid sequence homologies and relative concentrations by a dot hybridization procedure. Nucleic Acids Res. 1979;7(6):1541–52.
53. Southern EM. DNA microarrays. History and overview. Methods Mol Biol. 2001;170:1–15.
54. Hoheisel JD. (1994) application of hybridization techniques to genome mapping and sequencing. Trends Genet. 1994 Mar;10(3):79–83.
55. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949–51.
56. Hemmat M, Chen W, Anguiano A, Naggar ME, Racke FK, Jones D, et al. Submicroscopic deletion of 5q involving tumor suppressor genes (CTNNA1, HSPA9) and copy neutral loss of heterozygosity associated with TET2 and EZH2 mutations in a case of MDS with normal chromosome and FISH results. Mol Cytogenet. 2014;7:35.
57. Cheung KJ, Rogic S, Ben-Neriah S, Boyle M, Connors JM, Gascoyne RD, Horsman DE. SNP analysis of minimally evolved t(14;18)(q32;q21)-positive follicular lymphomas reveals a common copy-neutral loss of heterozygosity pattern. Cytogenet Genome Res. 2012;136(1):38–43.

58. Greisman HA, Hoffman NG, Yi HS. Rapid high-resolution mapping of balanced chromosomal rearrangements on tiling CGH arrays. J Mol Diagn. 2011;13(6):621–33.
59. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, Wang MD, Zhang K, Mitra RD, Church GM. Accurate multiplex polony sequencing of an evolved bacterial genome. Science. 2005;309(5741):1728–32.
60. Brenner S, Johnson M, Bridgham J, Golda G, Lloyd DH, Johnson D, et al. Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. Nat Biotechnol. 2000;18(6):630–4.
61. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387–402.
62. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. J Appl Genet. 2011;52(4):413–35.
63. Gullapalli RR, Lyons-Weiler M, Petrosko P, Dhir R, Becich MJ, LaFramboise WA. Clinical integration of next-generation sequencing technology. Clin Lab Med. 2012;32(4):585–99.
64. Vosberg S, Herold T, Hartmann L, Neumann M, Opatz S, Metzeler KH, et al. Close correlation of copy number aberrations detected by next-generation sequencing with results from routine cytogenetics in acute myeloid leukemia. Genes Chromosomes Cancer. 2016;55(7):553–67.
65. Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, et al. Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and Gene rearrangement in clinical cancer specimens. J Mol Diagn 2014 Jan. 2013;16(1):56–67.
66. Richter BG, Sexton DP. Managing and analyzing next-generation sequence data. PLoS Comput Biol. 2009;5(6):e1000369.
67. Batley J, Edwards D. Genome sequence data: management, storage, and visualization. BioTechniques. 2009;46(5):333–334, 336.
68. Sinard JH, Powell SZ, Karcher DS. Pathology training in informatics: evolving to meet a growing need. Arch Pathol Lab Med. 2014;138(4):505–11.
69. Haspel RL, Olsen RJ, Berry A, Hill CE, Pfeifer JD, Schrijver I, Kaul KL. Progress and potential: training in genomic pathology. Arch Pathol Lab Med. 2014;138(4):498–504.
70. Overby CL, Kohane I, Kannry JL, Williams MS, Starren J, Bottinger E, Gottesman O, Denny JC, Weng C, Tarczy-Hornoch P, Hripcsak G. Opportunities for genomic clinical decision support interventions. Genet Med. 2013;15(10):817–23.
71. Gray SW, Hicks-Courant K, Cronin A, Rollins BJ. Weeks JC (2014) Physicians' attitudes about multiplex tumor genomic testing. J Clin Oncol. 2014 May 1;32(13):1317–23.
72. Starren J, Williams MS, Bottinger EP. Crossing the omic chasm: a time for omic ancillary systems. JAMA. 2013;309(12):1237–8.
73. Guo J, Xu N, Li Z, Zhang S, Wu J, Kim DH, et al. Four-color DNA sequencing with 3′-O-modified nucleotide reversible terminators and chemically cleavable fluorescent dideoxynucleotides. Proc Natl Acad Sci U S A. 2008;105(27):9145–50.
74. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. Proc Natl Acad Sci U S A. 2006;103(52):19635–40.
75. Bentley DR, Balasubramanian S, Swerdlow HP, Smith GP, Milton J, Brown CG, et al. Accurate whole human genome sequencing using reversible terminator chemistry. Nature. 2008;456(7218):53–9.
76. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res. 2008;36(16):e105.
77. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 2012;40(10):e72.
78. Ross MG, Russ C, Costello M, Hollinger A, Lennon NJ, Hegarty R, Nusbaum C, Jaffe DB. Characterizing and measuring bias in sequence data. Genome Biol. 2013;14(5):R51.
79. Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke A. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries. Genome Biol. 2011;12(2):R18.

80. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, Pallen MJ. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434–9.
81. Kircher M, Heyn P, Kelso J. Addressing challenges in the production and analysis of illumina sequencing data. BMC Genomics. 2011;12:382.
82. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, et al. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475(7356):348–52.
83. Yeo ZX, Chan M, Yap YS, Ang P, Rozen S, Lee AS. Improving indel detection specificity of the ion torrent PGM benchtop sequencer. PLoS One. 2012;7(9):e45798.
84. Zeng F, Jiang R, Chen T. PyroHMMsnp: an SNP caller for ion torrent and 454 sequencing data. Nucleic Acids Res. 2013;41(13):e136.
85. Loman NJ, Quinlan AR. Poretools: a toolkit for analyzing nanopore sequence data. Bioinformatics. 2014;30(23):3399–401.
86. Merriman B, Ion Torrent R, Team D, Rothberg JM. Progress in ion torrent semiconductor chip based sequencing. Electrophoresis. 2012;33(23):3397–417.
87. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods. 2009;6(9):677–81.
88. Abel HJ, Al-Kateb H, Cottrell CE, Bredemeyer AJ, Pritchard CC, Grossmann AH, Wallander ML, Pfeifer JD, Lockwood CM, Duncavage EJ. Detection of gene rearrangements in targeted clinical next-generation sequencing. J Mol Diagn. 2014;16(4):405–17.
89. Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. Nat Methods. 2011;8(8):652–4.
90. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, et al. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proc Natl Acad Sci U S A. 2011;108(4):1513–8.
91. Schatz MC, Delcher AL, Salzberg SL. Assembly of large genomes using second-generation sequencing. Genome Res. 2010;20(9):1165–73.
92. Wang Y, Yang Q, Wang Z. The evolution of nanopore sequencing. Front Genet. 2014;5:449.
93. Shin SC, Ahn do H, Kim SJ, Lee H, Oh TJ, Lee JE, Park H. Advantages of single-molecule real-time sequencing in high-GC content genomes. PLoS One. 2013;8(7):e68824.
94. Vaughn CP, Robles J, Swensen JJ, Miller CE, Lyon E, Mao R, Bayrak-Toydemir P, Samowitz WS. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Hum Mutat. 2010;31(5):588–93.
95. Hansen MF, Neckmann U, Lavik LA, Vold T, Gilde B, Toft RK, Sjursen W. A massive parallel sequencing workflow for diagnostic genetic testing of mismatch repair genes. Mol Genet Genomic Med. 2014;2(2):186–200.
96. KF A, Underwood JG, Lee L, Wong WH. Improving PacBio long read accuracy by short read alignment. PLoS One. 2012;7(10):e46679.
97. Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, et al. Real-time DNA sequencing from single polymerase molecules. Science. 2009;323(5910):133–8.
98. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. Hum Mol Genet. 2010;19(R2):R227–40.
99. Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, et al. Origins of the E. coli strain causing an outbreak of hemolytic-uremic syndrome in Germany. N Engl J Med. 2011;365(8):709–17.
100. Chin CS, Sorenson J, Harris JB, Robins WP, Charles RC, Jean-Charles RR, Bullard J, Webster DR, Kasarskis A, Peluso P, Paxinos EE, Yamaichi Y, Calderwood SB, Mekalanos JJ, Schadt EE, Waldor MK. The origin of the Haitian cholera outbreak strain. N Engl J Med. 2011;364(1):33–42.
