The Journal of Molecular Diagnostics, Vol. 15, No. 1, January 2013

the Journal of Molecular Diagnostics
jmd.amjpathol.org

See related Commentary on page 27

Detection of *FLT3* Internal Tandem Duplication in Targeted, Short-Read-Length, Next-Generation Sequencing Data

David H. Spencer,* Haley J. Abel,† Christina M. Lockwood,* Jacqueline E. Payton,* Philippe Szankasi,‡ Todd W. Kelley,§ Shashikant Kulkarni,* John D. Pfeifer,*¶ and Eric J. Duncavage*¶

From the Divisions of Laboratory and Genomic Medicine and Molecular and Anatomic Pathology,¶ Department of Pathology and Immunology, and the Division of Statistical Genomics,† Department of Genetics and Center for Genome Sciences & Systems Biology, Washington University School of Medicine, St. Louis, Missouri; ARUP Laboratories,‡ Salt Lake City, Utah; and the Department of Pathology,§ University of Utah College of Medicine, Salt Lake City, Utah*

CME Accreditation Statement: This activity ("JMD 2013 CME Program in Molecular Diagnostics") has been planned and implemented in accordance with the Essential Areas and policies of the Accreditation Council for Continuing Medical Education (ACCME) through the joint sponsorship of the American Society for Clinical Pathology (ASCP) and the American Society for Investigative Pathology (ASIP). ASCP is accredited by the ACCME to provide continuing medical education for physicians.

The ASCP designates this journal-based CME activity ("JMD 2013 CME Program in Molecular Diagnostics") for a maximum of 48 AMA PRA Category 1 Credit(s)™. Physicians should only claim credit commensurate with the extent of their participation in the activity.

CME Disclosures: The authors of this article and the planning committee members and staff have no relevant financial relationships with commercial interests to disclose.

Accepted for publication August 24, 2012.

Address correspondence to Eric J. Duncavage, M.D., Department of Pathology and Immunology, Division of Anatomic and Molecular Pathology, Division of Laboratory and Genomic Medicine, 660 Euclid Ave, #8118, St. Louis, MO. E-mail: eduncavage@path.wustl.edu.

A recurrent somatic mutation frequently found in cytogenetically normal acute myeloid leukemia (AML) is internal tandem duplication (ITD) in the fms-related tyrosine kinase 3 gene (*FLT3*). This mutation is generally detected in the clinical laboratory by PCR and electrophoresis-based product sizing. As the number of clinically relevant somatic mutations in AML increases, it becomes increasingly attractive to incorporate *FLT3* ITD testing into multiplex assays for many somatic mutations simultaneously, using next-generation sequencing (NGS). However, the performance of most NGS analysis tools for identifying medium-size insertions such as *FLT3* ITD mutations is largely unknown. We used a multigene, targeted NGS assay to obtain deep sequence coverage (>1000-fold) of *FLT3* and 26 other genes from 22 *FLT3* ITD-positive and 29 ITD-negative specimens to examine the performance of several commonly used NGS analysis tools for identifying *FLT3* ITD mutations. ITD mutations were present in hybridization-capture sequencing data, and Pindel was the only tool out of the seven tested that reliably detected these insertions. Pindel had 100% sensitivity (95% CI = 83% to 100%) and 100% specificity (95% CI = 88% to 100%) in our samples; Pindel provided accurate ITD insertion sizes and was able to detect ITD alleles present at estimated frequencies as low as 1%. These data demonstrate that *FLT3* ITDs can be reliably detected in panel-based, next-generation sequencing assays. *(J Mol Diagn 2013, 15: 81—93; http://dx.doi.org/10.1016/j.jmoldx.2012.08.001)*

Recurrent somatic mutations play an important role in diagnosis, treatment, and prognosis in acute myeloid leukemia (AML). These mutations include not only the cytogenetic aberrations that define current classification schemes in AML, but also gene-level mutations that are valuable prognostic markers and help to guide complex treatment decisions. Such prognostic markers are especially important in cytogenetically normal AML, in which clinical outcomes are heterogeneous but risk stratification of patients based on prognosis is challenging.[1] Among the most common of these gene mutations are internal tandem duplication mutations, ranging in size from 15 bp to approximately 300 bp, in the juxtamembrane domain region of the fms-related tyrosine kinase 3 gene (*FLT3*); these mutations occur in approximately 20% to 30% of AML

Supported by funding from the Washington University Department of Pathology.

http://dx.doi.org/10.1016/j.jmoldx.2012.08.001

patients and have been associated with increased relapse risk and decreased overall survival in patients with cytogenetically normal AML.[2–4] As a result, more aggressive treatment is often considered in AML patients with a normal karyotype and *FLT3* ITD mutations.[5] Consequently, testing for these mutations is frequently part of the initial diagnostic workup for all patients with a new diagnosis of AML and a normal karyotype.

Recent genomic studies have demonstrated that recurrent mutations in several other genes may be informative prognostic markers in AML. For example, mutations in the *NPM1*, *CEBPA*, *WT1*, *KIT*, *DNMT3A*, *IDH1*, *IDH2, TET2, ASXL1*, *RUNX1*, *MLL*, and *NRAS* genes have all been found to occur in a significant fraction of AML patients, and their presence may be associated with differences in outcome.[5–13] Although focal testing of exons and known hotspots for mutations in some of these genes is currently available, the growing number of genes and the diverse spectrum of mutation types observed in AML make it expensive and impractical to conduct more extensive testing with current methods. In contrast, targeted next-generation sequencing (NGS) can provide comprehensive, unbiased mutational profiling of many genes simultaneously with a single method and for relatively little cost.[14] Nonetheless, a major obstacle to implementing these approaches is uncertainty about the performance of these methods in identifying the full spectrum of mutation types seen in cancer from NGS data.[15] Although NGS-based approaches are theoretically capable of identifying all types of mutations, the most common NGS analysis software tools have been rigorously evaluated only for their accuracy in detecting single-nucleotide substitutions. Methods for detecting insertions, deletions, and larger structural mutations either do not meet the rigorous standards required for a clinical diagnostic test or have not been systematically tested. In particular, the ability of any algorithm to reliably detect medium-size insertions (15 to 300 bp), such as *FLT3* ITDs, is largely unknown.

Because detecting *FLT3* ITD mutations is critical for any multiplex molecular testing strategy in AML, we sequenced *FLT3* and 26 other cancer-associated genes from ITD-positive and ITD-negative cancer specimens to determine how well *FLT3* ITD mutations can be identified in targeted, multigene NGS data. Our objectives were, first, to determine whether targeted sequencing via hybridization and capture with Illumina sequencing (Illumina, San Diego, CA) is an effective method for sequencing *FLT3* ITD insertion sequences and then also to find a robust bioinformatic tool for detecting these mutations that can be incorporated into the analysis pipeline of a comprehensive NGS-based test for somatic mutations in AML. Our results show that NGS can indeed be used to detect *FLT3* ITD mutations, and that directed application of specific bioinformatic tools to the *FLT3* ITD locus can result in an accurate alternative to standard PCR and capillary electrophoresis in the context of a multigene panel for mutational profiling in AML.

Materials and Methods

Samples

The DNA samples included in the present study were derived from a total of 51 unique specimens: 9 peripheral blood or bone marrow specimens from AML patients submitted to ARUP Laboratories (Salt Lake City, UT) for *FLT3* ITD testing, 15 blood or bone marrow specimens from AML patients from the Barnes-Jewish Hospital Molecular Diagnostics Laboratory (BJH MDL; St. Louis, MO), 12 fresh-frozen and 13 formalin-fixed lung adenocarcinoma specimens from the Siteman Cancer Center, Siteman Cancer Center Tissue Procurement Core (St. Louis, MO), and 2 cell lines. All human samples were deidentified and were approved for use in the present study by the Institutional Review Boards for Barnes-Jewish Hospital (IRB 7275) and Siteman Cancer Center at Washington University (IRB 2011-02311).

