

Integrated Genomic DNA/RNA Profiling of Hematologic Malignancies in the Clinical Setting

Jie He^{1*}, Omar Abdel-Wahab^{2,3*}, Michelle K. Nahas^{1*}, Kai Wang^{1*}, Raajit K. Rampal^{2,3}, Andrew M. Intlekofer^{4,5}, Jay Patel³, Andrei Krivstov^{4,6,11}, Garrett M. Frampton¹, Lauren E Young¹, Shan Zhong¹, Mark Bailey¹, Jared R. White¹, Steven Roels¹, Jason Deffenbaugh¹, Alex Fichtenholtz¹, Timothy Brennan¹, Mark Rosenzweig¹, Kimberly Pelak¹, Kristina M. Knapp⁶, Kristina W. Brennan¹, Amy L Donahue¹, Geneva Young¹, Lazaro Garcia¹, Selmira T. Beckstrom¹, Mandy Zhao¹, Emily White¹, Vera Banning¹, Jamie Buell¹, Kiel Iwanik¹, Jeffrey S Ross¹, Deborah Morosini¹, Anas Younes⁵, Alan M. Hanash⁷, Elisabeth Paietta⁸, Kathryn Roberts⁹, Charles Mullighan⁹, Ahmet Dogan¹⁰, Scott A. Armstrong^{4,6,11}, Tariq Mughal^{1,13}, Jo-Anne Vergilio¹, Elaine Labrecque¹, Rachel Erlich¹, Christine Vietz¹, Roman Yelensky¹, Philip J. Stephens¹, Vincent A. Miller¹, Marcel R. M. van den Brink^{7,12}, Geoff A. Otto^{1†}, Doron Lipson^{1†}, and Ross L. Levine^{2,3,6†}

* contributed equally

¹ Foundation Medicine, Cambridge, MA 02141

² Leukemia Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065

³ Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065

⁴ Cancer Biology and Genetics Program, Sloan Kettering Institute New York, NY 10065

⁵ Lymphoma Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065

⁶ Leukemia Center, Memorial Sloan Kettering Cancer Center, New York, NY 10065

⁷ Bone Marrow Transplant Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065

⁸ Department of Medicine (Oncology), Albert Einstein College of Medicine, Yeshiva University, New York, NY

⁹ Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN

¹⁰ Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY 10065

¹¹ Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, NY 10065

¹² Division of Hematologic Oncology, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065

¹³ Division of Hematology and Oncology, Tufts University Medical Center, Boston, MA 02111

† Corresponding authors: Geoff Otto gotto@foundationmedicine.com, Doron Lipson dlipson@foundationmedicine.com, and Ross L. Levine leviner@mskcc.org

Key Points

- Novel clinically-available comprehensive genomic profiling of both DNA and RNA in hematologic malignancies
- Profiling of 3696 clinical hematologic tumors identified somatic alterations that impact diagnosis, prognosis, and therapeutic selection

Abstract

The spectrum of somatic alterations in hematologic malignancies includes substitutions, insertions/deletions (indels), copy number alterations (CNAs) and a wide range of gene fusions; no current clinically available single assay captures the different types of alterations. We developed a novel next-generation sequencing-based assay to identify all classes of genomic alterations using archived formalin-fixed paraffin-embedded (FFPE), blood and bone marrow samples with high accuracy in a clinically relevant timeframe, which is performed in our CLIA-certified CAP-accredited laboratory. Targeted capture of DNA/RNA and next-generation sequencing reliably identifies substitutions, indels, CNAs and gene fusions, with similar accuracy to lower-throughput assays which focus on specific genes and types of genomic alterations. Profiling of 3696 samples identified recurrent somatic alterations that impact diagnosis, prognosis and therapy selection. This comprehensive genomic profiling approach has proved effective in detecting all types of genomic alterations, including fusion transcripts, which increases the ability to identify clinically-relevant genomic alterations with therapeutic relevance.

Introduction

In the past decade, our understanding of the somatic cancer genome has been greatly advanced through gene discovery studies¹⁻⁶. These studies delineated the genomic complexity in different types of human cancer, in different patients with the same tumor type, and within an individual's tumor. These efforts have identified recurrent somatic alterations that in some cases are specific to a particular tumor type but in many cases are shared across different malignancies. These discoveries have had clinical impact in the diagnosis of specific malignancies defined by recurrent somatic alterations³ and in the development of more precise prognostic schema¹. Most importantly these studies have identified disease alleles which have guided the use of molecularly targeted therapy². These data underscore the importance of genomic profiling in clinical oncology and led to the development of DNA-based genomic tests for cancer patients⁷.

Cytogenetic studies identified recurrent chromosomal translocations in a spectrum of hematologic malignancies, including acute myeloid leukemia (AML)⁸, acute lymphoblastic leukemia (ALL)⁹, chronic myeloid leukemia (CML)¹⁰, and non-Hodgkin lymphomas (NHL)¹¹, which impact clinical outcome and can guide therapeutic decisions¹². More recent studies have identified recurrent mutations and amplifications/deletions in a spectrum of hematologic malignancies, including AML¹³, ALL, multiple myeloma (MM), and lymphoma, creating a pressing need to develop comprehensive genomic assays to identify somatic alterations in hematologic malignancies.

Hematologic malignancies have a high frequency of rearrangements that lead to aberrant expression of oncogenes or to the expression of fusion transcripts that contribute to malignant transformation and maintenance. These include classical chromosomal translocations, inversions/duplications, interstitial deletions, and episomal fusions/amplifications that can give rise to disordered expression of full-length genes or of fusion transcripts. In many cases these discoveries have led to the use of targeted therapies in specific disease subsets. However, current diagnostic assays, including FISH and real-time PCR, are designed *ad hoc* to identify specific genomic alterations, and in some cases there are no assays that can reliably identify specific rearrangements. We sought to develop an integrated DNA/RNA profiling platform using targeted next-generation sequencing that could reliably identify single-nucleotide substitutions, insertions/deletions, copy number alterations, and rearrangements.

Methods

Description of workflow. DNA and RNA from each patient are extracted and made into barcoded libraries through separate workflow streams. The DNA and cDNA undergo library construction and hybrid selection on independent plates. DNA and RNA samples from the same patient then converge in an analysis pipeline using the plate names and shared specimen ID. Detailed protocols for DNA and RNA extraction, cDNA synthesis, library construction and hybrid selection are described in Supplementary Information.

Sequencing. The selected libraries are pooled and sequenced on the Illumina HiSeq2500 to ~500x unique coverage for DNA and to >3M unique on-target pairs for cDNA.

Base substitutions, indels and copy number analysis. Samples with median exon coverage in the range 150 – 250X are considered qualified while those with coverage less than 150x are considered failed. For base substitutions, the mutant allele frequency (MAF) cutoff is 1% for

known somatic variants (based on COSMIC v62) and 5% for novel somatic variants. For indels, the MAF cutoff is 3% for known somatic variants and 10% for novel somatic variants. Additional details of the methods were described previously⁷.

Rearrangement calling methods.

Customized alignment workflow was developed for fusion detection from RNA-seq. Raw sequence reads were aligned to whole transcriptome (refSeq) first, and reads with suboptimal mapping were aligned to whole genome references. Alignments to the two different references were then merged and calibrated based on the full genome reference (hg19) for fusion detection.

Gene rearrangements were detected by identifying clusters of chimeric read pairs from both DNA (pairs mapping over 10Kbp apart, or on different chromosomes) and RNA (pairs mapping to refseq sequences corresponding to different genes, or to genomic loci over 10kbp apart). Chimera clusters were filtered for repetitive sequence (average mapq>30) and by distribution of mapped positions (stdev>10). Identified rearrangements were then annotated according to the genomic loci of both clusters and categorized as gene fusions (e.g. *BCR-ABL1*), gene rearrangements (e.g. *IGH-BCL2*) or truncating events (e.g. *TP53* rearrangement). Rearrangement candidates were then filtered based on number of chimera reads supporting the rearrangement events (for documented fusions minimum 10 chimera reads required, for putative somatic driver rearrangements 50 chimera reads required).