101. Huddleston J, Ranade S, Malig M, Antonacci F, Chaisson M, Hon L, et al. Reconstructing complex regions of genomes using long-read sequencing technology. Genome Res. 2014;24(4):688–96.

102. Okoniewski MJ, Meienberg J, Patrignani A, Szabelska A, Matyas G, Schlapbach R. Precise breakpoint localization of large genomic deletions using PacBio and Illumina next-generation sequencers. BioTechniques. 2013;54(2):98–100.
103. Ritz A, Bashir A, Sindi S, Hsu D, Hajirasouliha I, Raphael BJ. Characterization of structural variants with single molecule and hybrid sequencing approaches. Bioinformatics. 2014;30(24):3458–66.
104. Koren S, Schatz MC, Walenz BP, Martin J, Howard JT, Ganapathy G, Wang Z, Rasko DA, McCombie WR, Jarvis ED, Phillippy AM. Hybrid error correction and de novo assembly of single-molecule sequencing reads. Nat Biotechnol. 2012;30(7):693–700. doi:10.1038/nbt.2280.
105. Archer J, Weber J, Henry K, Winner D, Gibson R, Lee L, et al. Use of four next-generation sequencing platforms to determine HIV-1 coreceptor tropism. PLoS One. 2012;7(11):e49602.
106. Jiao X, Zheng X, Ma L, Kutty G, Gogineni E, Sun Q, et al. A benchmark study on error assessment and quality control of CCS reads derived from the PacBio RS. J Data Mining Genomics Proteomics. 2013;4(3):1–5.
107. Stoddart D, Heron AJ, Klingelhoefer J, Mikhailova E, Maglia G, Bayley H. Nucleobase recognition in ssDNA at the central constriction of the alpha-hemolysin pore. Nano Lett. 2010;10(9):3633–7.
108. Manrao EA, Derrington IM, Laszlo AH, Langford KW, Hopper MK, Gillgren N, Pavlenok M, Niederweis M, Gundlach JH. Reading DNA at single-nucleotide resolution with a mutant MspA nanopore and phi29 DNA polymerase. Nat Biotechnol. 2012;30(4):349–53.
109. Tsutsui M, He Y, Furuhashi M, Rahong S, Taniguchi M, Kawai T. Transverse electric field dragging of DNA in a nanochannel. Sci Rep. 2012;2:394.
110. Singer A, McNally B, Torre RD, Meller A. DNA sequencing by nanopore-induced photon emission. Methods Mol Biol. 2012;870:99–114.
111. Derrington IM, Butler TZ, Collins MD, Manrao E, Pavlenok M, Niederweis M, Gundlach JH. Nanopore DNA sequencing with MspA. Proc Natl Acad Sci U S A. 2010;107(37):16060–5.
112. Laszlo AH, Derrington IM, Ross BC, Brinkerhoff H, Adey A, Nova IC, Craig JM, Langford KW, Samson JM, Daza R, Doering K, Shendure J, Gundlach JH. Decoding long nanopore sequencing reads of natural DNA. Nat Biotechnol. 2014;32(8):829–33.
113. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, Wain J. O'Grady J (2014) MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol. 2015 Mar;33(3):296–300.
114. Jain M, Fiddes IT, Miga KH, Olsen HE, Paten B, Akeson M. Improved data analysis for the MinION nanopore sequencer. Nat Methods. 2015;12:351–6.
115. Mikheyev AS, Tin MM. A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour. 2014;14(6):1097–102.
116. O'Donnell CR, Wang H, Dunbar WB. Error analysis of idealized nanopore sequencing. Electrophoresis. 2013;34(15):2137–44.
117. Cavelier L, Ameur A, Haggqvist S, Hoijer I, Cahill N, Olsson-Stromberg U, Hermanson M. Clonal distribution of BCR-ABL1 mutations and splice isoforms by single-molecule long-read RNA sequencing. BMC Cancer. 2015;15:45.
118. Kastner R, Zopf A, Preuner S, Proll J, Niklas N, Foskett P, Valent P, Lion T, Gabriel C. Rapid identification of compound mutations in patients with Philadelphia-positive leukaemias by long-range next generation sequencing. Eur J Cancer. 2014;50(4):793–800.
