Next-generation sequencing-based multigene mutational screening for acute myeloid leukemia using MiSeq: applicability for diagnostics and disease monitoring

Rajyalakshmi Luthra,[1] Keyur P. Patel,[1] Neelima G. Reddy,[1] Varan Haghshenas,[1] Mark J. Routbort,[1] Michael A. Harmon,[1] Bedia A. Barkoh,[1] Rashmi Kanagal-Shamanna,[1] Farhad Ravandi,[2] Jorge E Cortes,[2] Hagop M Kantarjian,[2] L. Jeffrey Medeiros,[1] and Rajesh R. Singh[1]

[1]Departments of Hematopathology and [2]Leukemia, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA

ABSTRACT

Routine molecular testing in acute myeloid leukemia involves screening several genes of therapeutic and prognostic significance for mutations. A comprehensive analysis using single-gene assays requires large amounts of DNA, is cumbersome and timely consolidation of results for clinical reporting is challenging. High throughput, next-generation sequencing platforms widely used in research have not been tested vigorously for clinical application. Here we describe the clinical application of MiSeq, a next-generation sequencing platform to screen mutational hotspots in 54 cancer-related genes including genes relevant in acute myeloid leukemia (*NRAS*, *KRAS*, *FLT3*, *NPM1*, *DNMT3A*, *IDH1/2*, *JAK2*, *KIT* and *EZH2*). We sequenced 63 samples from patients with acute myeloid leukemia/myelodysplastic syndrome using MiSeq and compared the results with those obtained using another next-generation sequencing platform, Ion-Torrent Personal Genome Machine and other conventional testing platforms. MiSeq detected a total of 100 single nucleotide variants and 23 *NPM1* insertions that were confirmed by Ion Torrent or conventional platforms, indicating complete concordance. *FLT3*-internal tandem duplications (n=10) were not detected; however, reanalysis of the MiSeq output by Pindel, an indel detection algorithm, did detect them. Dilution studies of cancer cell-line DNA showed that the quantitative accuracy of mutation detection was up to an allelic frequency of 1.5% with a high level of inter- and intra-run assay reproducibility, suggesting potential utility for monitoring response to therapy, clonal heterogeneity and evolution. Examples demonstrating the advantages of MiSeq over conventional platforms for disease monitoring are provided. Easy work-flow, high throughput multiplexing capability, 4-day turnaround time and simultaneous assessment of routinely tested and emerging markers make MiSeq highly applicable for clinical molecular testing in acute myeloid leukemia.

Introduction

Acute myeloid leukemia (AML) is commonly characterized by genetic aberrations, including gene mutations, copy number alterations and gene translocations.[1,2] In recent years, massively parallel next generation sequencing (NGS) technology has shown great potential in identifying AML-related genetic aberrations.[3-6] The current 2008 World Health Organization classification system requires genetic information, in addition to clinical, morphological and immunophenotypic findings, for the diagnosis and classification of AML. Several types of AML are now defined by the presence of recurrent chromosomal translocations that result in fusion genes, and two new provisional categories have been proposed based on the prognostic impact of these single gene defects: AML with mutated *NPM1* and AML with mutated *CEBPA*.[7] Proposed, new risk stratification models for AML patients incorporate data from cytogenetic studies and mutational profiling of several critical genes, such as *FLT3*, *NPM1*, *CEPBA*, *DNMT3A*, *IDH1*, *IDH2*, *KIT*, *MLL-PTD*, *TET2*, *RUNX1*, *ASXL1* and *TP53*.[8,9] For accurate risk stratification, it is important to analyze the mutations in several genes comprehensively because of complex interactions among different pathways in leukemogenesis.[9-11] Mutations in some of these genes, such as *RAS* and *FLT3*, can currently be targeted with inhibitors, and it will likely be possible to target other genes in the future. Recent literature also suggests that quantitative assessment of gene mutations may be useful to monitor minimal residual disease and predict the risk of relapse.[11] Routine molecular diagnostic testing of AML patients is, therefore, indispensable for rational therapeutic decisions and includes mutational screening of several genes, which are either targetable for therapy or useful for prognostication.[12]

In a clinical molecular diagnostic laboratory, multigene screening with conventional platforms proves challenging due to the need for relatively large quantities of DNA to assess one gene at a time and the coordination and compilation of the results from several analysis platforms into an integrated report. The massively parallel sequencing ability of NGS technologies has made high throughput and multiplexed sequencing of specific panels of genes or whole exomes/genomes feasible.[13-15] These technologies are currently being applied extensively to characterize genomic alterations in cancer[16-19] and are highly relevant for diagnostic purposes as alternatives to first-generation sequencing techniques.[20] However, inclusion of NGS technologies into the diagnostic arena has been delayed due to the lack of adequate guidelines and rigorous validation.

 doi:10.3324/haematol.2013.093765
The online version of this article has a Supplementary Appendix.
Manuscript received on June 24, 2013. Manuscript accepted on October 16, 2013.
Correspondence: rluthra@mdanderson.org or rsingh@mdanderson.org

MiSeq is a next-generation sequencer that is well-suited for various targeted sequencing applications and generates genomic sequence information with high accuracy.[21-27] In this study, we used MiSeq and a customized TruSeq Amplicon Cancer Panel (TSACP) to screen for mutations in 54 cancer-related genes in samples from 63 patients (60 with AML and 3 with myelodysplastic syndrome) in our Clinical Laboratory Improvement Amendments (CLIA)-certified laboratory. We demonstrate the value of this approach, which requires less DNA and provides data for multiple genes simultaneously in an integrated molecular diagnostic laboratory report with a short turnaround time.

Methods

Tumor samples and sequencing platforms

Sixty-three samples (62 bone marrow aspirates and 1 peripheral blood sample) from patients with a diagnosis of AML (n=60) or myelodysplasia (n=3) were assessed using a MiSeq sequencer (illumina Inc, San Diego, CA, USA). DNA was extracted using an Autopure extractor (QIAGEN/Gentra, Valencia, CA, USA) and quantified using a Qubit DNA BR assay kit (Life Technologies, Carlsbad, CA, USA). Of the 63 samples, 52 were selected from the archives of the Molecular Diagnostics Laboratory at The University of Texas MD Anderson Cancer Center based on their mutational status ascertained by conventional mutation detection platforms in the laboratory. Forty-four of the archival samples were sequenced using both the MiSeq and an Ion-Torrent Personal Genome Machine (IT-PGM) (Life Technologies). Eleven additional samples were sequenced on the MiSeq instrument in parallel (blind) with routinely used platforms, including Sanger sequencing, pyrosequencing and fragment analysis by capillary electrophoresis, performed as described previously.[28-30] To confirm mutations detected in regions covered by TSACP but not by pre-existing assays in our laboratory (AmpliSeq cancer panel, Sanger or pyro sequencing), 42 new Sanger sequencing assays were also developed.