In addition to the de-novo rearrangement detection method described above, reads were also separately aligned to a custom reference library generated based on common fusions and rearrangements. Fusions were detected based on the observation of reads aligned across the junction of rearrangement breakpoints. This method includes the detection of 6 common isoforms of *MLL-PTD*¹⁴ (Supplementary Table 3a).

Immunoglobulin heavy locus (*IGH*) rearrangements were detected by targeting rearrangement hotspots of both common immunoglobulin fusion partner genes (major and minor translocations involving *MYC*, *BCL2* and *CCND1*) as well as *IGH* breakpoint regions.

Expression Measurement. We first calculated read counts per million (ppm) for each gene, and then relative expression level was normalized by the median ppm of all 449 multiple myeloma patients. The statistical significance of expression between the rearrangement positive cohort and rearrangement negative cohort was determined by one-sided Wilcoxon rank-sum test.

Results

NGS-based Test for Detecting Genomic Alterations in DNA and RNA

The assay we have developed employs next-generation DNA and RNA sequencing, and builds upon the workflow that has been optimized for genomic profiling of DNA from patients with solid tumors⁷. Genomic profiling is accomplished by integration of data from targeted DNA and RNA profiling with up-to-date interpretation of the clinical significance, thereby achieving increased breadth and improved sensitivity to identify rearrangements which result in aberrant expression of fusion transcripts. The genes targeted by DNaseq and RNAseq are listed in Supplementary Table 1a-b, and the intervals of each exons/introns targeted by this assay are listed in Supplementary Table 1c-d. The clinical implications of targeted genes in ALL, AML, MDS/MPN, NHL and MM are annotated in Supplementary Table 1e. Uniform coverage across all genes are observed from a total of 108 Hapmap and hematologic tumor specimens. The distribution of coverage of 405 genes targeted in DNaseq is shown in Figure 1B and coverage of each gene is listed in Supplementary Table 1f. The distribution of exonic coverage (1244 exons) of all clinical relevant genes is shown in Supplementary Figure 1. Eight exons had consistent low coverage < 100x and were excluded from further analysis (Supplementary Table 1a). The read depth of each gene targeted in RNAseq from 61 validation samples is listed in Supplementary Table 1g. The current workflow is shown in (Figure 1A) and described in the Online Methods. The coverage of each target of genes involved in therapy, clinical trial and/or other prognostic and standard of care is summarized in Supplementary Table 1h.

Detection of DNA Sequence Variants and Copy Number Alterations (CNAs)

We first established analytic accuracy of detecting substitutions, shorts insertions and deletions (indels) and CNAs by comparing the performance of the new assay with a DNA-only assay that had previously undergone comprehensive validation across a large number of clinical samples. Compared with the previously validated assay, the new assay contained an additional 90 genes relevant to hematologic malignancies. Substitution, indel and copy-number detection were validated by reanalyzing 47 samples previously profiled with a validated test⁷ in which 169 alterations were identified in 55 genes common to both assays (102 substitutions, 59 indels and 8 copy-number alterations; 10 low-frequency subclonal variants were excluded from the analysis). The concordance between the two sets of results was 99.4%: 168/169 variants were concordant with a single discordant low frequency variant (Supplementary Table 2a). Additionally, MAF values measured using the two assays were highly correlated ($R^2 = 0.929$; Figure 2A). These data demonstrate that the current DNA profiling platform can accurately detect DNA sequence variants and CNAs, as established for the reference assay.

Detection of Genomic Rearrangements

To determine our ability to detect genomic rearrangements, we pooled 21 cell lines (Figure 3A, Supplemental table2b) with 28 known genomic rearrangements to create 39 mixtures with ratios of 10%, 20%, 25%, 33%, and 50% representing a range of clinically relevant tumor fractions. Sensitivity and specificity of the assay were determined by comparing the genomic rearrangements detected in the pooled samples relative to the constituent cell-lines. The sensitivity for fusion detection at 20-100% tumor fraction was 100% (161/161), and 98% at 10% tumor fraction (84/86) (Supplementary Table 2b). Very few false-positive calls were observed, with a positive predictive value of >98% (245/248, with 2/3 false calls being at a marginal level of detection). The combined DNA and RNA sequencing approach accurately detects a wide variety of genomic rearrangements and gene fusions. RNA sequencing enables efficient

detection of rearrangements and targets all 28 rearrangements with novel breakpoints in a large number of genes, whereas DNA sequencing allows high sensitivity for well characterized rearrangements affecting specific introns and rearrangements that do not result in expression of a fusion transcript, including those involving promoter regions. The distribution of rearrangements called by both DNA and RNA, RNA only and DNA only are shown in Figure 2B. RNA sequencing maintains high sensitivity in high tumor content (100% sensitivity in tumor content 25-100%); at low tumor fraction (10% or 20%), DNA sequencing is more sensitive in selected introns, including for example breakpoints involving the *MLL* and *PDGFRA* genes.

Concordance with Reference Platforms

We performed blinded comparisons with CLIA-certified diagnostic assays, including Sequenom, RT-PCR, FISH, and PCR fragment analysis, for 76 clinical specimens (Supplementary Table 2c) previously tested for 214 clinical relevant alterations in 11 genes that are known and routinely tested in clinical practice in AML, ALL and myelodysplastic syndromes (MDS) (165 short variants and 49 gene fusions, see Figures 2C-D). DNA and RNA were extracted from new unstained sections from the originally tested FFPE block. Coverage of DNA sequencing across 76 specimens is shown in Supplementary Figure 2 and Supplementary Table 2d. Of the 101 genomic alterations identified by Sequenom or gel-sizing, 100 were also called by our new assay (one positive *FLT3* ITD was not detected due to low mutant allele frequency). Of 113 negative genomic alterations reported previously, our assay confirmed 111 negative calls and identified two *IDH2* R140Q mutations that were not detected by the reference method likely due to the greater sensitivity of this approach than the orthogonal platform (Figure 2C, Supplementary Table 2e). These two *IDH2* alterations were later confirmed from an independent validation using AmpliSeq platform. Overall concordance was 99% (211/214) (Figure 2C). The concordance between the NGS-based test and reference testing, including positive and negative calls, is shown in Figure 2D. In addition to the concordance analysis, genomic profiling of the 76 test samples identified 126 additional somatic alterations which are not covered by available hot spot assays in the given disease type, including clinically relevant genomic alterations in *KRAS*, *TET2*, *EZH2*, and *DNMT3A*. (Supplementary Table 2e, Supplementary Figure 3).

A subset of samples with low frequency calls (<10%) were selected for an independent validation using either AmpliSeq assay (Supplementary Methods) or a CLIA-certified non-NGS mutation detection assay. Overall, 20 of 21 low frequency variants were confirmed from AmpliSeq assay and other hotspot clinical assay (Supplementary Table 2g).