DNA Preparation

DNA was extracted from whole blood and bone marrow specimens using either a QIAcube instrument (Qiagen, Valencia, CA), according to the manufacturer's recommendations, or a standard manual extraction protocol (Gentra Puregene; Qiagen). DNA from fresh-frozen and formalin-fixed, paraffin-embedded specimens was extracted using QIAamp DNeasy blood and tissue DNA extraction kits (Qiagen). Purity and concentration of the extracted DNA were measured using a NanoDrop 1000 spectrophotometer (Thermo Scientific, Wilmington, DE) and a Qubit fluorometer (Life Technologies, Carlsbad, CA) for all samples; DNA from formalin-fixed, paraffin-embedded specimens was assessed for size integrity using a PCR ladder assay, consisting of 100-, 200-, 300-, and 400-bp products. All specimens met minimum quality and purity requirements of 1 μg in less than 130 μL volume, with optical density $OD_{260/280}$ of 1.7 to 1.9 and $OD_{260/230}$ of >2. All DNA extracted from formalin-fixed, paraffin-embedded specimens demonstrated detectable amplification of fragments of at least 200 bp in size, according to a ladder PCR assay.[16]

FLT3 ITD Testing by PCR and Capillary Electrophoresis

All samples in the present study were tested for *FLT3* ITD mutations by PCR and capillary electrophoresis at ARUP Laboratories, BJH MDL, or the Laboratory for Personalized Molecular Medicine (San Diego, CA). Testing done at ARUP Laboratories or BJH MDL was performed with fluorescently labeled primers flanking the juxtamembrane domain region of *FLT3*, followed by capillary electrophoresis on an ABI 3730XL capillary sequencer (Life Technologies). Capillary data were analyzed using ABI GeneMapper software version 3.7 (Life Technologies); cases with peaks larger than the expected product size were called positive, with an ITD insertion of the size indicated by the software. The ITD allele fraction was determined using methods described by Zwann

et al[17] and by Meshinchi et al.[18] Briefly, ITD allele fraction was measured by dividing the ITD peak area by the sum of the peak areas for the ITD and wild-type peaks. This calculation was performed for all peaks that were identified; in cases with multiple traces from the same specimen, the mean allele fraction was calculated from all traces. To confirm linearity of the calculated ITD allelic ratio, a standard curve was constructed based on PCR and capillary electrophoresis results of serial dilutions of the ITD-containing MV4-11 cell line with *FLT3* wild-type DNA (Supplemental Figure S1).

Targeted Next-Generation Sequencing of *FLT3*

FLT3 sequencing was performed via targeted next-generation sequencing on a HiSeq 2000 sequencing system (Illumina), using a next-generation sequencing-based panel (WUCaMP27; Washington University Genomic Pathology Services, St. Louis, MO) for detecting somatic mutations in *FLT3* and in 26 other genes that are frequently mutated in cancer (Supplemental Table S1). Whole-genome sequencing libraries were prepared from 1 μg of input DNA fragmented to 200 to 250 bp using a Covaris E210 sonicator (Covaris, Woburn, MA). Fragmentation was verified on an Agilent 2100 bioanalyzer (Agilent Technologies, Santa Clara, CA), and then the fragmented DNA was purified with Agencourt AMPure XP beads (Beckman Coulter Genomics, Danvers, MA), end-repaired and A-tailed with Klenow DNA polymerase, and ligated to universal Illumina adapters. Library fragments were then bead-purified and analyzed for adequate ligation on an Agilent 2100 bioanalyzer (Agilent Technologies).

Limited cycle PCR with sample-specific, index-tagged primers was then performed to enrich for ligation products with the appropriate configuration (ie, ligation of one of each of the adapters on either end). Whole-genome libraries were enriched for exons plus 200 bp of flanking intronic sequence and 1 kbp flanking the first and last exon of the 27 genes targeted by the WUCaMP27 panel, including *FLT3*, using a custom SureSelect biotinylated cRNA probe set (Agilent Technologies). SureSelect reagents were prepared according to the manufacturer's instructions, and 500 ng of each indexed library was hybridized at 65°C for 24 hours. Captured library fragments were washed and purified from unbound material using Dynabeads MyOne streptavidin T1 beads (Life Technologies), and then were resuspended and purified by bead purification before a final limited-cycle PCR amplification.[19] Verification of library size and quantity was performed by electrophoresis using a bioanalyzer (Agilent Technologies). Enriched libraries were pooled (28 to 30 indexed libraries per pool) and were sequenced in multiplex on a HiSeq system (Illumina) using version 3 chemistry according to established protocols for paired-end 101-bp reads.

Bioinformatic Analysis

Base calls made using Casava software version 1.8 (Illumina) were converted from SCARF to FASTQ format. Paired-end reads were aligned to the reference genome (hg19, NCBI build GRCh37) using either Novoalign (Novocraft, Selangor, Malaysia) or the Burrows-Wheeler alignment tool (BWA) with default parameters.[20,21] Analysis of the resulting sorted BAM files for the presence of *FLT3* ITD mutations was performed using the following software tools: SAMtools with BCFtools (*http://samtools.sourceforge.net*), Dindel (*http://www.sanger.ac.uk/resources/software/dindel*), Genome Analysis Toolkit (GATK; (*http://www.broadinstitute.org/gatk*), C), Maq (*http://maq.sourceforge.net/maq-man.shtml*), CLC bio Genomics Workbench (deletion-insertion polymorphism detection and structural variation tools; CLC bio, Muehltal, Germany), SLOPE (*http://www-genepi.med.utah.edu/suppl/SLOPE/index.html*), BreakDancer (*http://breakdancer.sourceforge.net*), and Pindel (*https://trac.nbic.nl/pindel*), all of which were applied to the *FLT3* juxtamembrane domain region (hg 19, NCBI build GRCh37; chr13:28608000-28608600) using default parameters.[19,22–26] For Pindel, an additional filtering step was applied to remove indel variants with fewer than five supporting reads. Except for CLC bio Genomics Workbench, the software packages used for analysis were publicly available.

In addition, *de novo* fragment assembly was performed on read pairs mapped to the juxtamembrane domain region using a custom Perl script executing the Phrap assembly program (*http://www.phrap.org/phredphrapconsed.html*). Briefly, all reads mapped to the juxtamembrane region, including both reads from one-end-anchored read pairs, were extracted from BAM files and converted into FASTA format with associated decimal quality scores. The assembler Phrap then assembled the reads into contigs for local alignment with the reference subsequence chr13:28608000-28608600 using the BLAT sequence alignment program.[27] Any gap present in the resulting alignments with a mean coverage in the assembled contig of 10 or more was considered an ITD insertion and was parsed from the alignment file and converted into a final variant calling format (VCF) file. For Pindel, the allele fraction was calculated by dividing the number of supporting reads by the coverage in unique reads; for *de novo* assembly, the mean coverage depth of the assembled contig across the inserted sequence was divided by total coverage of *FLT3* exons 14 and 15 with unique reads.

Results

Targeted Multigene Next-Generation Sequencing

We sequenced *FLT3* along with 26 other cancer genes as part of a targeted, multigene test for detecting somatic mutations in AML and other malignancies using solution-phase hybridization and capture coupled with Illumina sequencing (WUCaMP27 assay). This combination of technologies was chosen because it has been shown to provide relatively unbiased enrichment for target sequences and robust detection of a broad spectrum of mutation types, including insertions, translocations, and large duplications.[14,28,29] A general overview of our sequencing method is shown in Figure 1. First,

whole-genome libraries were generated, followed by enrichment for all exons and up to 1000 bp of flanking noncoding sequence for each gene in the panel via solution-phase hybridization of the library fragments to biotinylated cRNA probes. Enriched libraries were indexed with unique index tags and sequenced in multiplex on an Illumina HiSeq system to obtain 101-bp, paired-end sequencing reads. The WUCaMP27 target size of 406 kbp allowed for multiplexing of ~30 cases per flow cell lane, to obtain an expected coverage of approximately 1000-fold.

We used the WUCaMP27 panel to sequence a total of 51 unique samples: 24 peripheral blood or bone marrow specimens from AML patients, 25 formalin-fixed or fresh-frozen solid tumor specimens, and 2 cancer cell lines (M4V-11 and GIST882).[30,31] Of these, 19 of the AML specimens and the MV4-11 cell line had been tested previously by PCR and capillary electrophoresis and found to have *FLT3* ITD insertions ranging from 17 to 185 bp. The *FLT3* ITD status of these cases was independently verified via PCR and capillary electrophoresis-based testing, as were all other samples analyzed in the present study (Table 1). Multiplex sequencing across three HiSeq flow cell lanes produced between 4.6 and 29.6 million mapped reads per case, of which 2.9 to 14.5 million mapped to the 406-kbp target region. This corresponds to an enrichment efficiency of 46% to 87% and an expected coverage of 739-fold to 3903-fold of exons in the target region for each case. The sequencing results for *FLT3* ITD positive and negative cases are summarized in Supplemental Table S2.