119. Grossmann V, Kohlmann A, Klein HU, Schindela S, Schnittger S, Dicker F, et al. Targeted next-generation sequencing detects point mutations, insertions, deletions and balanced chromosomal rearrangements as well as identifies novel leukemia-specific fusion genes in a single procedure. Leukemia. 2011;25(4):671–80.
120. Baer C, Kern W, Koch S, Nadarajah N, Schindela S, Meggendorfer M, Haferlach C, Haferlach T. Ultra-deep sequencing leads to earlier and more sensitive detection of the tyrosine kinase inhibitor resistance mutation T315I in chronic myeloid leukemia. Haematologica. 2016;101(7):830–8.

121. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. Proc Natl Acad Sci U S A. 2003;100(15):8817–22.
122. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437(7057):376–80.
123. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012:251364.
124. Thiede C, Koch S, Creutzig E, Steudel C, Illmer T, Schaich M, Ehninger G. Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). Blood. 2006;107(10):4011–20.
125. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 2008;18(7):1051–63.
126. Huang YF, Chen SC, Chiang YS, Chen TH, Chiu KP. Palindromic sequence impedes sequencing-by-ligation mechanism. BMC Syst Biol. 2012;6(Suppl 2):S10.
127. Tuna M, Amos CI. Genomic sequencing in cancer. Cancer Lett. 2013;340(2):161–70.
128. Kim T, Yoshida K, Kim YK, Tyndel MS, Park HJ, Choi SH, et al. Clonal dynamics in a single AML case tracked for 9 years reveals the complexity of leukemia progression. Leukemia. 2016;30(2):295–302.
129. Farrar JE, Schuback HL, Ries RE, Wai D, Hampton OA, Trevino LR, et al. Genomic profiling of pediatric acute myeloid leukemia reveals a changing mutational landscape from disease diagnosis to relapse. Cancer Res. 2016;76(8):2197–205.
130. Clark MJ, Chen R, Lam HY, Karczewski KJ, Euskirchen G, Butte AJ, Snyder M. Performance comparison of exome DNA sequencing technologies. Nat Biotechnol. 2011;29(10):908–14.
131. Wooderchak-Donahue WL, O'Fallon B, Furtado LV, Durtschi JD, Plant P, Ridge PG, et al. A direct comparison of next generation sequencing enrichment methods using an aortopathy gene panel- clinical diagnostics perspective. BMC Med Genet. 2012;5:50.
132. Shiba N, Yoshida K, Shiraishi Y, Okuno Y, Yamato G, Hara Y, et al. Whole-exome sequencing reveals the spectrum of gene mutations and the clonal evolution patterns in paediatric acute myeloid leukaemia. Br J Haematol. 2016;175(3):476–89.
133. Masetti R, Castelli I, Astolfi A, Bertuccio SN, Indio V, Togni M, Belotti T, Serravalle S, Tarantino G, Zecca M, Pigazzi M, Basso G, Pession A. Locatelli F (2016) genomic complexity and dynamics of clonal evolution in childhood acute myeloid leukemia studied with whole-exome sequencing. Oncotarget. 2016 Aug 30;7(35):56746–57.
134. Cottrell CE, Al-Kateb H, Bredemeyer AJ, Duncavage EJ, Spencer DH, et al. Validation of a next-generation sequencing assay for clinical molecular oncology. J Mol Diagn. 2014;16(1):89–105.
135. Tsongalis GJ, Peterson JD, de Abreu FB, Tunkey CD, Gallagher TL, Strausbaugh LD, et al. Routine use of the ion torrent AmpliSeq™ cancer hotspot panel for identification of clinically actionable somatic mutations. Clin Chem Lab Med. 2014;52(5):707–14.
136. Singh RR, Patel KP, Routbort MJ, Reddy NG, Barkoh BA, Handal B, et al. Clinical validation of a next-generation sequencing screen for mutational hotspots in 46 cancer-related genes. J Mol Diagn. 2013;15(5):607–22.