Detection of Complex Gene Rearrangements

In addition to rearrangements that result in expression of a fusion transcript containing a known or putative oncogene (Figure 3A), complex gene rearrangements are often identified in patients with hematologic malignancies. These include large intragenic rearrangements such as *MLL* partial tandem duplications (PTD)¹⁵, extra-genic rearrangements that involve the immunoglobulin genes, and smaller internal gene rearrangements, such as *FLT3* internal tandem duplications (ITD)¹⁶ that were assessed in the concordance with reference platforms above (Figure 2D). We profiled 12 samples that were *MLL*-PTD positive by PCR, spanning 3 different isoforms (Figure 3B), and 14 that were negative by PCR. The NGS-based assay identified *MLL*-PTDs in 11 of the 12 PCR positive samples and confirmed the negative PCR calls in all 14 samples, for an overall accuracy of 96.5% (25/26, the single missed PTD was present below the calling threshold). Rearrangements involving the *IGH* locus commonly occur in hematologic malignancies, including non-Hodgkin lymphoma¹⁷. The *IGH* rearrangement-

calling algorithm enables detection of known *IGH* rearrangements in addition to novel partners. *IGH* rearrangement detection was assessed using 10 cell-lines and 38 clinical samples harboring *IGH* rearrangements detected by FISH and/or karyotypic data (Figure 3C, Supplementary Table 3b-c). The 10 cell lines were fully concordant while the clinical samples had two discordant positive calls and one discordant negative call for an overall concordance of 94% (45/48) (Supplementary Table 3b-c).

Assay Reproducibility in Clinical Samples

To demonstrate assay reproducibility, we tested 13 clinical FFPE specimens in five replicates, each in three different batches with one batch including three intra-batch replicates. In total, the 65 samples contained 82 alterations (42 subs, 13 indels, 15 CNAs and 12 rearrangements) providing a comprehensive assessment of intra- and inter-batch reproducibility. Concordance between replicates was 97% for both inter- and intra-batch subsets (Supplementary Table 4) with discordant calls being ascribed to sensitivity limit of the assay. Long-term reproducibility was assessed through serial assessment of two pooled samples (UHRR for RNA, a mixture of 10 HapMap samples for DNA). These samples were sequenced repeatedly over 5 months; 303 SNPs (100% 134/134) in DNA and *BCR-ABL* fusion (100% 132/132) in RNA were successfully detected over the entire time series.

Workflow Compatibility with Different Input Materials

We next studied the assays performance using clinical FFPE, blood, and bone marrow aspirate samples. Blood and bone marrow were collected in EDTA for both DNA and RNA extractions. DNA and RNA were obtained from ten different blood, bone marrow, and FFPE specimens and nucleic acids were extracted according to protocols described in the online methods. Of these 40 samples, 39 of 40 exceeded performance specifications with the only failing sample coming from one of 10 FFPE blocks analyzed (Supplementary Table 5).

Clinical Experience to Date

The new assay was used to perform genomic profiling on 3696 hematologic malignancies submitted to our CLIA-certified, CAP-accredited, NYS-approved laboratory. 3433 out of 3696 (93%) specimens submitted for processing were successfully characterized, including 27% FFPE samples, 21% from bone marrow aspirates, 18% samples derived from blood and 34% samples from pre-extracted nucleic acids from relevant tumor tissues. This cohort included 39% MM, 22% NHL, 17% AML and 13% MDS/MPN specimens (Figure 4A). Different subtypes of NHL in our clinical experiences include DLBCL (41%), follicular lymphoma (13%), B-cell lymphoma (NOS) (8%) and mantle cell lymphoma (8%) (Figure 4B). Extracted DNA was sequenced to an average depth of 500x and RNA to an average of 6.9 M unique pairs. 263 out of 3696 (7%) specimens did not attain quality control specifications including failures in DNA preparation, or failures in RNA preparation, or blood sample older than acceptance criteria (one day old) and no evidence for tumor, or no somatic driver mutations reported under qualifying conditions such as low coverage (DNA < 250x), low tumor purity (< 20%), coverage bias impairing ability to call CNAs (Supplemental table 6a). Genomic profiling of hematologic malignancies demonstrates that only a relatively small number of genes are commonly mutated while many specimens harbor a wide range of rarer events (Figure 4C), including base substitutions, indels, copy number amplification/losses and rearrangements (Figure 4D). At least one driver alteration was identified in 3246/3433 (95%) tumor specimens, and 2650 (77%) cases harbored at least one alteration linked to a commercially-available targeted therapy or one that is in clinical development, including *NRAS* (14% of cases), *KRAS* (13%), *DNMT3A*

(7%), *CDKN2A* (7%), *IDH1/2* (5%), *BRAF* (4%) and *FLT3* (4%). In addition, 61% of cases harbored at least one alteration with known prognostic relevance in that tumor type, including *TP53* (19% of cases), *ASXL1* (9%), *TET2* (8%), *CDKN2B* (5%), *CREBBP* (5%), *MLL* (4.0%) and *NPM1* (2.0%) (Figure 4C). Disease-specific mutational landscapes of multiple myeloma (MM), non-Hodgkin lymphoma (NHL), acute lymphoblastic leukemia (ALL), AML and MDS are further presented in Figure 5 with associated clinical relevance, and a full listing of observed genes and mutation frequencies are listed in Supplemental Table 6b.

In total, we identified 1524 genomic rearrangements in 1256 of 3433 (37%) tumor specimens. These rearrangements involved genes implicated in chromatin/histone remodeling (20%), transcriptional regulation (16%), kinase/oncogene activation (15%), apoptosis regulation (13%), cell cycle regulation (13%) and truncation of tumor suppressors (5%), including novel in-frame fusions in kinase drug targets in *ALK*, *BRAF*, *FLT3*, *JAK2*, and *ROS1* (Figure 6A). In addition, we identified a novel translocation involving *NOTCH1*, as well as several novel *NF1* translocations that might confer sensitivity to MEK inhibitors.

Rearrangements involving the *IGH* locus were identified in 665/3433 tumor specimens, including rearrangements involving *CCND1/MYEOV* (28%), *BCL2* (25%), *MMSET/FGFR3* (17%), *MYC* (6%) and *BCL6* (4%) (Figure 6B, Supplementary Table 6c). Novel *IGH* rearrangements were observed in 17% of cases. The locations of the most common *IGH* rearrangement partners are summarized in Supplementary Figure 4. In addition, a sizable fraction of *IGH* rearrangements involved breakpoints in intergenic regions with no known oncogene within 7MB of the breakpoint. Genomic rearrangements involving *IGH* are believed to lead to over-expression of the partner oncogene by placing it under the regulatory control of the *IGH* locus¹⁸. Expression levels of *MMSET/FGFR3* (4p16.3) and *CCND1* genes were successfully measured for 449 multiple myeloma cases (Figure 6C). As depicted in Figure 6C, we observed over-expression of both *CCND1/FGFR3* and *CCND1* genes in samples harboring the respective rearrangements.

Identification of Different Classes of Alterations in a Single Gene Using an Integrated Genomic Profiling Assay

Previous studies have shown that a subset of oncogenes can be activated by different types of genomic alterations in different patients¹⁹. Constitutive activation of the *FLT3* receptor tyrosine kinase can occur in a significant proportion of leukemia patients as a result of somatic internal tandem duplications (ITD), point mutations (D835/I836) or less common chromosomal rearrangements^{16,20}. Somatic alterations involving the *FLT3* locus have been shown to be of prognostic importance in acute leukemia, and there are potent, specific *FLT3* kinase inhibitors in late-stage clinical trials²¹. We observed 89 alterations targeting *FLT3* in 84 specimens. This included 67 (75%) D835 and ITD mutations detected by existing clinical tests. We also identified other base substitutions (n=14), in-frame deletion (n=1), CNAs (n=6) and fusions involving *FLT3* (n=1) (Figure 6D). For example, one AML case that tested negative for D835Y harbored an I836 del, which has previously been associated with poor disease-free survival²². Four cases positive for *FLT3* ITD or D835 were found to also harbor concurrent low frequency V592A and S451F mutations, which have been functionally validated as activating the *FLT3* kinase.²³ These results indicate that the sequencing-based assay can robustly detect a wide range of alterations in driver genes that are not fully evaluated using conventional methods.