Sequencing of *FLT3* Genes with Internal Tandem Duplications

We first analyzed coverage statistics for *FLT3* exons 14 and 15 in all sequenced samples, to verify that our enrichment method provides adequate sequence coverage at this locus and that ITD insertions do not adversely affect enrichment, sequencing, or read mapping. The mean coverage of *FLT3* exons 14 and 15 did not differ significantly from the mean coverage for all exons across all sequenced samples (*FLT3* exons 14 and 15, 2328; all exons, 2055; $P = 0.11$ Student's *t*-test) or when only ITD-positive samples were considered (*FLT3* exons 14 and 15, 2021; all exons, 1652; $P = 0.08$ Student's *t*-test). In addition, coverage patterns for *FLT3* exons 13 to 15 were similar between ITD-positive and ITD-negative cases (Figure 2A). This demonstrates that a targeted-capture approach provides adequate coverage of *FLT3* and that ITD insertions do not grossly affect target enrichment, sequencing, or read-mapping ability.

Despite minimal overall effects on coverage of the *FLT3* juxtamembrane region using the assay, there was clear evidence of ITD insertions in the read alignments at this locus in ITD-positive cases. We observed two signals that

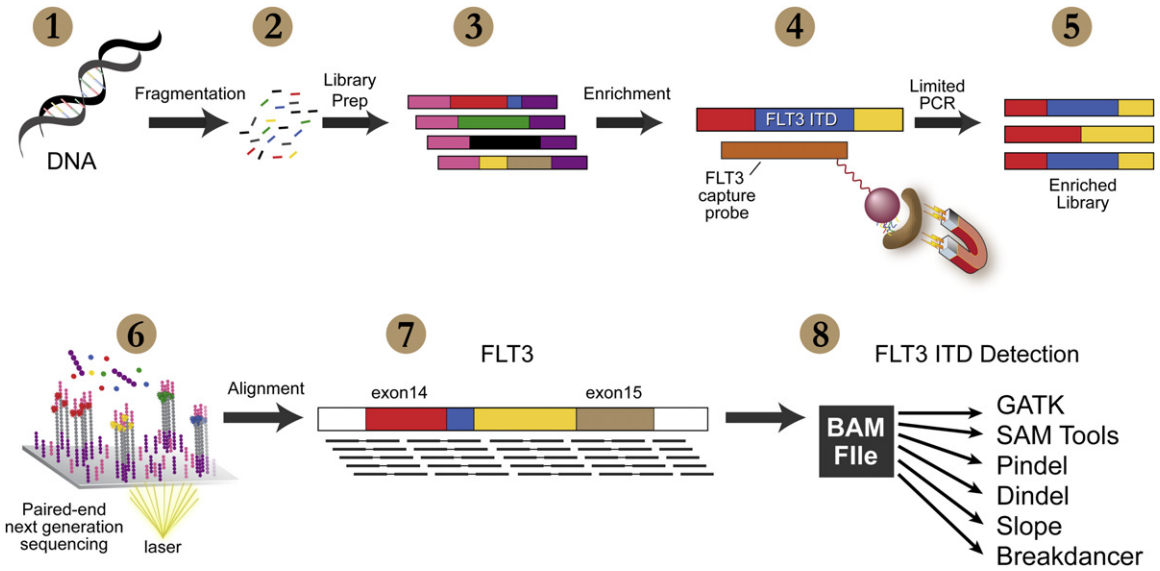

Figure 1 Experimental overview. **1:** Genomic DNA derived from bone marrow aspirates, fresh-frozen tumor tissue, or formalin-fixed paraffin-embedded tissue was extracted. **2:** The extracted genomic DNA was fragmented on a Covaris E210 instrument, producing a mean insert size of 270 bp. **3:** Indexed sequencing libraries were then prepared by linker ligation. *FLT3* Genomic DNA sequences are indicated in red, blue (corresponding to the ITD sequence), yellow, and brown; non-*FLT3* genomic DNA is indicated in black and green; the sequencing library adaptors are indicated in pink and purple. **4:** Sequencing libraries were enriched for genes of interest using a custom SureSelect panel (Agilent Technologies) consisting of 27 genes, including all exons of *FLT3*, and covering 406 kb in total (WUCaMP27). In this example, genomic DNA containing *FLT3* sequence (red, blue, and yellow) is captured by specific biotinylated probes (orange). The captured sequence is then separated from nontargeted DNA by the addition of streptavidin-coated paramagnetic beads. **5:** Enriched DNA was then eluted from capture beads and subjected to low-cycle amplification using primers targeting the adaptor linkers (linkers not shown). **6:** The resulting DNA was then sequenced in multiplex with 20 to 30 cases per lane on a HiSeq 2000 sequencing system (Illumina). **7:** Sequence data were then mapped to the reference genome (hg19, NCBI build GRCh37) using both BWA and Novoalign tools. In this example, aligned paired-end sequencing reads are shown in black. **8:** *FLT3* ITDs were detected from the mapped data using a variety of publicly available tools, including GATK, SAMtools, Pindel, Dindel, SLOPE, and BreakDancer. The results were then compared with conventional PCR-based findings for *FLT3* ITD detection.

Table 1 Summary of NGS versus PCR/Capillary Electrophoresis for *FLT3* ITD Detection

FLT3 ITD					
CE	NGS*	Cases (no.)	Mean blast percentage†	*FLT3* ITD size range (bp)	Mean *FLT3* ITD allelic ratio‡
Positive	Positive	20	52	17–185§	0.218§
Positive	Negative	0	NA	NA	NA
Negative	Positive	2	34	27–39¶	0.02¶
Negative	Negative	29	NA	NA	NA

*Positive by NGS was defined as detection of an ITD using either the Pindel tool or *de novo* assembly.
†Mean blast cell percentages reflect only data set 2 (cases 2-1 to 2-12). Percentages were not available from deidentified material (data set 1: cases 1-1 to 1-10).
‡The allelic ratio was calculated by dividing the area of the *FLT3* ITD peak by the total area of both the *FLT3* ITD peak and the wild-type allele peak.
§As reported by PCR followed by CE.
¶As reported by the Pindel tool on NGS data.
CE, capillary electrophoresis; ITD, internal tandem duplication; NA, not applicable; NGS, next-generation sequencing.

suggested the presence of insertions. First, there was an excess of one-end-anchored reads, in which one read in a read pair was successfully mapped to the *FLT3* region but the other was unmapped (Figure 2A). This suggested that a significant number of reads at this locus harbor novel sequence not present in the reference genome. Next, inspection of read alignments in the region revealed many reads with soft-clipped bases, which occur when reads are only partially aligned with the reference. These can easily be identified in the read alignments, because they have a high density of discrepancies in the unaligned portion (Figure 2B). Both phenomena demonstrate the presence of molecules containing the ITD insertion sequence in the enriched library, indicating that these sequences are successfully captured using a targeted NGS approach. These observations suggest that detection of ITD mutations should be possible with the appropriate bioinformatic methods.

Evaluation of Methods for Detecting *FLT3* ITD Insertions

With evidence that ITDs were present in NGS data, we next tested several widely available NGS software packages with insertion/deletion detection functionality to determine their ability to identify *FLT3* ITD insertions in a subset of the sequenced samples. The software tools we evaluated apply different methods for detecting potential insertions, but most (including the commonly used SAMtools, Maq, Dindel, and GATK) rely exclusively on indels initially identified during mapping and alignment of individual sequencing reads, followed by scoring of indel calls via probabilistic models (Figure 3). Because these methods are designed for detecting short insertions of 1 to 15 bp, we also tested SLOPE and BreakDancer, which identify larger insertions and deletions from chimeric reads or read pairs with discordant insert sizes, and Pindel, which uses a pattern-growth algorithm to detect a variety of indel variants using read pairs in which one member is partially or completely unaligned with the reference sequence.[19,23] We initially tested 10 ITD-positive cases (9 cases from AML patients and the *FLT3* ITD-positive M4V-11 cell line) with SAMtools, Maq, Dindel, GATK, CLC bio Genomics Workbench, SLOPE, BreakDancer, and Pindel. Default parameters were used throughout, and analysis was restricted to *FLT3* exons 14 and 15 (hg19, NCBI build GRCh37; chr13:28,608,000-28,608,600), because we were concerned only with detection of *FLT3* ITD insertions.