Customized TruSeq Amplicon Cancer Panel

The TSACP interrogates mutational hotspots in 48 cancer-related genes (listed in the *Online Supplementary Methods*). The TSACP consists of 212 pairs of probes designed to bind flanking genomic areas of interest. Additionally, seven more probe pairs were designed by illumina on our request and were spiked into the original TSCAP to interrogate mutational hotspots in three additional AML-related genes (*DNMT3A, IDH2* and *EZH2*) and three chronic lymphocytic leukemia-related genes (*XPO1, KLHL6* and *MYD88*). Library preparation and sequencing using MiSeq was performed according to the manufacturer's instruction (details provided in the *Online Supplementary Methods*).

Validation of the version 2 upgrade on MiSeq

The version 2 (V2) upgrade on MiSeq has changes in sequencing chemistry and imaging to allow sequencing twice as many samples as possible with version 1 (V1). To validate the accuracy and efficiency of the V2 upgrade, we sequenced in parallel 12 archival AML samples (additional to the above mentioned cohort) using MiSeq V1 and the V2 upgrade. To compensate for the higher sequencing capacity of V2, 12 additional samples were included in the sequencing run with V2, which facilitated optimal comparison.

Variant calling and data analysis

Human genome build 19 (hg19) was used as the reference. Alignment to the hg19 genome and variant calling was performed by MiSeq Reporter Software 1.3.17. Integrative Genomics Viewer (IGV)[31] was used to visualize the read alignment and confirm the variant calls. A sequencing coverage of 250X (bi-directional) and a minimum variant frequency of 5% in the background of wild-type (WT) were used as cutoffs for clinical reporting. Custom-developed, in-house software,[32] designated as OncoSeek, was used to interface the data with IGV and to annotate the sequence variants. For identification of *FLT3* internal tandem duplications (ITD), Pindel[33] analysis (source: *http://www.ebi.ac.uk/~kye/pindel/* and version Pindel0.2.4t) was performed using the BAM files.

Sequencing using the Ion Torrent Personal Genome Machine

Library preparation and sequencing on IT-PGM was performed as described earlier.[34] Details are provided in the *Online Supplementary Methods*.

Results

MiSeq sequencing run metrics

In 16 different V1 sequencing runs, the average cluster density of sequencing template generated per square millimeter (mm^2) in the flow cell ranged from 502,000 to 1,186,000 clusters/mm^2 with a median of 950,000 clusters/mm^2. The clusters which passed filter or clusters from which sequencing information could be obtained (without signal overlap from surrounding clusters) ranged from 460,000 to 1,065,000 clusters/mm^2 with a median value of 851,950/mm^2 (*Online Supplementary Figure S1A*). The total sequencing output for the runs ranged from 1.1 to 2.2 gigabases (Gb) with sequencing quality score > Q30 with a median value of 1.9 Gb (*Online Supplementary Figure S1B*). The total sequencing reads obtained from the run ranged from 4,079,435 to 9,570,965, with a median value of 7,424,774. The identified sequencing reads or reads pass filter (with chastity scores of ≥0.6) ranged from 3,716,549 to 8,080,211 reads with a median value of 6,641,052 (*Online Supplementary Figure S1C*) indicating that 72.4-96.1% of the reads (median, 95.3%) were 'identified reads' or reads that could be identified with the barcode indexes used (*Online Supplementary Figure S1D*). The average sequencing depth per base per sample in these runs was 1455X, which indicated adequate sequencing of the samples.

Sensitivity of mutation detection

To measure the mutation detection sensitivity, we sequenced H2122 cell line DNA (with a homozygous *KRAS* mutation, p.G12C and a heterozygous *MET* mutation, p.N375S) diluted into HL60 cell line DNA to provide different levels of H2122 DNA, including 100% (undiluted), 25% (1:3 dilution), 10% (1:9 dilution) and 5% (1:19 dilution). Sequencing results from two independent experiments showed efficient detection of these two mutations. The presence of mutations at all dilutions was clearly evident in the aligned sequencing reads for *KRAS* (p.G12C) (Figure 1A, upper panel) and *MET* (p.N375S) (Figure 1A, lower panel). The concordance between the average variant frequencies detected at different dilution levels and the expected frequencies are shown in Figure 1B, upper and lower panels.

Concordance between MiSeq and Ion Torrent Personal Genome Machine results

A high degree of concordance was observed for the 44

tumor samples analyzed in parallel using the MiSeq and IT-PGM sequencers. A total of 110 variants (single nucleotide variations and insertions) were detected by MiSeq, 102 of which were confirmed by IT-PGM. The eight variants not detected by IT-PGM were not covered by the AmpliSeq Cancer panel but were confirmed by Sanger sequencing (*Online Supplementary Table S1*). Several variants listed in this table are potential germ-line polymorphisms (*KDR* p.Q472H, *MET* p.N375S, *MET* p.T1010I and *KIT* p.M541L). They were also listed in the comparison study (*Online Supplementary Table S1*) as they help in establishing the overall mutation detection accuracy and specificity of MiSeq. A high degree of concordance in variant detection was also observed between MiSeq and Sanger sequencing, pyrosequencing and fragment analysis by capillary electrophoresis. For example, in a representative archival sample a *KRAS* p.G60V (GGT>GTT) mutation was detected by pyrosequencing (Figure 2A) and was detected at a variant frequency of 45.6% at a sequencing depth of 3931X by MiSeq (Figure 2B). IT-PGM also detected this mutation at a comparable frequency of 39% (3120X) (Figure 2C). Similarly, an *IDH2* p.R132H (CGT>CAT) (Figure 3A) mutation detected in a sample by Sanger sequencing was also detected by MiSeq (42.0%, 3,672X) (Figure 3B) and IT-PGM (43.1%, 3,318X) (Figure 3C). In a third example, a 4bp insertion (G>GTCTG) detected in exon 12 of *NPM1* by fragment analysis coupled to capillary electrophoresis at a frequency of 42.6% (Figure 4A) was detected by MiSeq (32.8%, 1,164X) (Figure 4B) and also by IT-PGM (26.5%, 2,249X) (Figure 4C).

Parallel comparison of mutation detection using MiSeq and routinely used testing platforms

Eleven samples were sequenced using the MiSeq in parallel (blind) with conventional techniques used routinely in our laboratory. In these samples, 14 sequence variants detected by MiSeq were confirmed by Sanger sequencing (n=8), pyrosequencing (n=5) and fragment analysis by capillary electrophoresis (n=1) indicating complete concordance (*Online Supplementary Table S2*).