Use of a Single Unified Test in B-cell ALL

Recent genomic studies in B-cell ALL have led to an increased understanding of the molecular pathogenesis of this B-cell malignancy. This includes the identification of recurrent substitutions, indels, and fusion genes that target B-cell differentiation. Most importantly, recent studies have shown that subsets of patients with high-risk, BCR-ABL-negative B-ALL have a “Philadelphia chromosome-like (Ph-like)” gene expression signature, and that the majority of cases of Ph-like ALL harbor somatic alterations in known signaling effectors, including in known and putative therapeutic targets²⁴⁻²⁶. However, current approaches to identify the substitution mutations, indels, and fusion genes that drive Ph-like ALL require the use of a series of genomic assays. We performed integrated DNA/RNA profiling on 16 cases classified as Ph-like ALL by gene expression profiling. We found deletions in *IKZF1* (7 of 16), *PAX5* (2 of 16) and *CDKN2A* (4 of 16) in a subset of these cases, consistent with previous reports²⁷. We also identified known activating mutations in *CRLF2*, *KRAS* and *JAK2* in some cases. Notably, we identified known and novel gene fusions in 9 of 16 cases. This included 2 *JAK2* fusions with different partner genes (*PAX5* and *SSBP2*), 6 known fusions of *CRLF2* and *EPOR* and a novel gene fusion involving *ABL2*, all of which would be expected to activate kinase signaling and to confer sensitivity to approved drugs. In one case of B-ALL, we identified a *FIP1L1-PDGRA* fusion, which has been shown to respond to imatinib in patients with hypereosinophilic syndrome²⁸ (Table 1). These data show that an integrated DNA/RNA profiling platform can identify a spectrum of alterations in kinase signaling pathway in B-ALL patients that can guide the use of molecularly targeted therapies. The result also achieved high concordance with an integrated research profiling approach which included genome sequencing, targeted DNA sequencing, RNA-sequencing, and SNP array analysis done on the same samples. Specifically, we observed complete concordance between our clinical genomic profiling assay and research-grade genomic analysis, with the exception a *CRLF2* mutation and a *CDKN2A/B* deletion that were detected only by the clinical sequencing assay.

Discussion

The identification of somatic genomic alterations with clinical relevance has increased the need for robust genomic profiling assays that can identify the different types of genomic alterations in clinically relevant cancer genes. Here we report the development of an integrated DNA/RNA target capture next-generation sequencing assay, which we have optimized to detect different classes of genomic events, including base substitutions, indels, CNAs, and chromosomal rearrangements. This allows the use of a single test to systematically profile clinical samples from patients with hematologic malignancies for actionable genomic lesions, in a clinically relevant time-frame and quality standard. As described in results, 93% specimens were successfully profiled in the clinical setting. 7 % of patients received failed report due to failures in DNA or RNA preparation, process of aged liquid specimen, and no somatic mutations detected under other qualifying conditions. One of the most common qualifying conditions is RNA preparation failure (132 cases) which highlighted the technical challenges of process and sequence RNA samples (Supplemental table 6a).

Previously described assays have largely focused on DNA-based approaches, or have used gene-specific RNA tests to detect fusion transcripts in a small set of genes. We show that targeted RNA capture sequencing can be used to identify a wide spectrum of fusion genes with known roles in malignant transformation and therapeutic response, and to identify novel fusions with likely diagnostic and therapeutic relevance. This approach can identify fusion transcripts present in as few as 10-20% of cells, and can be used to show that chromosomal rearrangements result in expression of in-frame, fusion transcripts. Moreover, combining DNA and RNA capture allows for the identification of full-length transcripts driven by promoter

rearrangements by DNA-sequencing, as well as fusion transcripts which are more easily identified by RNA profiling. This approach has immediate clinical value in hematologic malignancies, and we believe will show significant clinical value in epithelial tumors. Of note, this approach can also be used to reliably measure expression of genes in the target set, which will inform development and testing of genomic predictors that combine mutational data and gene expression/epigenetic data to inform prognosis and to identify novel therapeutic targets.

The clinical relevance of our DNA/RNA profiling²⁹ approach is underscored by our ability to identify genetic lesions with prognostic and therapeutic relevance in specific diseases. In the case of B-cell ALL, recent studies have identified novel genetic markers with prognostic relevance, and identified novel therapeutic targets in patients with high-risk disease^{24-27,30,31}. The challenge has been that the critical genes implicated in B-ALL can be altered by whole gene/intragenic deletions, DNA base pair substitutions and larger indels, as well as chromosomal, intergenic and cryptic rearrangements which lead to expression of fusion transcripts. Currently, most centers use an amalgam of DNA, FISH, and gene-specific RNA approaches to identify a subset of the most critical genetic lesions in B-ALL. Our assay provides a single profiling platform that can reliably identify all known actionable disease alleles relevant to B-ALL to improve diagnosis and risk-adapted therapy for B-ALL patients.

Most importantly, our profiling platform, allows for the increased use of genomic profiling to improve the diagnosis and treatment of cancer patients and to more precisely match patients with targeted therapies. These efforts are critical to more broadly expand precision medicine by assuring that cancer patients can be offered genomic profiling regardless of whether they are at tertiary care centers or receive care in other settings.

Authorship

Contribution: JH, OAW, DL, GAO and RL designed the study, wrote the manuscript, and developed and/or performed analysis. PJS and VAM designed the study and wrote the manuscript. OAW, RKR, AMI, JP, AK, AY, AH, EP, KR, CM, AD, SA, MRM and RL contribute clinical and genomic expertise. JH, MKN, KW, TB, GFF, LEY, SH, MB performed analysis. JH, LEY, JRW, SR, JD and AF developed computational tools. MKN, KWB, ALD, GY, EW and GAO developed the assay. LG, STB, MZ, VB, JB, and KI performed laboratory experiments. RY, TM, JAV, JSR and DM evaluated and edited the manuscript.

Conflict-of-interest disclosure: JH, MKN, KW, GFF, LEY, SH, MB, JRW, SR, JD, TB, MR, KP, KWB, ALD, GY, LG, STB, MZ, EW, VB, JB, KI, JSS, DM, TM, JAV, EL, RE, CV, RY, PJS, VAM, GAO, and DL are employees and equity holders of Foundation Medicine, Inc. OAW, AD, SA, MRM and RLL are consultants for Foundation Medicine, Inc.

Correspondence: Geoff Otto, e-mail: gotto@foundationmedicine.com, Doron Lipson, e-mail dlipson@foundationmedicine.com, and Ross L. Levine, e-mail leviner@mskcc.org

Figure Legends

Figure 1 – Workflow of comprehensive combined DNA and RNA genomic analysis in clinical specimens and coverage distribution of all exons.

A. DNA and RNA are extracted from fresh blood and bone marrow aspirate (BMA) specimens procured in EDTA or FFPE biopsy/surgical specimens; cDNA from 300-500ng of RNA and 50-200 ng of DNA undergoes whole-genome shotgun library construction. cDNA libraries are hybrid capture selected for 265 genes known to be rearranged in RNA and DNA libraries are hybrid capture selected for 405 genes known to be altered in DNA. Hybrid-capture selected libraries are sequenced to high depth using the Illumina HiSeq2500 platform; sequence data is processed using a customized analysis pipeline designed to accurately detect multiple classes of genomic alterations: base substitutions, short insertions/deletions, copy-number alterations and gene fusions; detected mutations are annotated according to clinical significance and reported. B. Coverage distribution in DNaseq of all genes in 108 validation samples including 4 Hapmap controls and 104 hematologic tumor specimens.

Figure 2 – Alteration Calling Accuracy and Concordance with Reference Platforms.