137. Beadling C, Neff TL, Heinrich MC, Rhodes K, Thornton M, Leamon J, Andersen M, Corless CL. Combining highly multiplexed PCR with semiconductor-based sequencing for rapid cancer genotyping. J Mol Diagn. 2013;15(2):171–6.
138. Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, et al. Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. J Mol Diagn. 2013;15(2):234–47.
139. Escaramís G, Tornador C, Bassaganyas L, Rabionet R, Tubio JM, Martínez-Fundichely A, et al. PeSV-fisher: identification of somatic and non-somatic structural variants using next generation sequencing data. PLoS One. 2013;8(5):e63377.
140. Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics. 2009;25(21):2865–71.

141. Krumm N, Sudmant PH, Ko A, O'Roak BJ, Malig M, Coe BP, Quinlan AR, Nickerson DA, Eichler EE, Project NES. Copy number variation detection and genotyping from exome sequence data. Genome Res. 2012;22(8):1525–32.
142. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Gorlich D, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. Blood. 2016;128(5):686–98.
143. Kluk MJ, Lindsley RC, Aster JC, Lindeman NI, Szeto D, Hall D, Kuo FC. Validation and implementation of a custom next-generation sequencing clinical assay for hematologic malignancies. J Mol Diagn. 2016;18(4):507–15.
144. McKerrell T, Moreno T, Ponstingl H, Bolli N, Dias JM, Tischler G, et al. Development and validation of a comprehensive genomic diagnostic tool for myeloid malignancies. Blood. 2016;128(1):e1–9.
145. Grasso C, Butler T, Rhodes K, Quist M, Neff TL, Moore S, Tomlins SA, Reinig E, Beadling C, Andersen M. Corless CL (2015) assessing copy number alterations in targeted, amplicon-based next-generation sequencing data. J Mol Diagn. 2015 Jan;17(1):53–63.
146. Kadri S, Zhen CJ, Wurst MN, Long BC, Jiang ZF, Wang YL, Furtado LV, Segal JP. Amplicon Indel hunter is a novel bioinformatics tool to detect large somatic insertion/deletion mutations in amplicon-based next-generation sequencing data. J Mol Diagn. 2015;17(6):635–43.
147. Vidriales MB, Perez-Lopez E, Pegenaute C, Castellanos M, Perez JJ, Chandia M, et al. Minimal residual disease evaluation by flow cytometry is a complementary tool to cytogenetics for treatment decisions in acute myeloid leukaemia. Leuk Res. 2016;40:1–9.
148. Zhou Y, Othus M, Araki D, Wood BL, Radich JP, Halpern AB, et al. Pre- and post-transplant quantification of measurable ('minimal') residual disease via multiparameter flow cytometry in adult acute myeloid leukemia. Leukemia. 2016;30(7):1456–64.
149. Duployez N, Nibourel O, Marceau-Renaut A, Willekens C, Helevaut N, Caillault A, et al. Minimal residual disease monitoring in t(8;21) acute myeloid leukemia based on RUNX1-RUNX1T1 fusion quantification on genomic DNA. Am J Hematol. 2014;89(6):610–5.
150. Kohlmann A, Klein HU, Weissmann S, Bresolin S, Chaplin T, Cuppens H, et al. The Interlaboratory RObustness of next-generation sequencing (IRON) study: a deep sequencing investigation of TET2, CBL and KRAS mutations by an international consortium involving 10 laboratories. Leukemia. 2011;25(12):1840–8.
151. Grossmann V, Roller A, Klein HU, Weissmann S, Kern W, Haferlach C, Dugas M, Haferlach T, Schnittger S, Kohlmann A. Robustness of amplicon deep sequencing underlines its utility in clinical applications. J Mol Diagn. 2013;15(4):473–84.
152. Salipante SJ, Fromm JR, Shendure J, Wood BL, Wu D. Detection of minimal residual disease in NPM1-mutated acute myeloid leukemia by next-generation sequencing. Mod Pathol. 2014;27(11):1438–46.
153. Kohlmann A, Nadarajah N, Alpermann T, Grossmann V, Schindela S, Dicker F, et al. Monitoring of residual disease by next-generation deep-sequencing of RUNX1 mutations can identify acute myeloid leukemia patients with resistant disease. Leukemia. 2014;28(1):129–37.