Detection of insertions by MiSeq

In our archival set of 44 samples, 23 samples had a 4 bp insertion in *NPM1* and two samples had *FLT3*-ITD. The

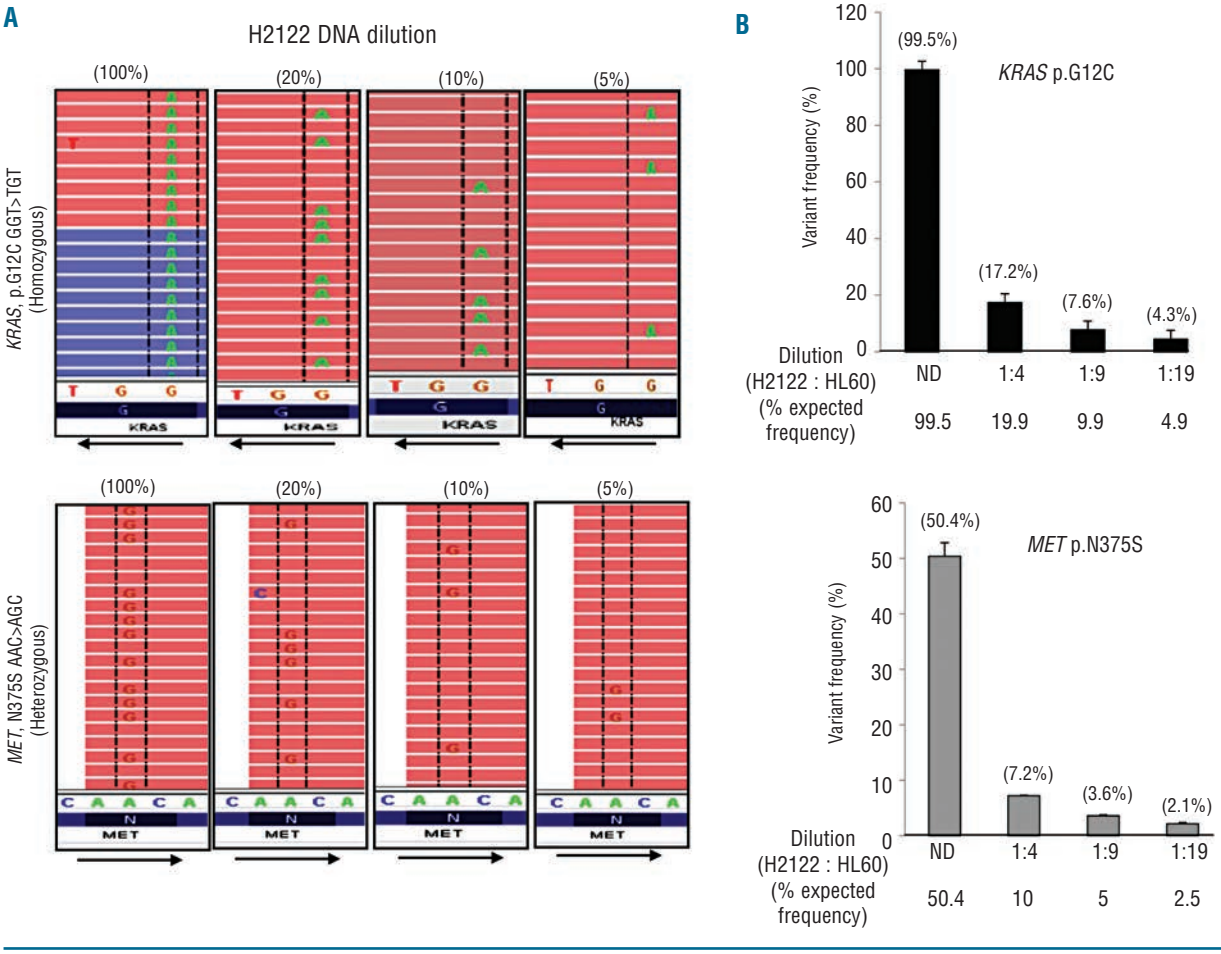

Figure 1. Sensitivity of detection using MiSeq. (A) A representative image of the aligned sequencing reads as visualized in Integrative Genome Viewer shows a progressive decrease in the single nucleotide variant detected in the background of wild-type sequence in H2122 cell line DNA sequentially diluted into HL60 cell line DNA. With dilution, a clear and proportional decrease in the homozygous *KRAS* (GGT>TGT, p.G12C) and a heterozygous *MET* (AAC>AGC, p.N375S) mutation are evident. *KRAS* gene has the reverse orientation on chromosome 12. However, by default, the Integrative Genome Viewer exhibits aligned reads in 'forward' orientation. Hence the substituted nucleotide appears as 'A' instead of 'T' in the reads (CCA>ACA or GGT>TGT). The directions of gene orientation are indicated by the arrows. (B) The expected variant frequencies and the average variant frequency of the *KRAS* (p.G12C) and *MET* (p.N375S) mutations detected in two independent sensitivity analyses.

MiSeq reporter detected all 4 bp insertions in *NPM1* but did not detect the two *FLT3*-ITD (samples n. 3 and 40, *Online Supplementary Table S1*). These ITD were also not detected by IT-PGM. To further assess the ability of MiSeq reporter to call *FLT3*-ITD, we sequenced an additional eight archival samples with known *FLT3*-ITD of various sizes and allele frequencies as detected by capillary electrophoresis. MiSeq did not call any of the *FLT3*-ITD present in these samples. However, using Pindel, an indel detection algorithm, we were able to detect and map each of the ITD in the MiSeq sequencing output with complete concordance with the capillary electrophoresis-based fragment analysis method (*Online Supplementary Table S3*).