A. Correlation of short variant mutant allele frequencies (MAFs) for substitutions and indels called by new versus previously validated assays. MAFs of 102 substitutions and 59 indels were compared in pairs and showed significant correlation ($R^2=0.929$); B. Summary of sensitivity of fusion calling in cell line mixes, separated by detection method (DNA, RNA or both); C. High concordance (99%) observed between NGS and Sanger sequencing or PCR validation results. D. Concordance of alterations between the new assay and other CLIA-certified reference platforms. The count of positive call and negative call from reference test and our NGS-base assay from the 11 genes/regions are highlighted by different colors: dark green (alteration was called negative by reference test but positive by our NGS test), light green (alteration was called positive in reference test and negative in our NGS test), orange (alteration was called negative in both reference test and negative test) and green (alteration was called positive in both reference test and negative test).

Figure 3 – Combined DNA and RNA Sequencing for Detection of Complex Rearrangements.

A. Circos plot showing broad range of fusions in cell line mixes used for validation of fusion calling. B. Three *MLL*-PTD isoforms that were tested (from top to bottom): duplication of *MLL* exons 2-8, exons 2-10 and exons 4-8. C. Breakpoints on the *IGH* and its partner genes for the 13 clinical concordance samples including 5 *MYC-IGH*, 2-*BCL6-IGH*, 5 *CCND1-IGH* and 11 *BCL2-IGH* rearrangements.

Figure 4 – Overview of clinical cases profiled by combined DNA/RNA next-generation sequencing assay.

A. Tumor type distribution of all hematologic tumor cases profiled by the NGS-based DNA/RNA test. B. Disease ontology distribution of 297 non-Hodgkin lymphoma tumors profiled by the assay. PMBCL = primary mediastinal (thymic) large B-cell lymphoma, AITL = angioimmunoblastic T-cell lymphoma. C. Genomic alterations detected from the 1931 clinical specimens and distribution of alteration types. Top 50 most frequent altered genes are displayed. Dark green: base substitutions, orange: indels, purple: copy number alterations

(CNAs), light green: rearrangement. D. Distribution of alteration types including base substitutions, indels, copy number alterations and rearrangements in the clinical specimens.

Figure 5 – Genomic landscape in ALL, AML, MDS, NHL and MM.

Observed frequencies of most mutated genes in A. Acute lymphoblastic leukemia (ALL), B. Acute Myeloid Leukemia (AML), C. Myelodysplastic Syndromes (MDS), D. Non-Hodgkin Lymphoma (NHL), and E. Multiple Myeloma (MM). Mutation frequency is grouped by function effect, substitution/indel (green), focal amplification (orange), homozygous deletion (purple), rearrangement/fusion (light green) and truncation (pink). The numbers of cases included in this analysis are 102 cases in ALL, 274 cases in AML, 130 cases in MDS, 297 cases in NHL, and 753 cases in MM. In The clinical relevance including therapeutic (red), clinical trial (orange), prognostic (green), and diagnostic (purple) are highlighted on top of each gene.

Figure 6 – Clinical Highlights

A. Novel fusions involving druggable targets such as *ALK*, *BRAF*, *FLT3*, *JAK2* and *ROS1* (5' partner blue, 3' partner including kinase domain of target gene orange; ALCL: anaplastic large cell lymphoma; MM: multiple myeloma; HL: Hodgkin lymphoma; B-ALL: B-cell acute lymphoblastic leukemia). B. Distribution of IGH rearrangement partners. C. (i,ii) Overexpression of (i) *FGFR3* and (ii) *WHSC1* in t(4;14) positive multiple myeloma (MM) cases (n=61) compared with t(4;14) negative MM cases (n=388). The violin plot shows the distribution of the log expression ratios of *FGFR3* and *WHSC1* in each group respectively. For both cases, the difference is statistically significant (*FGFR3*, $p=8.7e^{-22}$; *WHSC1*, $p=5.28e^{-32}$, one-sided Wilcoxon rank-sum test). (iii) Overexpression of *CCND1* in all t(11;14) positive cases (n=67) compared with t(11;14) negative cases (n=3095). The difference is also statistically significant ($p=6.00e^{-38}$, one-sided Wilcoxon rank-sum test). D. Overview of short variants [single nucleotide variants (SNVs), internal tandem duplicates (ITDs), indels] and large-scale aberrations including long deletions, copy number gains, and fusions detected in *FLT3*. The FLT3 protein is drawn with its immunoglobulin-like (Ig, green), transmembrane (Tm, purple) and tyrosine kinase (Tk, orange) domains illustrated. Positions and protein effect of all short variants are indicated. For ITDs (dark green), samples with ITD at the same positions are aggregated together regardless of the actual insert sequence. The ranges of large-scale aberrations are indicated with cyan bars. Commonly tested “hotspot” mutations are shaded in light yellow.

Table 1 Summary of genomic alterations identified in Ph-like B-cell Acute Lymphoblastic Leukemia Cases.

Summary of genomic alterations including kinase fusions, kinase mutation and deletions in *IKZF1*, *PAX5*, *EBF1* and *CDKN2A/B* in 16 Ph-like ALL patients. * represent mutations that were detected only by the clinical sequencing assay.

References

1. Bejar R, Stevenson K, Abdel-Wahab O, et al. Clinical effect of point mutations in myelodysplastic syndromes. *N Engl J Med*. 2011;364(26):2496-2506.
2. Flaherty KT, Puzanov I, Kim KB, et al. Inhibition of mutated, activated BRAF in metastatic melanoma. *N Engl J Med*. 2010;363(9):809-819.
3. Tiacci E, Trifonov V, Schiavoni G, et al. BRAF mutations in hairy-cell leukemia. *N Engl J Med*. 2011;364(24):2305-2315.
4. Sjoblom T, Jones S, Wood LD, et al. The consensus coding sequences of human breast and colorectal cancers. *Science*. 2006;314(5797):268-274.
5. Cancer Genome Atlas Research N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-1068.
6. Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007;446(7132):153-158.
7. Frampton GM, Fichtenholtz A, Otto GA, et al. Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol*. 2013;31(11):1023-1031.
8. Nucifora G, Birn DJ, Erickson P, et al. Detection of DNA rearrangements in the AML1 and ETO loci and of an AML1/ETO fusion mRNA in patients with t(8;21) acute myeloid leukemia. *Blood*. 1993;81(4):883-888.
9. Golub TR, Barker GF, Bohlander SK, et al. Fusion of the TEL gene on 12p13 to the AML1 gene on 21q22 in acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 1995;92(11):4917-4921.
10. Rowley JD. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature*. 1973;243(5405):290-293.
11. Dalla-Favera R, Martinotti S, Gallo RC, Erikson J, Croce CM. Translocation and rearrangements of the c-myc oncogene locus in human undifferentiated B-cell lymphomas. *Science*. 1983;219(4587):963-967.
12. Byrd JC, Mrozek K, Dodge RK, et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood*. 2002;100(13):4325-4336.
13. Welch JS, Ley TJ, Link DC, et al. The origin and evolution of mutations in acute myeloid leukemia. *Cell*. 2012;150(2):264-278.
14. Caligiuri MA, Strout MP, Schichman SA, et al. Partial tandem duplication of ALL1 as a recurrent molecular defect in acute myeloid leukemia with trisomy 11. *Cancer Res*. 1996;56(6):1418-1425.
15. Schnittger S, Kinkelin U, Schoch C, et al. Screening for MLL tandem duplication in 387 unselected patients with AML identify a prognostically unfavorable subset of AML. *Leukemia*. 2000;14(5):796-804.
16. Armstrong SA, Mabon ME, Silverman LB, et al. FLT3 mutations in childhood acute lymphoblastic leukemia. *Blood*. 2004;103(9):3544-3546.
17. Bernicot I, Douet-Guilbert N, Le Bris MJ, et al. Characterization of IGH rearrangements in non-Hodgkin's B-cell lymphomas by fluorescence in situ hybridization. *Anticancer Res*. 2005;25(5):3179-3182.
18. Vincent-Fabert C, Fiancette R, Cogne M, Pinaud E, Denizot Y. The IgH 3' regulatory region and its implication in lymphomagenesis. *Eur J Immunol*. 2010;40(12):3306-3311.
19. Polsky D, Cordon-Cardo C. Oncogenes in melanoma. *Oncogene*. 2003;22(20):3087-3091.