The results achieved with each software tool are summarized in Table 2. Only SLOPE and Pindel detected any insertions in these ITD-positive cases. SLOPE detected insertions in 8 of the 10 cases (80%; 95% CI = 44% to 97%), and Pindel detected insertions in all 10 cases (100%; 95% CI = 69% to 100%). The two cases in which SLOPE failed to detect insertions (cases 1-1 and 1-7) were shown to have multiple ITD alleles by PCR and capillary electrophoresis (Table 2), suggesting that the software program may have difficulty resolving sequencing reads that contain multiple insertions. In contrast, Pindel successfully detected insertions in these complex cases; for two of the three cases (1-6 and 1-7), the program predicted insert sizes that were similar to results obtained from PCR (Table 2). The ITD sizes by PCR for the 10 cases ranged from 21 to 93 bp and had estimated allele fractions between 5% and 50%, suggesting that Pindel can detect ITD insertions over a range of these parameters. In addition to ITD insertions, Pindel also occasionally reported insertions (and deletions) of less than 5 bp, which is smaller than what is typically seen in *FLT3* ITD insertions. These variants were supported by only a few reads; inspection of the aligned reads at these positions suggested that the calls were false positives. In subsequent analysis, therefore, we added a filtering step to remove indel calls with fewer than five supporting reads.

Although the Pindel results were generally compatible with PCR and capillary electrophoresis, in two instances the pattern of insertions did not agree between methods. In case 1-6, Pindel identified three insertions (72, 27, and 21 bp), whereas only two insertions (69 and 21 bp) were reported from PCR and capillary electrophoresis data (Table 2). Close inspection of the electrophoretic traces for this case revealed a small peak consistent with a 27-bp insertion, which supports the Pindel result with three ITD alleles in the sample (Figure 4). In case 1-4, the other discordant case, only one insertion was identified by Pindel, whereas two

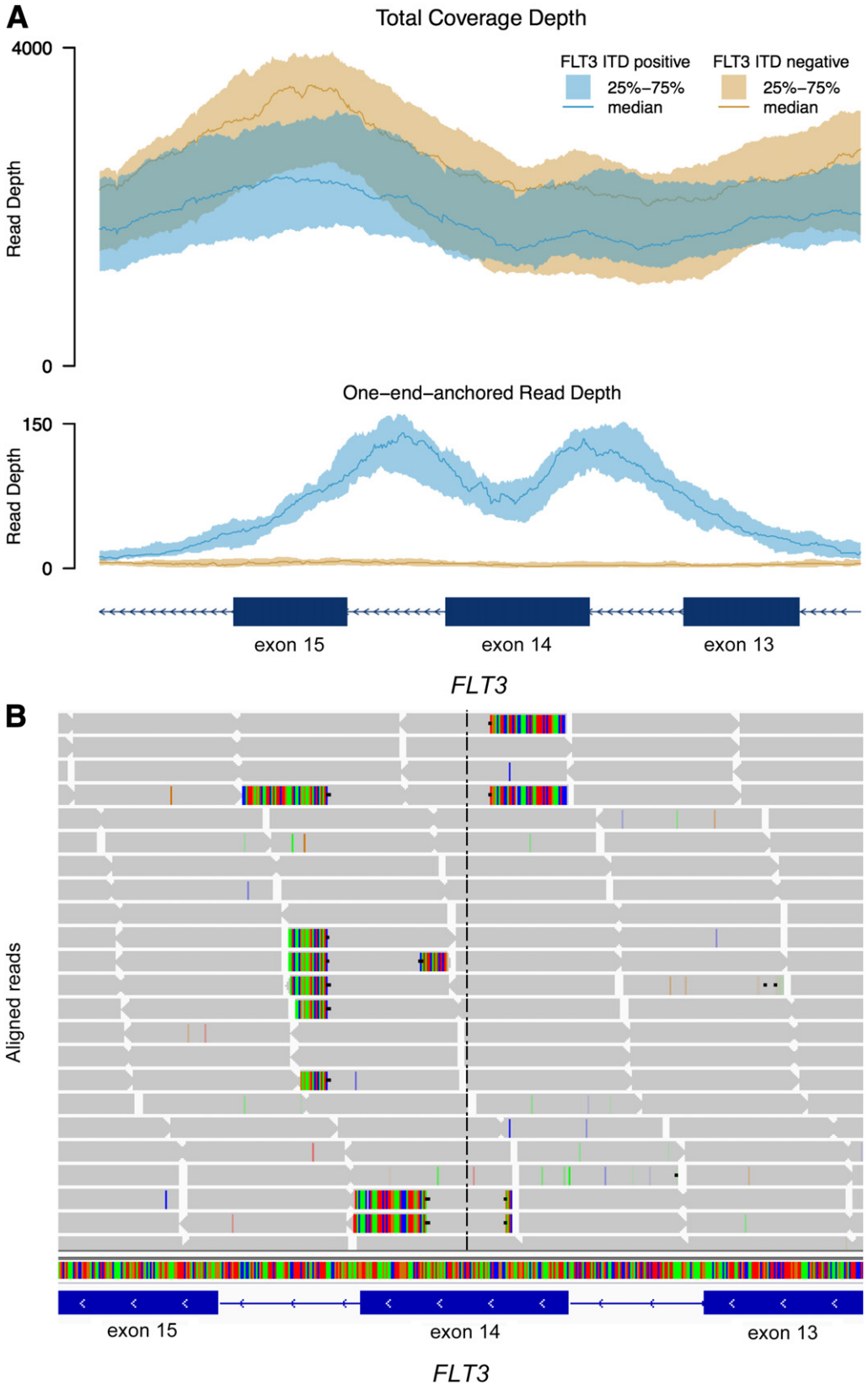

Figure 2 Coverage data and read alignments at the *FLT3* locus from NGS data from samples with and without ITD insertion. **A**: Shown are total median and interquartile range of raw coverage depth values for *FLT3* ITD-positive and *FLT3* ITD-negative cases (read depth 0—4000), along with depth in one-end-anchored reads (read depth 0—150), in which one read of a paired read did not map. The presence of an insertion results in an excess of one-end-anchored reads in the *FLT3* ITD-positive cases. **B**: A subset of aligned reads in case 1-8 with an *FLT3* ITD insertion. Aligned reads are shown in gray; multicolored bars in the reads indicate discrepancies with the reference sequence. Blocks of discrepancies in a subset of the reads are the result of the ITD insertion.