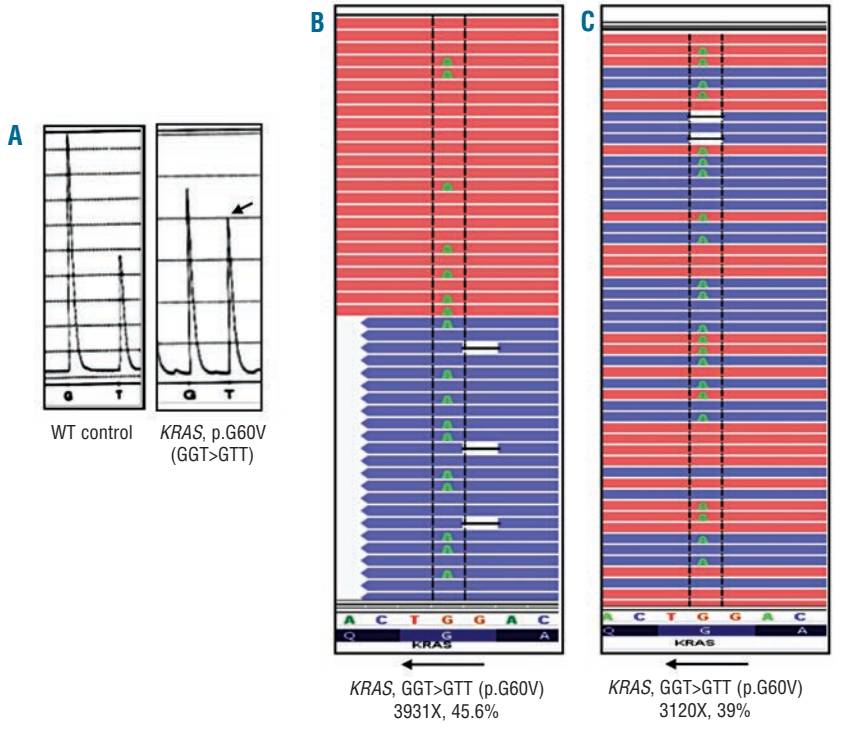

Figure 2. Concordance of MiSeq, pyrosequencing and Ion Torrent Personal Genome Machine (IT-PGM) findings. A representative sample in which a *KRAS* mutation (p.G60V) identified by pyrosequencing was also detected and called by both MiSeq and the IT-PGM. (A) The pyrosequencing results showing the base change (GGT>GTT, indicated by the arrow) in *KRAS* in comparison with wild-type control. (B) The same mutation change is evident in the aligned reads from the MiSeq sequencing output at a very high coverage of 3,931X and showing a variant frequency of 45.6%. (C) Sequencing on the IT-PGM also confirms the same mutation at a sequencing depth and variant frequency comparable with the MiSeq. The *KRAS* gene has the reverse orientation on chromosome 12. As the default setting on Integrative Genome Viewer exhibits aligned reads in 'forward' orientation, the substituted nucleotide appears as 'A' instead of 'T' in the reads (CCA>CAA or GGT>GTT). The orientation of the gene is depicted by the arrows in panels (B) and (C).

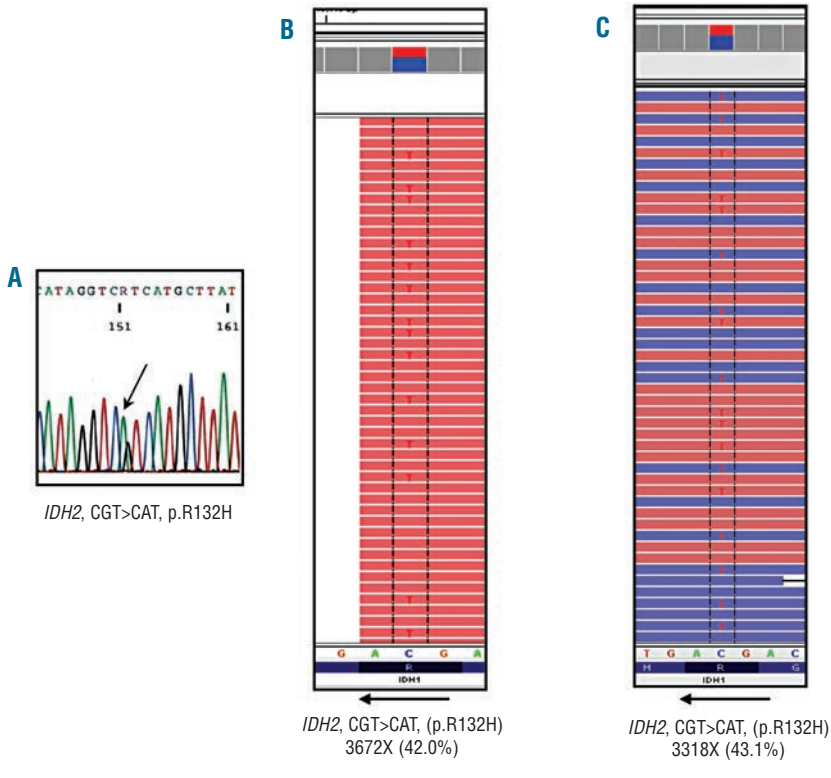

Figure 3. Concordance of MiSeq, Sanger sequencing and Ion Torrent Personal Genome Machine (IT-PGM). A mutation in *IDH2* (p.R132H) resulting from a CGT>CAT substitution originally detected by Sanger sequencing (A) is also detected clearly in the MiSeq sequencing output and (B) sequencing using the IT-PGM with comparable coverage and mutational frequency with both platforms. *IDH2* has reverse orientation on chromosome 15. The default setting on Integrative Genome Viewer shows aligned reads in forward orientation only. Hence, the substituted nucleotide appears as 'T' instead of 'A' in the reads (CGT>CAT or GCA>GTA). The orientation of the gene is depicted by the arrows in panels (B) and (C).

Comparison of version 1 and version 2 sequencing on MiSeq

During our study a MiSeq upgrade was introduced with changes in the sequencing chemistry and imaging (referred to as version 2 or V2), which doubles the throughput capacity in comparison with version 1. To validate the upgrade, we sequenced 12 samples using MiSeq V1 and V2 in parallel. To compensate for the higher sequencing capacity of V2 and to facilitate an optimal comparison, we included 12 additional samples in the sequencing run with V2. A comparison of the run metrics showed comparable cluster density and clusters pass filter between the two versions (*Online Supplementary Figure S2A*). A higher sequencing capacity of V2 was evident from the 4.7 Gb sequencing output in comparison with the 2.1 Gb output of V1, whereas the sequencing quality (Q30) remained comparable (*Online Supplementary Figure S2B*). The increased sequencing capacity of V2 was evident in the doubling of the total sequencing reads and reads pass filter with the total overall percentage of reads identified remaining comparable (*Online Supplementary Figure S2C*, upper panel). Similarly, total reads and average coverage per base per sample remained highly comparable between V1 and V2 (*Online Supplementary Figure S2C*, lower panel). Furthermore, a high degree of concordance was observed in the sequencing coverage and mutations detected by V1 and V2 (*Online Supplementary Table S4*) indicating that the V2 upgrade maintained the same efficiency and accuracy as V1.

Sensitivity of mutation detection using version 2 sequencing chemistry

Sensitivity analysis was performed using several sequentially diluted samples of DNA from the cell line DLD1 (harboring 8 heterozygous mutations) into normal control DNA. Across six different sequencing runs each of eight heterozygous (monoallelic) mutations present in DLD1 were successfully detected in four sequential dilutions (50%, 25%, 10% and 5%) with minimal variation in the overall variant frequencies. The lowest detection limit was seen for the *IDH1* p.G97D variant detected at a frequency of 1.5% in the 1:19 (5%) diluted DLD1 sample, indicating high detection sensitivity (Figure 5). *Online Supplementary Table S5* provides a detailed summary of the dilution studies with H2122 (V1) and DLD1 DNA (V2) with expected and detected variant frequencies of the mutations at each dilution level.