154. Malmberg EB, Stahlman S, Rehammar A, Samuelsson T, Alm SJ, Kristiansson E, et al. Patient-tailored analysis of minimal residual disease in acute myeloid leukemia using next generation sequencing. Eur J Haematol. 2017;98(1):26–37.
155. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A, et al. Group UKNCRIAW. Assessment of minimal residual disease in standard-risk AML. N Engl J Med. 2016; 374(5):422–433.
156. Bartram J, Mountjoy E, Brooks T, Hancock J, Williamson H, Wright G, Moppett J, Goulden N, Hubank M. Accurate sample assignment in a multiplexed, ultrasensitive, high-throughput sequencing assay for minimal residual disease. J Mol Diagn. 2016;18(4):494–506.
157. Genovese G, Kähler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371(26):2477–87.
158. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371(26):2488–98.

159. Wong TN, Miller CA, Klco JM, Petti A, Demeter R, Helton NM, et al. Rapid expansion of preexisting nonleukemic hematopoietic clones frequently follows induction therapy for de novo AML. Blood. 2016;127(7):893–7.
160. Schmitt MW, Fox EJ, Prindle MJ, Reid-Bayliss KS, True LD, Radich JP, Loeb LA. Sequencing small genomic targets with high efficiency and extreme accuracy. Nat Methods. 2015;12(5):423–5.
161. Young AL, Wong TN, Hughes AE, Heath SE, Ley TJ, Link DC, Druley TE. Quantifying ultra-rare pre-leukemic clones via targeted error-corrected sequencing. Leukemia. 2015;29(7): 1608–11.
162. Hiatt JB, Pritchard CC, Salipante SJ, O'Roak BJ, Shendure J. Single molecule molecular inversion probes for targeted, high-accuracy detection of low-frequency variation. Genome Res. 2013;23(5):843–54.
163. Hokland P, Ommen HB, Mule MP, Hourigan CS. Advancing the minimal residual disease concept in acute myeloid leukemia. Semin Hematol. 2015;52(3):184–92.
164. Hourigan CS, Karp JE. Minimal residual disease in acute myeloid leukaemia. Nat Rev Clin Oncol. 2013;10(8):460–71.
165. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun. 2016;7:12484.
166. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat med. 014 Dec. 2014;20(12):1472–8.
167. Babayan A, Alawi M, Gormley M, Muller V, Wikman H, McMullin RP, et al. Comparative study of whole genome amplification and next generation sequencing performance of single cancer cells. Oncotarget. 2016; doi:10.18632/oncotarget.10701. [Epub ahead of print].
168. Sabina J, Leamon JH. Bias in whole genome amplification: causes and considerations. Methods Mol Biol. 2015;1347:15–41.
169. Pugh TJ, Delaney AD, Farnoud N, Flibotte S, Griffith M, Li HI, Qian H, Farinha P, Gascoyne RD, Marra MA. Impact of whole genome amplification on analysis of copy number variants. Nucleic Acids Res. 2008;36(13):e80.
170. Paguirigan AL, Smith J, Meshinchi S, Carroll M, Maley C, Radich JP. Single-cell genotyping demonstrates complex clonal diversity in acute myeloid leukemia. Sci Transl Med. 2015;7(281):281re282.
171. Silva FP, Swagemakers SM, Erpelinck-Verschueren C, Wouters BJ, Delwel R, Vrieling H, van der Spek P, Valk PJ, Giphart-Gassler M. Gene expression profiling of minimally differentiated acute myeloid leukemia: M0 is a distinct entity subdivided by RUNX1 mutation status. Blood. 2009;114(14):3001–7.
172. Bullinger L, Rucker FG, Kurz S, Du J, Scholl C, Sander S, Corbacioglu A, Lottaz C, Krauter J, Frohling S, Ganser A, Schlenk RF, Dohner K, Pollack JR, Dohner H. Gene-expression profiling identifies distinct subclasses of core binding factor acute myeloid leukemia. Blood. 2007;110(4):1291–300.
173. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional heterogeneity and lineage commitment in myeloid progenitors. Cell. 2015;163(7):1663–77.