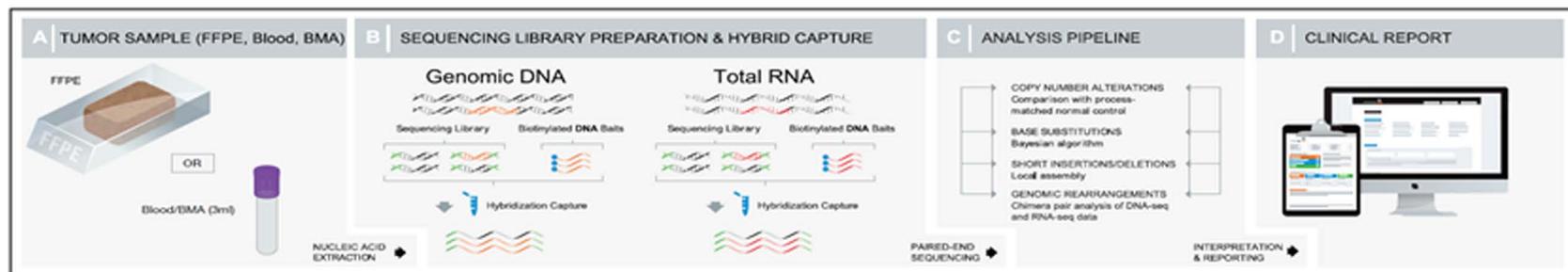
20. Wellmann S, Moderegger E, Zelmer A, et al. FLT3 mutations in childhood acute lymphoblastic leukemia at first relapse. *Leukemia*. 2005;19(3):467-468.
21. Smith CC, Wang Q, Chin CS, et al. Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. *Nature*. 2012;485(7397):260-263.
22. Whitman SP, Ruppert AS, Radmacher MD, et al. FLT3 D835/I836 mutations are associated with poor disease-free survival and a distinct gene-expression signature among younger adults with de novo cytogenetically normal acute myeloid leukemia lacking FLT3 internal tandem duplications. *Blood*. 2008;111(3):1552-1559.
23. Frohling S, Scholl C, Levine RL, et al. Identification of driver and passenger mutations of FLT3 by high-throughput DNA sequence analysis and functional assessment of candidate alleles. *Cancer Cell*. 2007;12(6):501-513.
24. Roberts KG, Morin RD, Zhang J, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell*. 2012;22(2):153-166.
25. Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;371(11):1005-1015.
26. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009;360(5):470-480.
27. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446(7137):758-764.
28. Cools J, DeAngelo DJ, Gotlib J, et al. A tyrosine kinase created by fusion of the PDGFRA and FIP1L1 genes as a therapeutic target of imatinib in idiopathic hypereosinophilic syndrome. *N Engl J Med*. 2003;348(13):1201-1214.
29. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-1068.
30. Yoda A, Yoda Y, Chiaretti S, et al. Functional screening identifies CRLF2 in precursor B-cell acute lymphoblastic leukemia. *Proc Natl Acad Sci U S A*. 2010;107(1):252-257.
31. Bercovich D, Ganmore I, Scott LM, et al. Mutations of JAK2 in acute lymphoblastic leukaemias associated with Down's syndrome. *Lancet*. 2008;372(9648):1484-1492.

Table 1

Specimen ID	Kinase fusion	Kinase CNA/sequence mutation	RAS pathway	Other lesion	<i>IKZF1</i> deletion or mutation	<i>PAX5</i> deletion or mutation	<i>EBF1</i> alteration	<i>CDKN2A/B</i> deletion
ALL1	<i>PAX5-JAK2</i>	Unknown			Deletion	No	Yes	No
ALL2								
ALL3	<i>IGH@-CRLF2</i>	<i>CRLF2</i> F232C*			Deletion(Single exon deletion)	No	Yes	No
ALL4								
ALL5	<i>SSBP2-JAK2</i>	Unknown			Deletion	No	No	Yes
ALL6								
ALL7	<i>RSCD1-ABL2</i>	Unknown			Deletion	No	No	No
ALL8	<i>IGH@-EPOR</i>		<i>KRAS</i> G12D		Deletion	Deletion and mutation	No	Yes
ALL11				<i>CREBBP</i> R1081fs <i>ETV6</i> focal deletion	Deletion(single exon deletion)	No	No	Yes*
ALL12								
ALL13				<i>IGH-CRLF2</i>				
ALL14	<i>P2RY8-CRLF2</i>	<i>JAK2</i> T875N			Deletion	Deletion	No	Yes
ALL15								
ALL16	<i>IGH@-CRLF2</i>	Unknown		<i>PAX5-MEGF9</i>			No	
ALL17				<i>FIP1L1-PDGFR</i>				
ALL18								

Figure 1

A

**Sample requirements**

- Tumor content: $\geq 20\%$ of nucleated cellularity; $\geq 30,000$ nucleated cells
- FFPE: Fifteen 5 micron sections of a $5 \times 5\text{mm}^2$ piece of tissue
- Peripheral blood/bone marrow aspirate: each collected in an EDTA tube

Laboratory process highlights

- Requires $\geq 50\text{ng}$ of dsDNA (quantified by PicoGreen[®])
- Requires $\geq 300\text{ng}$ of RNA (quantified by RiboGreen[®])
- RNA is converted to cDNA
- Fragmentation by sonication (Covaris[®]) and 'with-bead' library construction
- Hybridization capture with biotinylated DNA oligonucleotides
- 49x49 paired-end sequencing on the Illumina HiSeq platform
- DNA to $\sim 500\text{x}$ average unique coverage
- RNA to $>3\text{M}$ unique pairs

Analysis methods highlights

- Sensitivity to variants present at $>5\%$ mutant allele frequency
- Detection of long (40bp+) indel variants using de-Brujin graph-based local assembly
- CGH-like analysis of read-depth for CNAs assessment
- Genomic rearrangements and gene fusions are identified by analyzing chimeric read pairs from DNA-seq and RNA-seq

Reporting approach

- Interpretation without a matched normal
- Germline variants from 1000 Genomes project (dbSNP135) removed
 - Known driver alterations (COSMIC v62, internal CNA and fusion list) identified as biologically significant
 - A concise summary of the biomedical literature and current clinical trials is provided per alteration