Inter- and intra-run reproducibility

Inter-run reproducibility was assessed by sequencing a 1:9 (10%) dilution of DLD1 DNA in 11 separate multiplexed sequencing runs. The results showed that each of the eight expected mutations was consistently detected in every run with minimal variability in variant frequencies (*Online Supplementary Figure S3A*). Intra-run assay reproducibility was assessed by sequencing the 1:9 (10%) DLD1 DNA with 24 different barcode indexes and sequencing them on the same multiplexed sequencing run. Again, each mutation in the 24 samples analyzed was

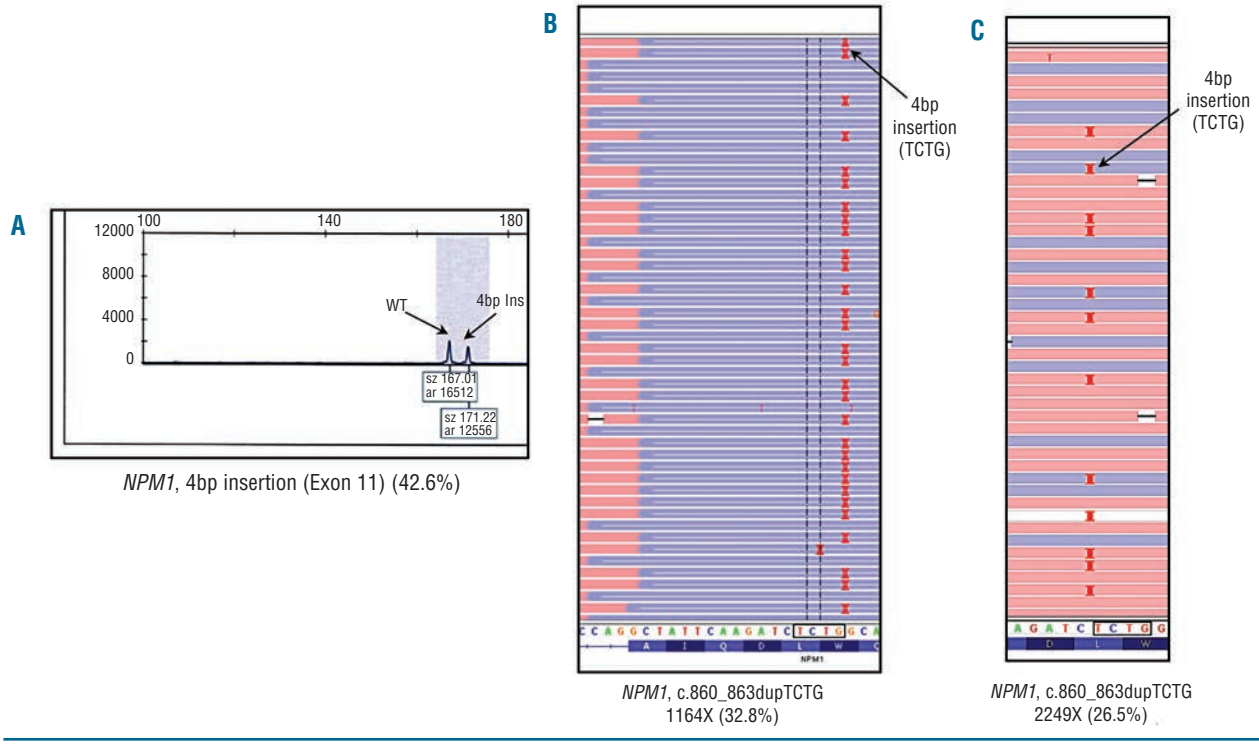

Figure 4. Ability of sequencing using MiSeq to detect insertions in *NPM1*. (A) A characteristic 4 bp insertion (G>GTCTG) in exon 12 of the *NPM1* gene is evident in this sample as detected by capillary electrophoresis and is present at a ratio of 0.426 or 42.6%. (B) Sequencing of the same sample using MiSeq also detected and called this mutation at a sequencing depth of 1,164X and a variant frequency of 32.8%. (C) The Ion Torrent Personal Genome Machine (IT-PGM) also detected and called the same 4 bp insertion at a sequencing depth of 2,249X and a variant frequency of 26.5% in the background of wild-type sequence. The presence of the insertions in the aligned reads is indicated as characteristic red bars in Integrative Genome Viewer as seen in panels (B) and (C) (indicated by the arrows).

consistently detected with comparable variant frequencies, indicating high inter-assay reproducibility (*Online Supplementary Figure S3B*).

Disease monitoring or assessment of minimal residual disease

We evaluated the utility of highly sensitive sequencing on MiSeq in monitoring minimal residual disease using four representative cases with archival samples analyzed and reported in our laboratory by routinely used platforms and subsequently sequenced by MiSeq for this study. For patient 1 (Figure 6A), we used MiSeq to sequence five consecutive samples taken pre-treatment (day 1) and post-treatment (days 10, 185, 455, and 475) and compared the results with those obtained using conventional test platforms. Initially (day 1) the sample had two mutations, *IDH2* p.R140Q and a 4 bp insertion in exon 12 of *NPM1*, which were detected by Sanger sequencing analysis and fragment analysis, respectively. MiSeq also detected these mutations. In subsequent samples taken 10 and 185 days after treatment, MiSeq detected the *IDH2* p.R140Q mutation at allele frequencies of 5.5% and 2.8%, respectively, and the *NPM1* mutation at allele frequencies of 6% and 2%, respectively. By contrast, Sanger sequencing did not detect any *IDH2* mutation in either the day 10 or day 185 sample and was reported negative, highlighting the greater sensitivity of MiSeq. Capillary electrophoresis did detect the 4 bp insertions in *NPM1* in these two samples at frequencies of 13.6% and 10.5%, respectively. In later samples obtained at 455 and 475 days, MiSeq detected a *FLT3* point mutation p.D835Y as well as the *IDH2* and *NPM1* mutations originally detected in the sample, which were also, detected using the routine alternative platforms (Figure 6A).

In the second patient (Figure 6B), mutations in four genes (*NRAS*, *NPM1*, *DNMT3A* and *FLT3*) were detected at diagnosis and in the day 35 sample. At 55 and 95 days the *NRAS* and *NPM1* mutations had decreased drastically while the *FLT3*-ITD and *DNMT3A* mutations persisted, indicating the presence of distinct clonal populations with different somatic mutations. Furthermore, at 95 days two low frequency *FLT3* p.D835 mutations appeared, increasing the total number of mutations to be followed for disease monitoring to six and indicating the need to follow multiple mutations, which can be conveniently achieved on MiSeq.

In the third example (Figure 6C), two *NRAS* mutations (p.G12S and p.G12C) were detected at diagnosis. The p.G12S mutation decreased considerably with treatment indicating that the two mutations occurred in distinct clones. An activating *FLT3* kinase domain mutation (p.Y842H)[35] with implications in therapy response was observed at low levels at diagnosis and its allelic frequency increased progressively over time. As this mutation is not routinely tested for in AML, it was not tested by conventional platforms but was identified by virtue of multigene screening on MiSeq.

In the fourth sample (Figure 6D), four somatic mutations (*IDH2*, *NRAS*, *NPM1* and *DNMT3A*) were detected at the onset of the disease, but only the *DNMT3A* mutation remained detectable after 33 days.