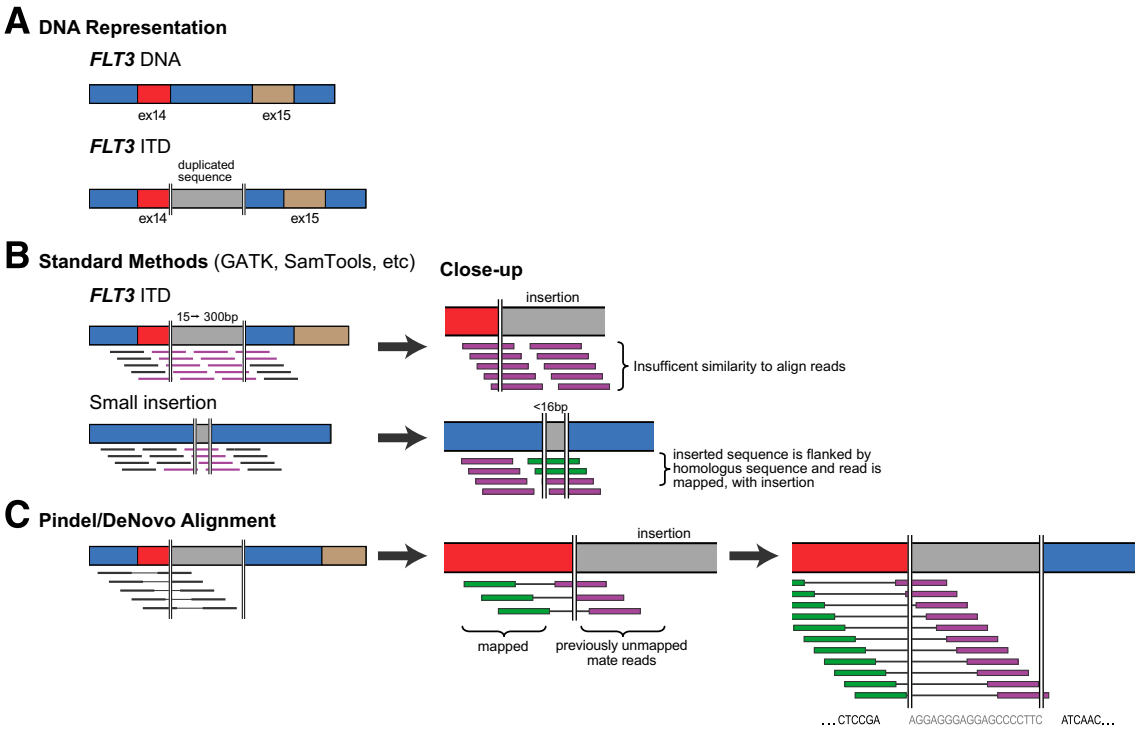

Figure 3 Informatics of insertion detection. **A:** The *FLT3* ITD is an insertion of duplicated and nonduplicated sequence that occurs between exons 13 and 14. These insertions (shown in gray) range in size from 15 to ~300 bp. **B:** Standard methods for finding insertions, including the Genome Analysis Toolkit (GATK), SAMtools, and Dindel, apply probabilistic models to make insertion calls based on data obtained during the initial read mapping and alignment process. Because of the difficulty associated with aligning short reads, only small insertion events (generally <15% of the total read length) can be identified by this approach (aligned reads are shown in green; unaligned reads, in purple). Such reads generally have sufficient homology in the regions flanking the insertion to permit accurate alignment. Large insertions (>16 bp), including the *FLT3* ITD, are too long to be detected by this method. **C:** Using a paired-end approach, software such as Pindel and *de novo* alignment can reliably detect larger insertions, including the *FLT3* ITD. In this approach, mate-pairs are identified in which one end is mapped, but the other is not. The unmapped mates are then assembled to form contigs with partial homology to the reference sequence, using a pattern-growth algorithm (Pindel) or *de novo* assembly with a custom script (unpublished data) executing Phrap assembly software. This method allows for the detection of much larger insertions.

ITD alleles were identified by PCR (Table 2). The missed 84-bp insertion was present at a relatively high frequency, according to PCR and capillary electrophoresis data. Given that other insertions with similar size and allele fraction were detected by Pindel, why this insertion was missed is unclear.

Given the performance of Pindel in this set of samples, we were interested in exploring other methods that could also take advantage of partially aligned and unmapped reads to detect *FLT3* ITD insertions. To this end, we used the publicly available Phrap software version 1.090518 (and a custom script; unpublished data) to assemble all reads at the *FLT3* ITD locus *de novo*, including partially aligned and one-end-anchored reads, and then identified insertions in the resulting contigs from alignments with *FLT3* ITD locus from the human reference sequence. We called cases ITD-positive if any gap was identified in these alignments with mean coverage of >10 and used the gap size as the ITD allele size (as described under *Materials and Methods*). This approach was applied to the same 10 cases with known ITD mutations that were analyzed (see above), and it successfully detected insertions in all 10 cases, including the three with multiple ITD alleles (Table 2). The patterns of insertion mutations were generally concordant with PCR-based results, and in most cases insertion sizes differed from PCR by only a few base pairs. In addition, *de novo* assembly detected the 27-bp insertion in case 1-6 that was initially missed by PCR but identified by Pindel, providing additional support for this finding. Although assembly showed good overall concordance with the other two methods, in case 1-1, assembly identified an extra 27-bp insertion, and in a second case (1-7) the insertion sizes differed by ~40 bp and ~170 bp from both PCR and Pindel (Table 2), as discussed below. Despite these few discordant results, assembly still resulted in 100% sensitivity (95% CI = 69% to 100%) for ITD-positive cases, which provides additional support for methods that take into account unmapped and one-end-anchored reads in the detection of *FLT3* ITD insertions.

Performance Characteristics of Pindel and *de Novo* Assembly Assessed on Additional Cases

We next used the three best-performing tools from the first data set (Pindel, *de novo* assembly, and SLOPE) to test an additional 15 AML specimens and 26 fresh-frozen and

Table 2 *FLT3* ITD Detection by Different NGS Software Tools

Case	PCR/CE	Pindel	Assembly	SAMtools	GATK	SLOPE*	Dindel†	Maq†	CLC bio†	BreakDancer†
Data set 1										
1-1	**18**	**21**	**21, 27**	ND	ND	ND	ND	ND	ND	ND
1-2	**42**	**45**	**45**	ND	ND	**POS**	ND	ND	ND	ND
1-3	**93**	**96**	**96**	ND	ND	**POS**	ND	ND	ND	ND
1-4	**81, 51**	**54**	**84, 54**	ND	ND	**POS**	ND	ND	ND	ND
1-5	**36**	**39**	**39**	ND	ND	**POS**	ND	ND	ND	ND
1-6	**69, 21**	**72, 21, 27**	**72, 30, 27**	ND	ND	**POS**	ND	ND	ND	ND
1-7	**87, 66**	**84, 69**	**258, 28**	ND	ND	ND	ND	ND	ND	ND
1-8	**57**	**54**	**54**	ND	ND	**POS**	ND	ND	ND	ND
1-9	**17**	**18**	**18**	ND	ND	**POS**	ND	ND	ND	ND
1-10	**30**	**30**	**30**	ND	ND	**POS**	ND	ND	ND	ND
Data set 2†										
2-1	**60**	**63**	**126**	ND	ND	ND				
2-2	**72**	**72**	**72**	ND	ND	**POS**				
2-3	**70**	**85**	**87**	ND	ND	ND				
2-4	**60, 100**	**63, 102**	**63, 102**	ND	ND	ND				
2-5	**57**	**60**	**60**	ND	ND	**POS**				
2-6	**49**	**33, 51**	**33, 51**	ND	ND	ND				
2-7	**97**	**95**	**99**	ND	ND	ND				
2-8	**37, 63, 185**	**39, 66, 186**	**66, 186**	ND	ND	ND				
2-9	**57**	**60**	**60**	ND	ND	ND				
2-10	**37**	**39**	**39**	ND	ND	ND				
2-11‡	ND	**27**	ND	ND	ND	ND				
2-12‡	ND	**39**	**39**	ND	ND	**POS**				

Cases in which an *FLT3* ITD was detected are highlighted in bold. Numeric values express ITD insertion size in base pairs (bp).

*The SLOPE tool does not directly report the insertion size or sequence. Cases that showed evidence of an insertion event were called positive.

†Given the uniformly poor performance of Dindel, MAQ, CLC bio Genomic Workbench, and BreakDancer in the first set of cases, these tools were not tested on the second set of cases. As both SAMTools and the GATK are commonly used general purpose analysis tools, they were run in all cases.

‡Cases 2-11 and 2-12 were considered low-positive. *FLT3* ITD was not initially identified in the Barnes-Jewish Hospital Molecular Diagnostics Laboratory (BJH MDL) by PCR-based methods; however, these cases were sent to a second laboratory (ARUP Laboratories), where the findings of low-positive *FLT3* ITD were confirmed by PCR.

ND, not detected; POS, positive.

formalin-fixed solid tumor specimens for *FLT3* ITD mutations. The AML specimens in this set included 10 cases previously found to be ITD-positive by PCR, although our analysis was performed masked to *FLT3* ITD status. Both Pindel and *de novo* assembly correctly identified insertions in all 10 of the additional *FLT3* ITD-positive AML cases in data set 2, and detected multiple insertions in two cases with more than one allele identified by PCR (Table 2). There were no false-positive *FLT3* ITDs identified by Pindel or *de novo* assembly in 26 ITD-negative solid tumor specimens (Supplemental Table S3), although there were two potential false positives, two cases in which Pindel detected ITD insertions but that had been called ITD-negative by conventional methods. These cases were subsequently found to harbor low-level ITD alleles (discussed below). These data suggest that both Pindel and *de novo* assembly have a high specificity when applied to the *FLT3* ITD region. Although SLOPE performed well in data set 1, it identified only 2/10 (20%) ITDs in data set 2. The reason for this low sensitivity is unclear, but may be related to the lower overall depth of coverage obtained in data set 2.