B

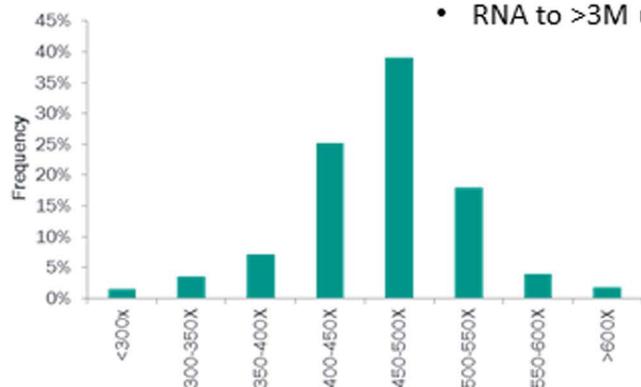


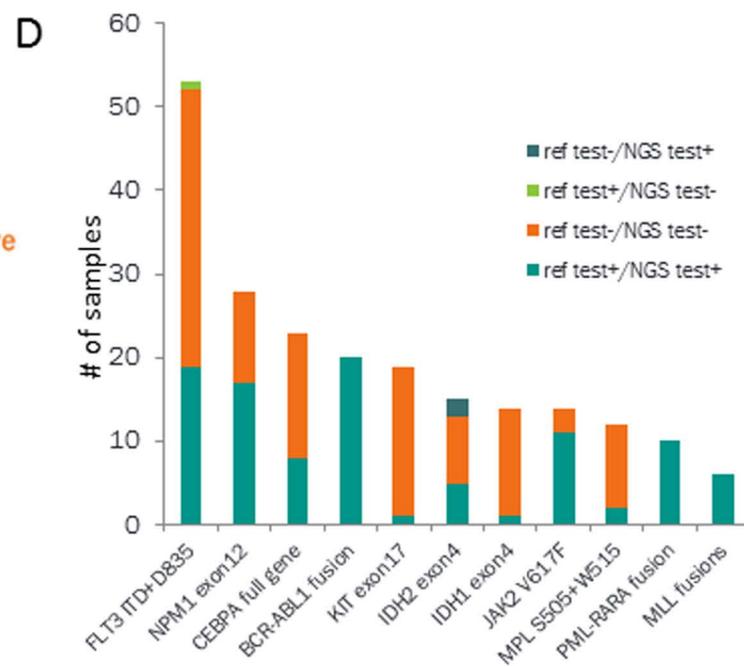
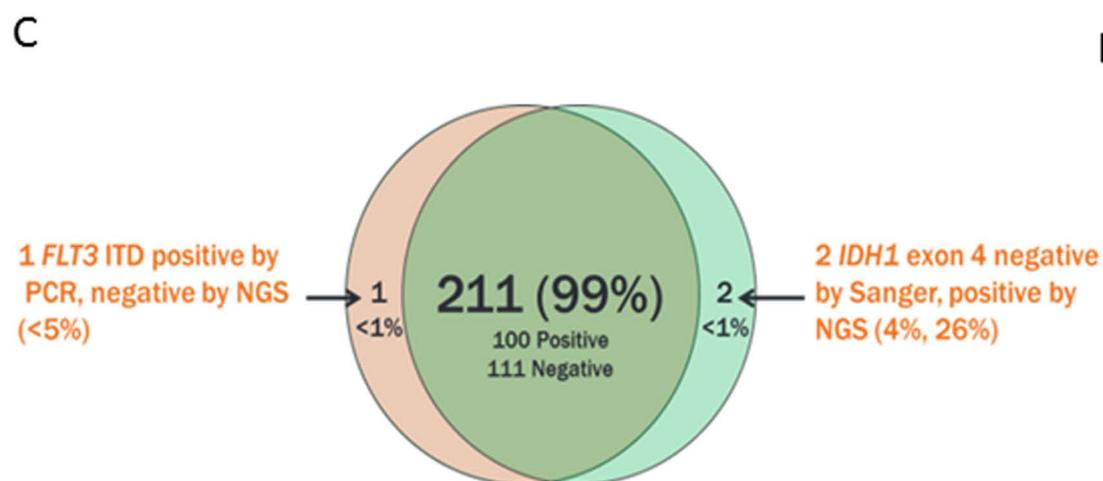
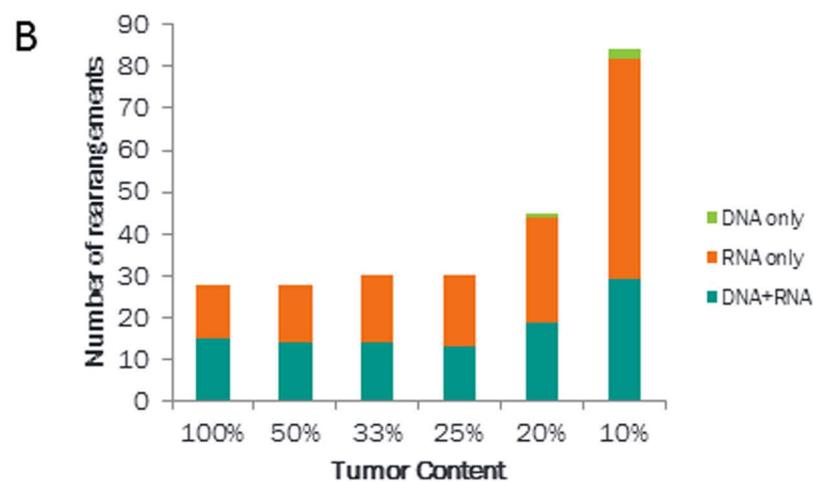
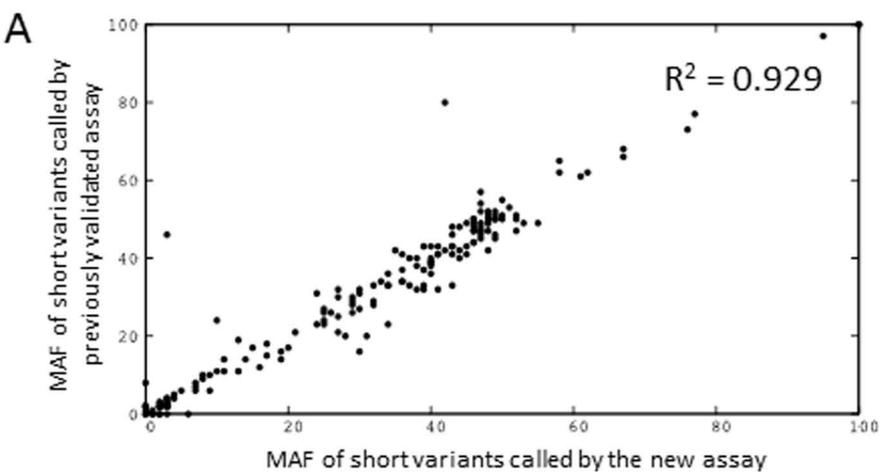
Figure 2