In these cases, the ability to monitor multiple somatic mutations at high sensitivity using a single platform, MiSeq, highlights the advantages of the NGS platforms which not only help in conserving precious samples but also in improving analyses and clinical reporting workflows.

Discussion

AML is a genetically heterogeneous disease and history has shown that identification of genomic abnormalities enables separation into prognostically meaningful subgroups with therapeutic implications. However, more accurate identification of genetic lesions is needed to improve disease classification, risk stratification and treatment strategies. Comprehensive routine screening using conventional or first-generation sequencing platforms in a clinical molecular diagnostics laboratory is very challenging since it requires large amounts of DNA and is labor-intensive, expensive, and logistically difficult in terms of

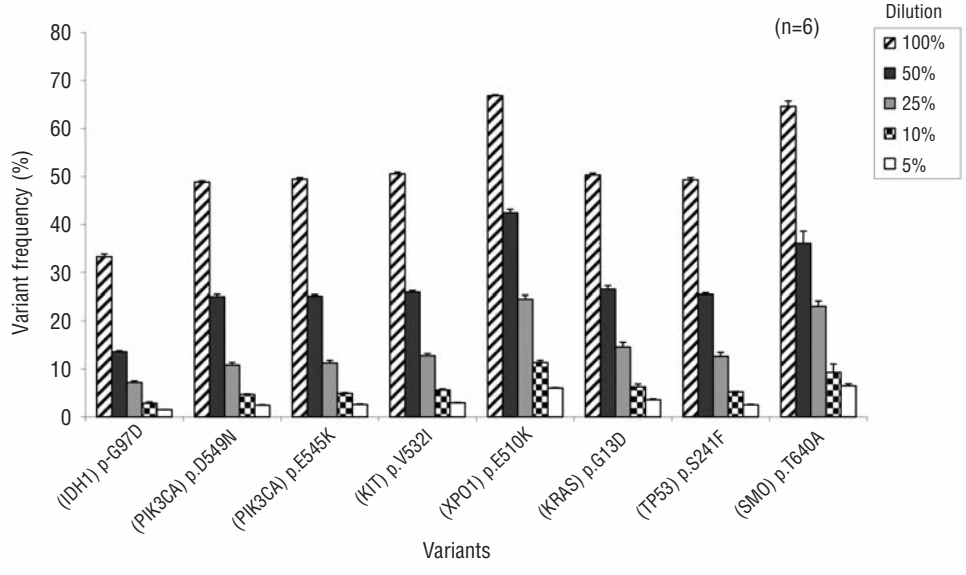

Figure 5. Assay sensitivity of the V2 upgrade. Undiluted (100%) and samples of DLD1 cell line DNA sequentially diluted into normal female control DNA (50%, 25%, 10% and 5%) were sequenced in six different sequencing runs and the ability to detect the eight different heterozygous (monoallelic) mutations present in each dilution was tested. These mutations were successfully detected in every dilution tested in all of the sequencing runs with minimal variation in the variant frequency detected.

data integration in real-time. Next-generation technologies are highly advantageous with their decreasing cost, high throughput and multiplexing capacity, thereby facilitating parallel and targeted sequencing of all genes of interest on a routine basis. However, sample preparation, sequencing and variant calling components of NGS technologies need to be validated before they can be applied for routine diagnostic purposes.

In recent years, NGS studies have revealed numerous previously unrecognized or under-recognized molecular changes of therapeutic and prognostic significance. Hence, upfront screening for genetic abnormalities and their subsequent monitoring are essential for optimal management of patients. Here we have rigorously tested the applicability of a NGS-based screen using MiSeq by analyzing both archival samples with known mutations and samples assessed prospectively. For single nucleotide variants, complete concordance was observed between MiSeq and all other sequencing methods including the IT-PGM, another NGS platform. MiSeq also successfully detected all 4 bp insertions in *NPM1* in this cohort. MiSeq software, however, could not identify *FLT3*-ITD. A similar deficiency was also observed for IT-PGM. To detect *FLT3*-ITD using MiSeq, it was necessary to re-analyze the aligned sequencing data using another program, namely Pindel.

The sequencing results using MiSeq showed a high level of inter- and intra-run reproducibility with consistent detection of mutations. Dilution studies with two cell lines harboring a total of ten mutations across seven chromosomes showed that MiSeq could detect every mutation at each dilution level tested with a variant frequency as low as 1.5%. This high detection sensitivity is advantageous for better monitoring of the efficacy of therapy as evident in one of the examples included in the study in which an *IDH2* mutation missed by Sanger sequencing in two longitudinal samples was detected by MiSeq (Figure 6A). This mutation recurred at high levels subsequently indicating re-emergence of the same clone. The high detection sensitivity of MiSeq also has the potential to detect new clones at earlier time points than possible with less sensitive conventional mutation detection methods. Furthermore, the high throughput capacity of MiSeq