Taken together with the results from our initial analysis, Pindel and *de novo* assembly identified similar ITD allele sizes (defined as alleles with a size difference of <15 bp). Pindel identified 25 of 26 ITD alleles identified by PCR and capillary electrophoresis (96%; 95% CI = 80% to 100%) and *de novo* assembly identified 22 of the 26 (85%; 95% CI = 65% to 96%); both tools correctly identified ITD insertions in five cases with multiple alleles. Using the PCR results as the gold standard, the overall sensitivity for identifying an ITD-positive specimen (ie, detection of any insertion in an ITD-positive case) was 100% (20/20; 95% CI = 83% to 100%) for both Pindel and *de novo* assembly, and each had 100% specificity (0 ITDs detected in 29 confirmed ITD-negative specimens; 95% CI = 88% to 100%). This level of accuracy matches our own experience validating single-nucleotide variant (SNV) detection for clinical mutation detection (data not shown) and is similar to published reports on the accuracy of SNV detection using standard tools,[26,32]

Discrepancies Between Conventional and NGS-Based Methods

Although the majority of NGS-based ITD calls agreed with PCR and capillary electrophoresis, a few results were

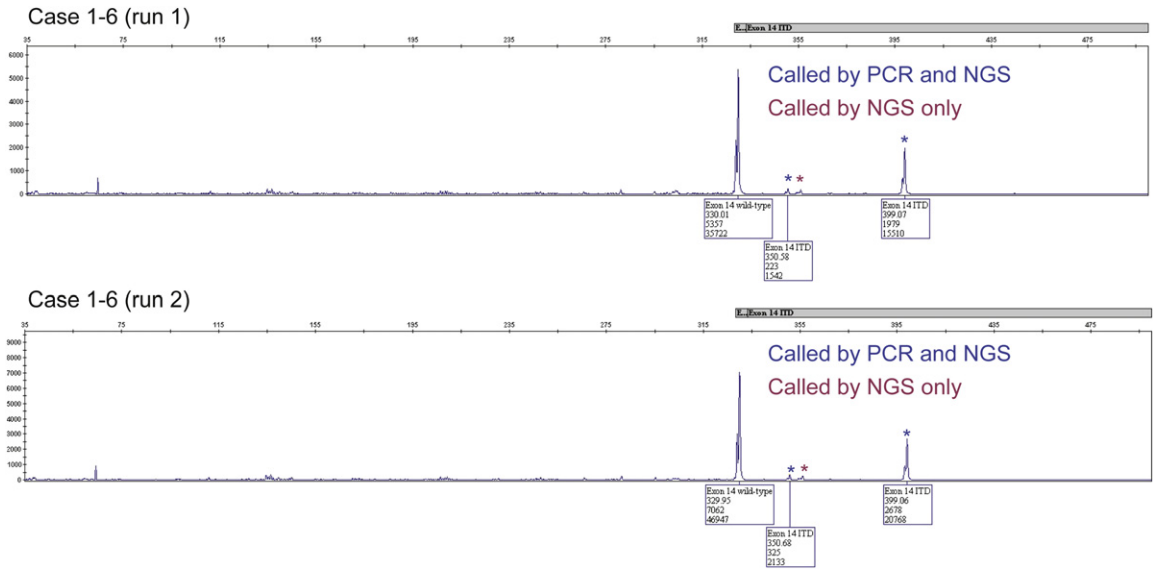

Figure 4 Example of *FLT3* ITD called by NGS but not PCR. In case 1-6, both Pindel and *de novo* alignment identified a third *FLT3* ITD (27 bp) from NGS data (**red asterisk**), corresponding to an insertion of 5′-TCTCTGAAATCAACGTAGAAGTACTCA-3′. Retrospective analysis of the capillary traces (performed at ARUP Laboratories) confirmed the presence of a small peak (~27 bp) in replicate PCR-based testing called by PCR and NGS (**blue asterisk**).

discordant. We considered differences in ITD allele size and number (described above) to be minor discrepancies, because these would be unlikely to affect patient management. However, there were two potential false positive cases, in which AML samples previously determined to be ITD-negative by PCR and capillary electrophoresis were called positive by NGS. In one instance (case 2-11), both Pindel and *de novo* assembly detected a 39-bp insertion; in the other (case 2-12), Pindel, but not *de novo* assembly, detected a 27-bp insertion. In both cases, the supporting reads of the insertion was low: Pindel reported 24 supporting reads for case 2-11 and only 5 supporting reads for case 2-12, compared with a mean coverage of ~2400 for the *FLT3* ITD locus in each case. This suggested that the frequency of these two insertions was low; similar to the cases (described above) with low-level ITD alleles that were missed by PCR, these ITDs could be real but below the limit of detection of PCR-based methods. Indeed, manual review of the capillary electrophoresis data showed that both cases had small peaks corresponding to insertions of the same size as those detected in NGS data. DNA from these two discrepant cases was sent to ARUP Laboratories for repeat testing, using a different laboratory-derived test based on PCR and capillary electrophoresis, which confirmed the presence of low-allele-fraction *FLT3* ITDs corresponding to the ITD size identified by NGS (Supplemental Figure S2). This result provides evidence that that Pindel may be more sensitive than conventional methods when applied to high-coverage NGS data.

FLT3 ITD Size and Allele Fraction by NGS

Because some studies have correlated ITD allele fraction and insertion size with outcome in AML,[33,34] we compared these metrics calculated from the Pindel and *de novo* assembly with those from PCR and capillary electrophoresis, to determine the concordance between the different methods. There was generally good concordance in ITD size between NGS-based methods and PCR (Figure 5A). All but one of the insertions detected by Pindel was within 3 bp of the PCR-based estimate, and 21 of the 26 ITD insertions identified by *de novo* assembly were within 9 bp. However, in two cases the sizes predicted by *de novo* assembly differed appreciably from the PCR-based size estimate. In one case with two ITD alleles (case 1-7), *de novo* assembly identified two insertions that differed by ~40 bp and ~170 bp from the PCR-based size; in another case with a single ITD allele, the size differed by ~60 bp. ITD insertions are of course in tandem (as is implicit in the term itself), and so misalignment or misassembly of duplicated sequence could explain this difference. Indeed, for both of these size-discrepant cases, alignment of the inserted sequence in the assembled contigs with the *FLT3* juxtamembrane region showed that only a portion of the putative insertion aligned, implicating misassembly of reads with tandem duplications as a cause for the discrepancy. Despite these two discordant results using the assembly method, ITD size from both Pindel and assembly were largely concordant with PCR across a range of ITD allele sizes.

We also compared the ITD allele fraction calculated from Pindel and *de novo* assembly to the estimate from peak heights obtained from PCR and capillary electrophoresis. Although NGS data can provide a digital count of the frequency of single nucleotide substitutions and smaller indels contained within individual reads, this calculation is more complicated for the larger *FLT3* ITD insertion mutations, which may span multiple reads. To estimate ITD allele fraction from our data, we used the number of

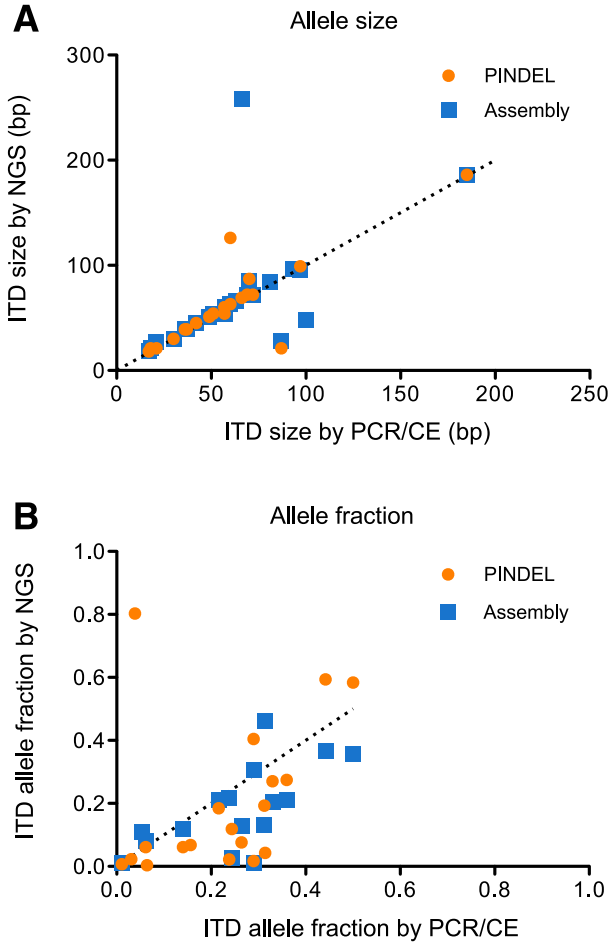

Figure 5 Comparison of ITD size and allele fraction between PCR and capillary electrophoresis and NGS methods for a subset of ITD-positive cases. **A:** Size of the ITD insertion detected from NGS data using Pindel and *de novo* assembly, compared with size from PCR and capillary electrophoresis. **B:** The ITD allele fraction calculated from the areas under the mutant and wild-type peaks from PCR and capillary electrophoresis, compared with Pindel (Pearson's correlation coefficient = 0.37; 95% CI = −0.1 to 0.71; P = 0.11) and assembly (Pearson's correlation coefficient = 0.65; 95% CI = 0.23 to 0.87; P = 0.006). The dashed line indicates a perfect fit ($y = x$), indicating complete agreement between PCR and NGS methods.