174. Yan P, Frankhouser D, Murphy M, Tam HH, Rodriguez B, Curfman J, et al. Genome-wide methylation profiling in decitabine-treated patients with acute myeloid leukemia. Blood. 2012;120(12):2466–74.
175. Metzeler KH, Heilmeier B, Edmaier KE, Rawat VP, Dufour A, Dohner K, et al. High expression of lymphoid enhancer-binding factor-1 (LEF1) is a novel favorable prognostic factor in cytogenetically normal acute myeloid leukemia. Blood. 2012;120(10):2118–26.
176. Kuhnl A, Valk PJ, Sanders MA, Ivey A, Hills RK, Mills KI, et al. Downregulation of the Wnt inhibitor CXXC5 predicts a better prognosis in acute myeloid leukemia. Blood. 2015;125(19):2985–94.
177. de Jonge HJ, Valk PJ, Veeger NJ, ter Elst A, den Boer ML, Cloos J, et al. High VEGFC expression is associated with unique gene expression profiles and predicts adverse prognosis in pediatric and adult acute myeloid leukemia. Blood. 2010;116(10):1747–54.

178. Piazza R, Cecchetti C, Pirola A, Donandoni C, Fontana D, Mezzatesta C, et al. RNA-seq is a valuable complement of conventional diagnostic tools in newly diagnosed AML patients. Am J Hematol. 2015;90(12):E227–8.
179. Panagopoulos I, Gorunova L, Zeller B, Tierens A, Heim S. Cryptic FUS-ERG fusion identified by RNA-sequencing in childhood acute myeloid leukemia. Oncol Rep. 2013;30(6):2587–92.
180. Zhou JB, Zhang T, Wang BF, Gao HZ, Xu X. Identification of a novel gene fusion RNF213SLC26A11 in chronic myeloid leukemia by RNA-Seq. Mol Med Rep. 2013;7(2): 591–7.
181. Guglielmelli P, Bisognin A, Saccoman C, Mannarelli C, Coppe A, Vannucchi AM, Bortoluzzi S. Small RNA sequencing uncovers new miRNAs and moRNAs differentially expressed in normal and primary Myelofibrosis CD34+ cells. PLoS One. 2015;10(10):e0140445.
182. Wilhelm BT, Briau M, Austin P, Faubert A, Boucher G, Chagnon P, Hope K, Girard S, Mayotte N, Landry JR, Hebert J, Sauvageau G. RNA-seq analysis of 2 closely related leukemia clones that differ in their self-renewal capacity. Blood. 2011;117(2):e27–38.
183. Spinelli R, Pirola A, Redaelli S, Sharma N, Raman H, Valletta S, et al. Identification of novel point mutations in splicing sites integrating whole-exome and RNA-seq data in myeloproliferative diseases. Mol Genet Genomic Med. 2013;1(4):246–59.
184. Brooks AN, Choi PS, de Waal L, Sharifnia T, Imielinski M, Saksena G, et al. A pan-cancer analysis of transcriptome changes associated with somatic mutations in U2AF1 reveals commonly altered splicing events. PLoS One. 2014;9(1):e87361.
185. Li XY, Yao X, Li SN, Suo AL, Ruan ZP, Liang X, Kong Y, Zhang WG, Yao Y. RNA-Seq profiling reveals aberrant RNA splicing in patient with adult acute myeloid leukemia during treatment. Eur Rev Med Pharmacol Sci. 2014;18(9):1426–33.
186. Rathe SK, Moriarity BS, Stoltenberg CB, Kurata M, Aumann NK, Rahrmann EP, et al. Using RNA-seq and targeted nucleases to identify mechanisms of drug resistance in acute myeloid leukemia. Sci Rep. 2014;4:6048.
187. Swanson L, Robertson G, Mungall KL, Butterfield YS, Chiu R, Corbett RD, et al. Barnacle: detecting and characterizing tandem duplications and fusions in transcriptome assemblies. BMC Genomics. 2013;14:550.
188. Francis RW, Thompson-Wicking K, Carter KW, Anderson D, Kees UR, Beesley AH. FusionFinder: a software tool to identify expressed gene fusion candidates from RNA-Seq data. PLoS One. 2012;7(6):e39987.
189. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly parallel genome-wide expression profiling of individual cells using Nanoliter droplets. Cell. 2015;161(5):1202–14.
190. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell. 2015;161(5):1187–201.
191. Nestorowa S, Hamey FK, Pijuan Sala B, Diamanti E, Shepherd M, Laurenti E, Wilson NK, Kent DG, Gottgens B. A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. Blood. 2016;128(8):e20–31.
192. Spencer DH, Young MA, Lamprecht TL, Helton NM, Fulton R, O'Laughlin M, et al. Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. Leukemia. 2015;29(6):1279–89.
193. Haladyna JN, Yamauchi T, Neff T, Bernt KM. Epigenetic modifiers in normal and malignant hematopoiesis. Epigenomics. 2015;7(2):301–20.
194. Fong CY, Morison J, Dawson MA. Epigenetics in the hematologic malignancies. Haematologica. 2014;99(12):1772–83.
195. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059–74.
196. Wakita S, Yamaguchi H, Omori I, Terada K, Ueda T, Manabe E, et al. Mutations of the epigenetics-modifying gene (DNMT3a, TET2, IDH1/2) at diagnosis may induce FLT3-ITD at relapse in de novo acute myeloid leukemia. Leukemia. 2013;27(5):1044–52.
197. Ahn JS, Kim HJ, Kim YK, Jung SH, Yang DH, Lee JJ, et al. Adverse prognostic effect of homozygous TET2 mutation on the relapse risk of acute myeloid leukemia in patients of normal karyotype. Haematologica. 2015;100(9):e351–3.

198. Voso MT, Lo-Coco F, Fianchi L. Epigenetic therapy of myelodysplastic syndromes and acute myeloid leukemia. Curr Opin Oncol. 2015;27(6):532–9.
199. Gill H, Leung AY, Kwong YL. Molecularly targeted therapy in acute myeloid leukemia. Future Oncol. 2016;12(6):827–38.
200. Bravo GM, Lee E, Merchan B, Kantarjian HM, Garcia-Manero G. Integrating genetics and epigenetics in myelodysplastic syndromes: advances in pathogenesis and disease evolution. Br J Haematol. 2014;166(5):646–59.
201. Gronbaek K, Muller-Tidow C, Perini G, Lehmann S, Bach Treppendahl M, et al.; Epigenomics Study on MDS, Aml CABM (2012) A critical appraisal of tools available for monitoring epigenetic changes in clinical samples from patients with myeloid malignancies. Haematologica 97(9):1380–8.
202. McDevitt MA. Clinical applications of epigenetic markers and epigenetic profiling in myeloid malignancies. Semin Oncol. 2012;39(1):109–22.
203. Holstege H, Pfeiffer W, Sie D, Hulsman M, Nicholas TJ, Lee CC, et al. Somatic mutations found in the healthy blood compartment of a 115-yr-old woman demonstrate oligoclonal hematopoiesis. Genome Res. 2014;24(5):733–42.
204. Swisher EM, Harrell MI, Norquist BM, Walsh T, Brady M, Lee M, et al. Somatic mosaic mutations in PPM1D and TP53 in the blood of women with ovarian carcinoma. JAMA Oncol. 2016;2(3):370–2.
205. Tang G, Goswami RS, Liang CS, Bueso-Ramos CE, Hu S, DiNardo C, Medeiros LJ. Isolated del(5q) in patients following therapies for various malignancies may not all be clinically significant. Am J Clin Pathol. 2015;144(1):78–86.
206. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, Ebert BL. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. Blood. 2015;126(1):9–16.
207. Bianchi DW, Chudova D, Sehnert AJ, Bhatt S, Murray K, Prosen TL, Garber JE, Wilkins-Haug L, Vora NL, Warsof S, Goldberg J, Ziainia T, Halks-Miller M. Noninvasive prenatal testing and incidental detection of occult maternal malignancies. JAMA. 2015;314(2):162–9.
208. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. Cell-free DNA comprises an in vivo Nucleosome footprint that informs its tissues-of-origin. Cell. 2016;164(1–2):57–68.