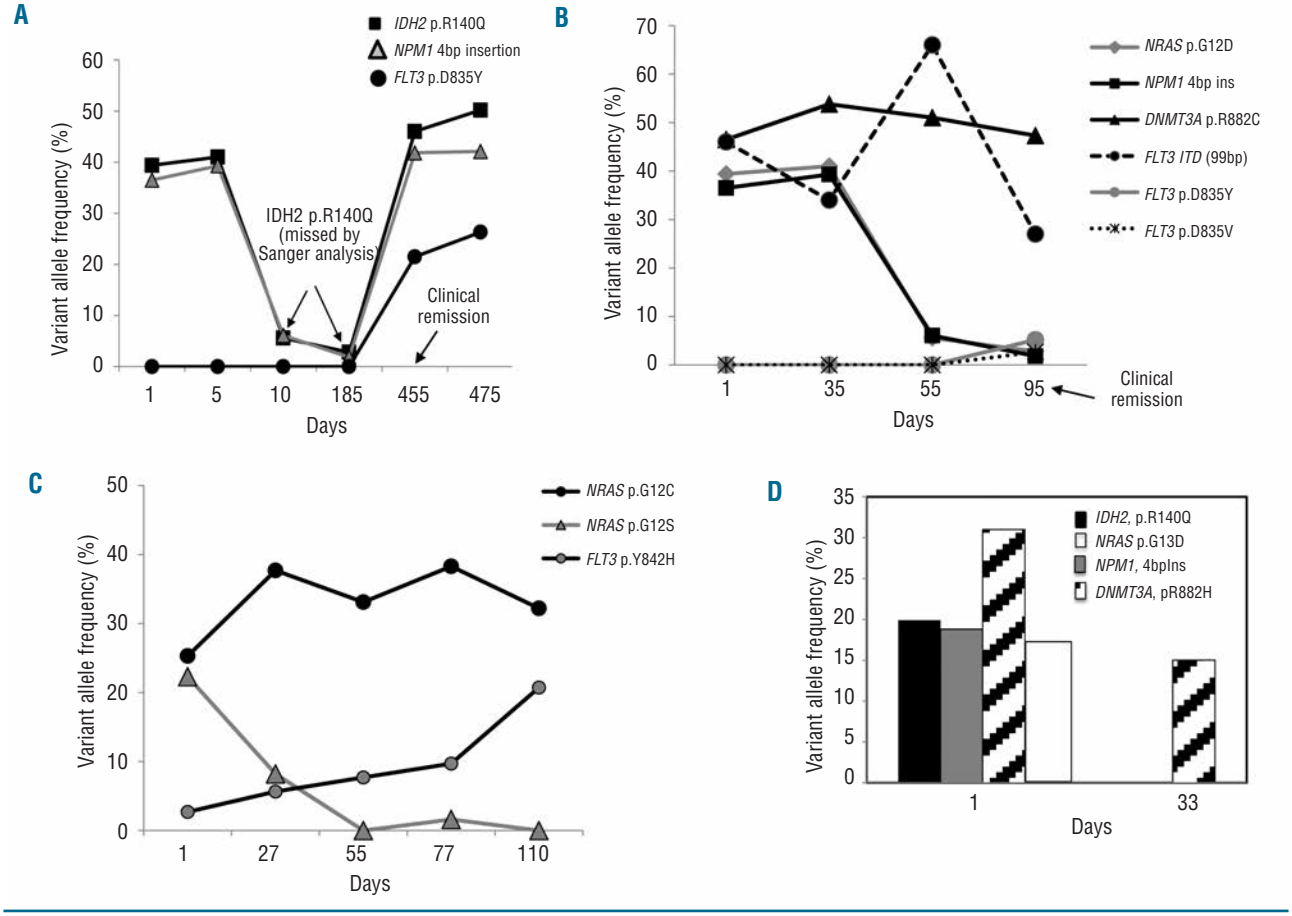

Figure 6. Application for disease monitoring. (A) *IDH2* and *NPM1* mutations detected in a patient at the onset of the disease were missed subsequently when present at low levels (days 10 and 185) when monitored by Sanger sequencing but were effectively detected and called by MiSeq. The time point of clinical remission is indicated by the arrow (B) Mutations in *DNMT3A* and *NRAS*, and insertions in *NPM1* and *FLT3* were detected at diagnosis and followed subsequently. Two low-lying *FLT3* D835 mutations appeared at day 95 indicating the need to follow multiple mutations. The *FLT3* ITD percentages as detected by capillary electrophoresis are plotted. The time point of clinical remission is indicated by the arrow (C) Two *NRAS* mutations were detected, one of which disappeared with treatment indicating distinct clones. A *FLT3* Y842H activating mutation, not routinely tested for was also detected by multigene screening on MiSeq (D) Mutations were detected in four genes at diagnosis warranting testing of each of them after treatment. The mutiplexing capacity of MiSeq would help in consolidating them on to a single platform. Data presented in panels (C) and (D) were from patients with refractory AML.

allows for the screening of multiple genes over the course of therapy thereby maximizing the chances of detecting any other emergent clone that would be missed using classical single-gene screening approaches.

A summary of all somatic mutations detected by MiSeq in our study is provided in *Online Supplementary Figure S4*. It is interesting to note that 23% of the mutations were found in genes which are not routinely tested in AML but are important emerging markers with implications in disease outcome. For instance, *JAK2* p.V617F mutations have been found to occur at low levels in AML with an aberrant karyotype, being more prevalent in AML patients with a preceding myeloproliferative neoplasm,[36] and are associated with aggressive disease and recurrence.[37] Similarly, mutations in *PTPN11*, which encodes a non-receptor tyrosine phosphatase SHP-2, occur at low levels in a subset of AML in both adults and children.[38,39] Furthermore, loss of one copy of the *TP53* gene and mutation of the other has been reported in 78% of AML patients with a complex karyotype,[40] whereas the prevalence in AML cases with a normal karyotype is very low (~2%).[41] It is also noteworthy that two cases in our study harbored the *FLT3* Y842H mutation, which is implicated in resistance to targeted therapy but is not yet routinely tested for in AML. Three of the genes added to the TSACP (*XPO1*, *KLHL6* and *MYD88*) are genes which have been found to be mutated in chronic lymphocytic leukemia; however it is of interest to note that application of this test for routine screening of 875 cases of AML/myelodysplastic syndrome in our laboratory identified three mutations in *XPO1* (p. E571K) and *KLHL6* (p. A93G and p.A9G), indicating that these mutations also occur in AML/myelodysplastic syndrome, albeit at lower frequencies. On the whole, the above mentioned examples demonstrate the advantages of higher sensitivity and simultaneous screening of multiple genes by NGS in identifying clonal heterogeneity and additional genomic aberrations in markers not routinely screened for in AML.

It was not feasible to test the mutation detection ability of this assay for each of the 54 genes because of the lack of positive samples from patients or cell lines. However, in our set of 63 AML/myelodysplastic syndrome samples, 100 single nucleotide variants and 33 insertions were detected by this platform and confirmed by alternate platforms (IT-PGM, capillary electrophoresis and Sanger sequencing) indicating complete concordance and high specificity of mutation detection. We also detected a total of 73 somatic mutations in seven genes routinely tested in AML/myelodysplastic syndrome (*Online Supplementary Figure S4*). Furthermore, we established the overall performance of the test (specificity, sensitivity and reproducibility) for routine clinical use. Our experience has shown that the average sample preparation using TSACP takes less than 9 h (including library preparation, concentration normalization, sample multiplexing and loading on MiSeq), with only 4 h of hands-on time and provides high multiplexing capacity of up to 12 samples (V1) or 24 samples (V2) per run. All subsequent sequencing and analysis steps involving cluster generation, sequencing, base-calling, sequence alignment and variant calling are performed on board and take 28 h (27 h for sequencing and 1 h for analysis) making MiSeq a very self-sufficient sequencer. An average turnaround time of 72 h, starting from DNA extraction to completion of sequencing, can be achieved if batching of samples is not required and is very desirable in medium and high-volume laboratories to offset costs. Furthermore, the convenience and efficiency of screening several routinely tested and AML- associated genes using only 250 ng of DNA and reporting the data simultaneously in a single clinical report enable efficient use of patients' samples and resources in a molecular diagnostic laboratory. On the whole the overall simplicity of sample preparation, the self-sufficiency of the MiSeq sequencer and the associated high specificity, sensitivity and accuracy make it a highly applicable next-generation sequencer for routine mutation screening of AML samples in a clinical molecular diagnostics laboratory.

Authorship and Disclosures

Information on authorship, contributions, and financial & other disclosures was provided by the authors and is available with the online version of this article at www.haematologica.org.