supporting reads reported by Pindel and the mean coverage depth in the assembled contigs over the interval that corresponded to the alignment gap for *de novo* assembly; both measures were divided by the depth in unique reads over the *FLT3* exons 14 and 15 region to arrive at an allele fraction estimate. The estimates from both methods had relatively poor correlation with results from PCR and capillary electrophoresis (Figure 5B). The allele fraction estimates from Pindel were statistically uncorrelated with the PCR estimate (Pearson's correlation coefficient = 0.19; 95% CI = −0.1 to 0.71; P = 0.11) (R statistics package; *http://www.r-project.org*), although this was likely due to an outlier case with a 185-bp ITD allele that had many supporting reads because of its large size and resulted in an overestimation of the allele fraction. The estimates from *de novo* assembly had better agreement with PCR-based results (Pearson's correlation coefficient = 0.65; 95% CI = 0.23 to 0.87; P = 0.006) and were less sensitive to ITD allele size, because our allele fraction calculation used mean coverage over the gap length and therefore was normalized to the length of the inserted sequence. These results suggest that ITD allele fraction from NGS may be useful for a qualitative interpretation or relative comparisons in a single case with the same ITD alleles over time, although reporting precise estimates may be problematic.

In addition to determining *FLT3* ITD allele fraction, we compared the ITD allele fraction with blast cell percentage (Supplemental Figure S3). We observed little correlation between the two measures, suggesting presence of the ITD in a subclonal population. We also examined the allele frequencies of all SNVs within the *FLT3* capture region. Most of these SNVs were present at frequencies near 50% or 100%, consistent with constitutional heterozygous or homozygous variation, respectively. A few SNVs were present at other frequencies, suggestive of copy number variation in a clonal population, but this also was independent of the *FLT3* ITD allelic ratio.

Effect of Alignment Method and Read Length on *FLT3* ITD Detection

Given the added expense and sequencing time required to generate the paired-end 101-bp reads used in the present study, we modified our sequencing data to simulate both single-end reads and shorter paired-end reads, to determine how well they perform in detecting *FLT3* ITD insertions. Using single-end reads, we were unable to run Pindel using any read length, because the software requires paired-end reads in which one end (the mapped read) serves as an anchor sequence. We next simulated 42-bp and 76-bp paired-end reads, using data from the second set of 10 cases (data set 2), by truncating the read length; we then remapped the resulting reads, using Novoalign. When Pindel was applied to these simulated data sets, insertions were detected in only 5 of the 10 cases for both the 42-bp and 76-bp paired-end reads, compared with detection of all 10 for the initial data set of 101-bp paired-end reads. Of the five cases in which insertions were detected, most were larger ITD alleles (85, 95, and 186 bp); smaller insertions (<72 bp) were largely missed. There did not appear to be a significant correlation between *FLT3* ITD detection with shorter reads and allelic ratio, arguing that the ability to detect ITD insertions with shorter reads cannot be overcome simply by increasing coverage.

Finally, we sought to determine the effect of the choice of initial alignment software on the ability to detect the *FLT3* ITD insertions. We aligned sequence data from the initial 10 ITD-positive cases, using both Novoalign and BWA with default parameters for paired-end reads, and analyzed the resulting alignments using Pindel as described in *Materials and Methods* (command line arguments are summarized in Supplemental Table S4). For the 10 ITD-positive cases from data set 2 (Table 2), *FLT3* ITD insertions were identified in

all 10 cases from the Novoalign-aligned data and in 8 of 10 cases using the BWA-aligned data. The two cases missed in the BWA-aligned data were relatively large (85 bp, case 2-3; 95 bp, case 2-7); however, the largest ITD in the set (186 bp, case 2-8) was detected in both BWA and Novoalign data. Interestingly, the BWA-aligned data produced far more soft-clipped reads, corresponding to the ITD event, than did Novoalign (Supplemental Figure S4).

Discussion

In the present study, we used hybridization and capture enrichment (WUCaMP27 panel) with Illumina sequencing to sequence *FLT3* and 26 other genes, to determine whether *FLT3* ITD mutations can be accurately identified from targeted, multigene NGS data. We generated approximately 1000-fold coverage of *FLT3* from 51 specimens, including 19 cases and the M4V-11 cell line, known to be ITD-positive, and 29 ITD-negative specimens used as negative controls, and 2 AML cases initially PCR tested as ITD negative, but subsequently shown to have low level ITDs by NGS and more sensitive PCR. We demonstrated that ITD insertions were well represented in these data, and indeed signals of these mutations were clearly evident in NGS read alignments, including an excess of reads with unaligned, soft-clipped bases, and read pairs with one end that could not be mapped to the reference genome. Because these signals serve as indicators of the presence of insertions in NGS data, we believe that they could be applied more broadly to aid in the clinical interpretation of other loci in the genome where recurrent insertion mutations are known to occur.

We tested the ability of several NGS analysis tools to identify *FLT3* ITD insertions and found that only one of the publicly available software packages, Pindel, reliably detected the ITD mutations. Pindel was 100% sensitive for ITD insertions in the 20 ITD-positive specimens in the present study and resulted in no false positives in 29 specimens determined to be ITD-negative by standard PCR methods. Accurate identification of all ITD alleles was achieved in most cases with multiple alleles by PCR, including cases in which the ITD allele fraction was as low as 1%. This excellent sensitivity for low-frequency mutations was also demonstrated by the detection of ITD insertions in two cases that were initially called ITD-negative by conventional methods but for which review of the electrophoretic traces confirmed the presence of ITD alleles. We also showed that alignment of contigs from *de novo* assembly of reads mapped to *FLT3* exons 14 and 15 was equally successful as Pindel at identifying these mutations.

The fact that Pindel and *de novo* assembly identified ITD insertions that were missed by several other indel detection tools in common use can be attributed to differences in the algorithms used. The majority of the available indel detection tools are able to find only insertions that are identified during the read-mapping process that are less than approximately 15 to 20 bp (using 100-bp read lengths) and that are contained entirely within individual read alignments. Most of the other tools are designed to detect large structural mutations from confidently mapped read pairs with discordant insert sizes or orientations that suggest a potential discrepancy with the reference sequence. In contrast, both Pindel and *de novo* assembly include reads that are partially or completely unaligned with the reference. These reads are critical for identifying medium-size insertions, such as those seen in *FLT3* ITD mutations, because this size range is too large to be detected by indels within single reads and often too small for detection by analysis of discordant insert sizes. Although we expect the identification of such mutations to become easier and more reliable as NGS technologies permit longer read lengths, Pindel and similar approaches will likely continue to be useful for detecting insertion mutations or in any other situation in which analysis of unmapped reads can facilitate mutation detection.

Although Pindel and *de novo* assembly performed very well at detecting *FLT3* ITD insertions in the present study, it is important to emphasize that the sensitivity and specificity we achieved cannot be generalized beyond this particular locus. We applied Pindel and assembly to a narrow genomic region, containing only coding sequence where indel mutations are expected to be rare. General application of Pindel to larger regions, and especially to noncoding or repetitive sequences where indels are more likely to occur, will likely result in the identification of a substantial number of indel variants, many of which may be false positives. In some instances, manual review of the sequence context at putative indel variants for microsatellites, other repetitive sequences, and features suggestive of insertion mutations (such as excess soft-clipped and one-end-anchored reads) can aid in interpretation. Nonetheless, the performance of any indel detection software can be assessed only using a set of known indels obtained from orthogonal methods.

Panel-based NGS approaches to molecular diagnostics in cancer have recently gained popularity, because of their ability to sequence large regions for relatively little cost. In theory, such panels offer increased sensitivity to a larger spectrum of mutations, compared with current testing methods, which are often targeted to specific mutational hotspots that may include a single codon or nucleotide position. Nonetheless, there are many challenges to developing NGS assays, including identification and validation of bioinformatic tools for accurate detection of all mutation types seen in cancer. Here, we have shown that accurate detection of critical *FLT3* ITD mutations is possible in the context of deep sequencing with a targeted multigene panel using NGS. These findings are a step forward in the validation of targeted multigene next generation sequencing as a viable option for comprehensive molecular diagnostic testing of leukemia, and of cancer in general.