References

1. Owen C, Barnett M, Fitzgibbon J. Familial myelodysplasia and acute myeloid leukaemia a review. Br J Haematol. 2008;140(2):123-32.
2. Sanders MA, Valk PJ. The evolving molecular genetic landscape in acute myeloid leukaemia. Curr Opin Hematol. 2013;20(2):79-85.
3. Ley TJ, Ding L, Walter MJ, McLellan MD, Lamprecht T, Larson DE, et al. DNMT3A mutations in acute myeloid leukemia. N Engl J Med. 2010;363(25):2424-33.
4. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature. 2008;456(7218):66-72.
5. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-10.
6. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, et al. The origin and evolution of mutations in acute myeloid leukemia. Cell. 2012;150(2):264-78.
7. Swederlow SH, Campo E, Harris NLE. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues, International Agency for Research on Cancer Press, Lyon, France. 2008.
8. Grossmann V, Schnittger S, Kohlmann A, Eder C, Roller A, Dicker F, et al. A novel hierarchical prognostic model of AML solely based on molecular mutations. Blood. 2012;120(15):2963-72.
9. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med. 2012;366(12):1079-89.
10. Cancer Genome Atlas Research Network. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-74.
11. Estey EH. Acute myeloid leukemia: 2013 update on risk-stratification and management. Am J Hematol. 2013;88(4):318-27.
12. Burnett A, Wetzler M, Lowenberg B. Therapeutic advances in acute myeloid leukemia. J Clin Oncol. 2011;29(5):487-94.
13. Mardis ER. Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet. 2008;9:387-402.
14. Mardis ER. The impact of next-generation sequencing technology on genetics. Trends Genet. 2008;24(3):133-41.
15. Metzker ML. Sequencing technologies - the next generation. Nat Rev Genet. 2010;11(1):31-46.
16. Loyo M, Li RJ, Bettegowda C, Pickering CR, Frederick MJ, Myers JN, et al. Lessons learned from next-generation sequencing in head and neck cancer. Head Neck. 2013;35(3):454-63.
17. Mardis ER. Applying next-generation sequencing to pancreatic cancer treatment. Nat Rev Gastroenterol Hepatol. 2012;9(8):477-86.
18. Ramsay AJ, Martinez-Trillos A, Jares P, Rodriguez D, Kwarciak A, Quesada V. Next-generation sequencing reveals the secrets of

the chronic lymphocytic leukemia genome. Clin Transl Oncol. 2013;15(1):3-8.
19. Toffanin S, Cornella H, Harrington A, Llovet JM. Next-generation sequencing: path for driver discovery in hepatocellular carcinoma. Gastroenterology. 2012;143(5):1391-3.
20. Ku CS, Cooper DN, Roukos DH. Clinical relevance of cancer genome sequencing. World J Gastroenterol. 2013;19(13):2011-8.
21. Harismendy O, Schwab RB, Bao L, Olson J, Rozenzhak S, Kotsopoulos SK, et al. Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing. Genome Biol. 2011;12(12): R124.
22. Wang C, Krishnakumar S, Wilhelmy J, Babrzadeh F, Stepanyan L, Su LF, et al. High-throughput, high-fidelity HLA genotyping with deep sequencing. Proc Natl Acad Sci USA. 2012;109(22):8676-81.
23. Eyre DW, Golubchik T, Gordon NC, Bowden R, Piazza P, Batty EM, et al. A pilot study of rapid benchtop sequencing of Staphylococcus aureus and Clostridium difficile for outbreak detection and surveillance. BMJ Open. 2012;2(3). pii: e001124.
24. Liu L, Li Y, Li S, Hu N, He Y, Pong R, et al. Comparison of next-generation sequencing systems. J Biomed Biotechnol. 2012;2012: 251364.
25. Loman NJ, Misra RV, Dallman TJ, Constantinidou C, Gharbia SE, Wain J, et al. Performance comparison of benchtop high-throughput sequencing platforms. Nat Biotechnol. 2012;30(5):434-9.
26. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics. 2012;13:341.
27. Junemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31(4): 294-6.
28. Singh RR, Bains A, Patel KP, Rahimi H, Barkoh BA, Paladugu A, et al. Detection of high-frequency and novel DNMT3A mutations in acute myeloid leukemia by high-resolution melting curve analysis. J Mol Diagn. 2012;14(4):336-45.
29. Verma S, Greaves WO, Ravandi F, Reddy N, Bueso-Ramos CE, O'Brien S, et al. Rapid detection and quantitation of BRAF mutations in hairy cell leukemia using a sensitive pyrosequencing assay. Am J Clin Pathol. 2012;138(1):153-6.
30. Deqin M, Chen Z, Nero C, Patel KP, Daoud EM, Cheng H, et al. Somatic deletions of the polyA tract in the 3' untranslated region of epidermal growth factor receptor are common in microsatellite instability-high endometrial and colorectal carcinomas. Arch Pathol Lab Med. 2012;136(5):510-6.
31. Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29(1):24-6.
32. Routbort MJ, Patel KP, Singh RR, Aldape K, Reddy NG, Barkoh BA, et al. OncoSeek - a versatile annotation and reporting system for next generation sequencing-based clinical mutation analysis of cancer specimens. Abstract - Association For Molecular Pathology, Annual Conference. J Mol Diagn. 2012; 14(6).
33. 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.
34. 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.
35. Bagrintseva K, Schwab R, Kohl TM, Schnittger S, Eichenlaub S, Ellwart JW, et al. Mutations in the tyrosine kinase domain of FLT3 define a new molecular mechanism of acquired drug resistance to PTK inhibitors in FLT3-ITD-transformed hematopoietic cells. Blood. 2004;103(6):2266-75.
36. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, et al. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. Cancer Cell. 2005;7(4):387-97.
37. Illmer T, Schaich M, Ehninger G, Thiede C. Tyrosine kinase mutations of JAK2 are rare events in AML but influence prognosis of patients with CBF-leukemias. Haematologica. 2007;92(1):137-8.
38. Hou HA, Chou WC, Lin LI, Chen CY, Tang JL, Tseng MH, et al. Characterization of acute myeloid leukemia with PTPN11 mutation: the mutation is closely associated with NPM1 mutation but inversely related to FLT3/ITD. Leukemia. 2008;22(5):1075-8.
39. Tartaglia M, Martinelli S, Iavarone I, Cazzaniga G, Spinelli M, Giarin E, et al. Somatic PTPN11 mutations in childhood acute myeloid leukaemia. Br J Haematol. 2005;129(3):333-9.
40. Rucker FG, Schlenk RF, Bullinger L, Kayser S, Teleanu V, Kett H, et al. TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. Blood. 2012;119(9):2114-21.
41. Fernandez-Mercado M, Yip BH, Pellagatti A, Davies C, Larrayoz MJ, Kondo T, et al. Mutation patterns of 16 genes in primary and secondary acute myeloid leukemia (AML) with normal cytogenetics. PLoS One. 2012;7(8):e42334.