Acknowledgments

We thank Mark Watson for assistance with tissue bank specimens; Mark Johnson and Paul Cliften for their

assistance in using the high performance computing cluster; Rob Mitra, Dayna Oschwald, Richard Head, and the Genome Technology Access Core Laboratory (GTAC) for providing sample preparation and sequencing services; Karen Seibert, Andy Bredemeyer, and the Washington University Genome Pathology Services (WU-GPS) for providing sequencing services; and Jeffery Klco for his critique of this manuscript.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2012.08.001*.

References

1. Walter MJ, Graubert TA, Dipersio JF, Mardis ER, Wilson RK, Ley TJ: Next-generation sequencing of cancer genomes: back to the future. Per Med 2009, 6:653
2. Nakao M, Yokota S, Iwai T, Kaneko H, Horiike S, Kashima K, Sonoda Y, Fujimoto T, Misawa S: Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia 1996, 10: 1911−1918
3. Abu-Duhier FM, Goodeve AC, Wilson GA, Care RS, Peake IR, Reilly JT: Genomic structure of human FLT3: implications for mutational analysis. Br J Haematol 2001, 113:1076−1077
4. Kottaridis PD, Gale RE, Frew ME, Harrison G, Langabeer SE, Belton AA, Walker H, Wheatley K, Bowen DT, Burnett AK, Goldstone AH, Linch DC: The presence of a FLT3 internal tandem duplication in patients with acute myeloid leukemia (AML) adds important prognostic information to cytogenetic risk group and response to the first cycle of chemotherapy: analysis of 854 patients from the United Kingdom Medical Research Council AML 10 and 12 trials. Blood 2001, 98:1752−1759
5. Estey EH: Acute myeloid leukemia: 2012 update on diagnosis, risk stratification, and management. Am J Hematol 2012, 87:89−99
6. Wouters BJ, Löwenberg B, Erpelinck-Verschueren CA, van Putten WL, Valk PJ, Delwel R: Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood 2009, 113:3088−3091
7. 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, 363:2424−2433
8. Hou HA, Huang TC, Lin LI, Liu CY, Chen CY, Chou WC, Tang JL, Tseng MH, Huang CF, Chiang YC, Lee FY, Liu MC, Yao M, Huang SY, Ko BS, Hsu SC, Wu SJ, Tsay W, Chen YC, Tien HF: WT1 mutation in 470 adult patients with acute myeloid leukemia: stability during disease evolution and implication of its incorporation into a survival scoring system. Blood 2010, 115:5222−5231
9. Mardis ER, Ding L, Dooling DJ, Larson DE, McLellan MD, Chen K, et al: Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009, 361:1058−1066
10. Gaidzik VI, Bullinger L, Schlenk RF, Zimmermann AS, Röck J, Paschka P, Corbacioglu A, Krauter J, Schlegelberger B, Ganser A, Späth D, Kündgen A, Schmidt-Wolf IG, Götze K, Nachbaur D, Pfreundschuh M, Horst HA, Döhner H, Döhner K: RUNX1 mutations in acute myeloid leukemia: results from a comprehensive genetic and clinical analysis from the AML study group. J Clin Oncol 2011, 29: 1364−1372
11. Shen Y, Zhu YM, Fan X, Shi JY, Wang QR, Yan XJ, Gu ZH, Wang YY, Chen B, Jiang CL, Yan H, Chen FF, Chen HM, Chen Z, Jin J, Chen SJ: Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. Blood 2011, 118:5593−5603
12. Döhner K, Tobis K, Ulrich R, Fröhling S, Benner A, Schlenk RF, Döhner H: Prognostic significance of partial tandem duplications of the MLL gene in adult patients 16 to 60 years old with acute myeloid leukemia and normal cytogenetics: a study of the Acute Myeloid Leukemia Study Group Ulm. J Clin Oncol 2002, 20:3254−3261
13. Bacher U, Haferlach T, Schoch C, Kern W, Schnittger S: Implications of NRAS mutations in AML: a study of 2502 patients. Blood 2006, 107:3847−3853
14. Duncavage EJ, Abel HJ, Szankasi P, Kelley TW, Pfeifer JD: Targeted next generation sequencing of clinically significant gene mutations and translocations in leukemia. Mod Pathol 2012, 25:795−804
15. Mardis ER: The $1,000 genome, the $100,000 analysis? Genome Med 2010, 2:84
16. van Dongen JJ, Langerak AW, Brüggemann M, Evans PA, Hummel M, Lavender FL, Delabesse E, Davi F, Schuuring E, García-Sanz R, van Krieken JH, Droese J, González D, Bastard C, White HE, Spaargaren M, González M, Parreira A, Smith JL, Morgan GJ, Kneba M, Macintyre EA: Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia 2003, 17:2257−2317
17. Zwaan CM, Meshinchi S, Radich JP, Veerman AJ, Huismans DR, Munske L, Podleschny M, Hählen K, Pieters R, Zimmermann M, Reinhardt D, Harbott J, Creutzig U, Kaspers GJ, Griesinger F: FLT3 internal tandem duplication in 234 children with acute myeloid leukemia: prognostic significance and relation to cellular drug resistance. Blood 2003, 102:2387−2394
18. Meshinchi S, Alonzo TA, Stirewalt DL, Zwaan M, Zimmerman M, Reinhardt D, Kaspers GJ, Heerema NA, Gerbing R, Lange BJ, Radich JP: Clinical implications of FLT3 mutations in pediatric AML. Blood 2006, 108:3654−3661
19. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R: Dindel: accurate indel calls from short-read data. Genome Res 2011, 21:961−973
20. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009, 25:1754−1760
21. Li H, Durbin R: Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics 2010, 26:589−595
22. Li H: A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. Bioinformatics 2011, 27:2987−2993
23. 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: 2865−2871
24. Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl M, Zhang Q, Locke DP, Shi X, Fulton RS, Ley TJ, Wilson RK, Ding L, Mardis ER: BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. Nat Methods 2009, 6:677−681
25. Abel HJ, Duncavage EJ, Becker N, Armstrong JR, Magrini VJ, Pfeifer JD: SLOPE: a quick and accurate method for locating non-SNP structural variation from targeted next-generation sequence data. Bioinformatics 2010, 26:2684−2688
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010, 20:1297−1303
27. Kent WJ: BLAT−the BLAST-like alignment tool. Genome Res 2002, 12:656−664
28. Welch JS, Westervelt P, Ding L, Larson DE, Klco JM, Kulkarni S, Wallis J, Chen K, Payton JE, Fulton RS, Veizer J, Schmidt H, Vickery TL, Heath S, Watson MA, Tomasson MH, Link DC, Graubert TA, DiPersio JF, Mardis ER, Ley TJ, Wilson RK: Use of

whole-genome sequencing to diagnose a cryptic fusion oncogene. JAMA 2011, 305:1577–1584
29. 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:R18
30. Stong RC, Korsmeyer SJ, Parkin JL, Arthur DC, Kersey JH: Human acute leukemia cell line with the t(4;11) chromosomal rearrangement exhibits B lineage and monocytic characteristics. Blood 1985, 65:21–31
31. Tuveson DA, Willis NA, Jacks T, Griffin JD, Singer S, Fletcher CD, Fletcher JA, Demetri GD: STI571 inactivation of the gastrointestinal stromal tumor c-KIT oncoprotein: biological and clinical implications. Oncogene 2001, 20:5054–5058
32. Pritchard CC, Smith C, Salipante SJ, Lee MK, Thornton AM, Nord AS, Gulden C, Kupfer SS, Swisher EM, Bennett RL, Novetsky AP, Jarvik GP, Olopade OI, Goodfellow PJ, King MC, Tait JF, Walsh T: ColoSeq provides comprehensive Lynch and polyposis syndrome mutational analysis using massively parallel sequencing. J Mol Diagn 2012, 14:357–366
33. Mead AJ, Gale RE, Hills RK, Gupta M, Young BD, Burnett AK, Linch DC: Conflicting data on the prognostic significance of FLT3/TKD mutations in acute myeloid leukemia might be related to the incidence of biallelic disease. Blood 2008, 112:444–445. author reply 445
34. Gale RE, Green C, Allen C, Mead AJ, Burnett AK, Hills RK, Linch DC, Medical Research Council Adult Leukaemia Working Party: The impact of FLT3 internal tandem duplication mutant level, number, size, and interaction with NPM1 mutations in a large cohort of young adult patients with acute myeloid leukemia. Blood 2008, 111: 2776–2784