Figure 3

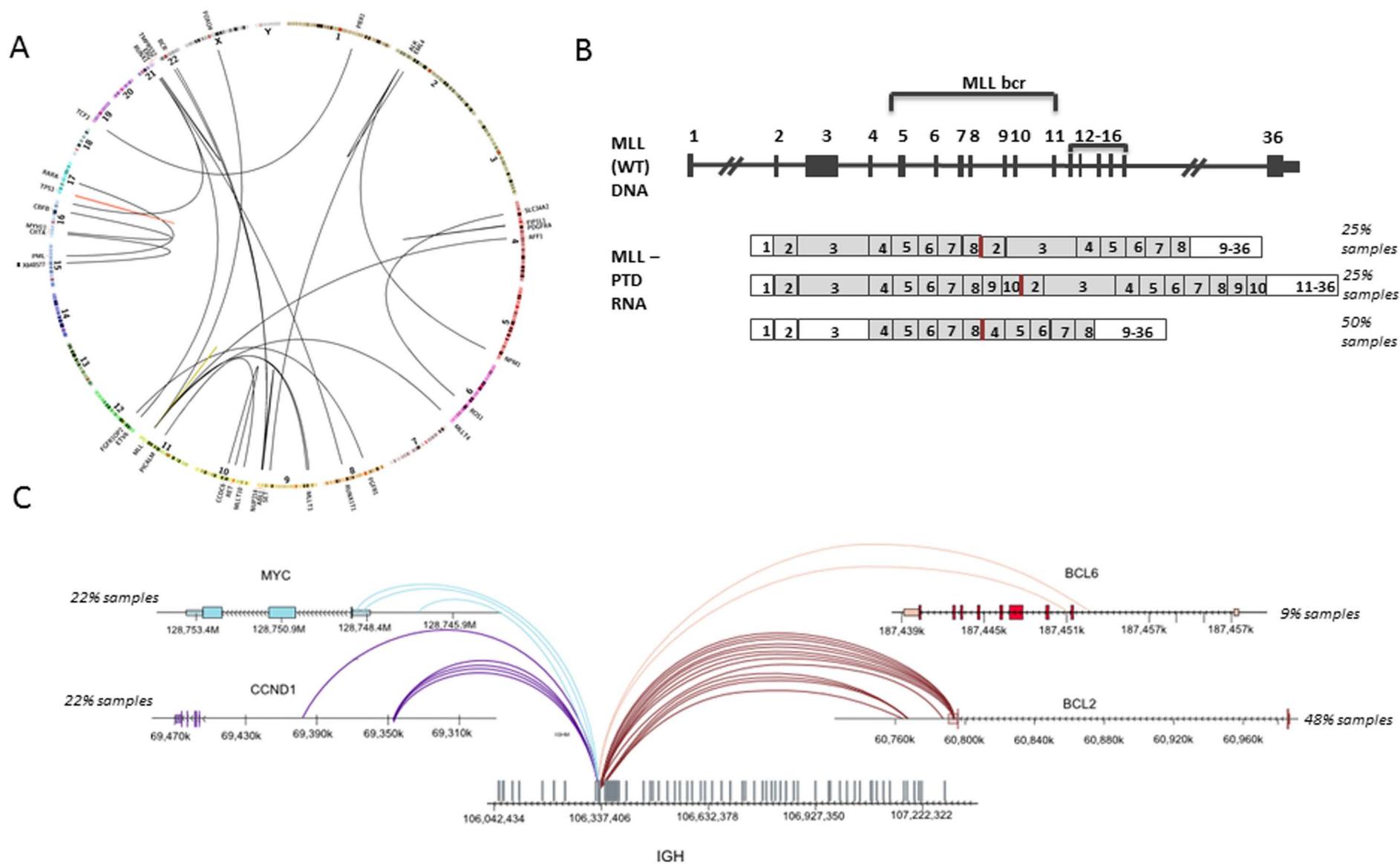
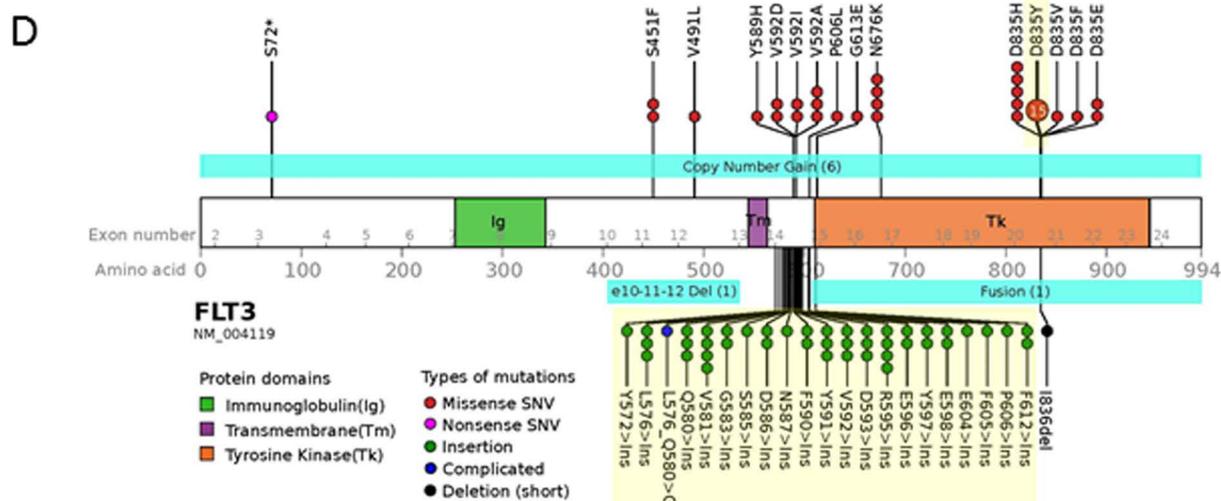
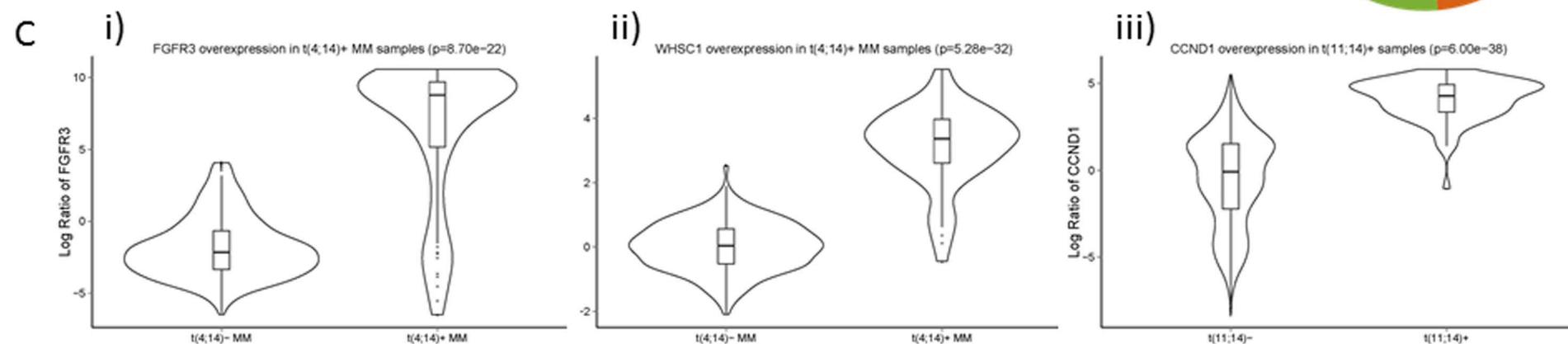
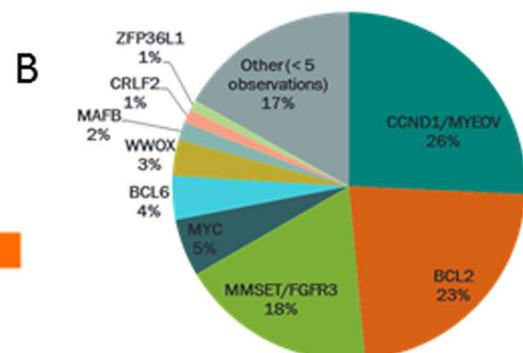
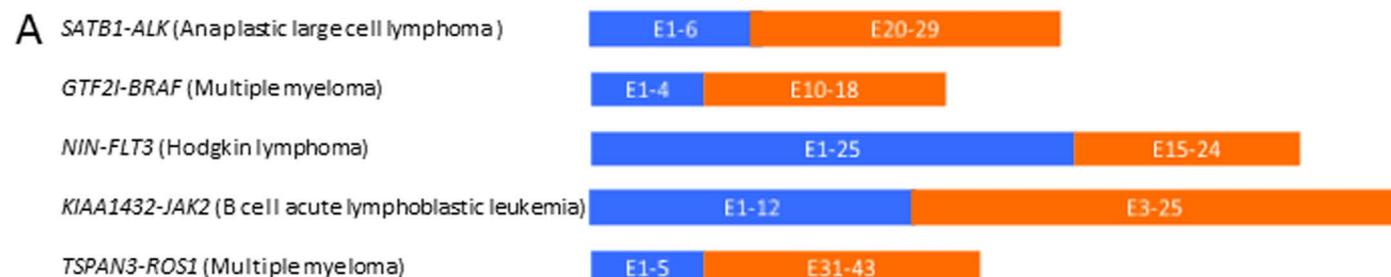


Figure 6





blood[®]

Prepublished online March 10, 2016;
doi:10.1182/blood-2015-08-664649

Integrated genomic DNA/RNA profiling of hematologic malignancies in the clinical setting

Jie He, Omar Abdel-Wahab, Michelle K. Nahas, Kai Wang, Raajit K. Rampal, Andrew M. Intlekofer, Jay Patel, Andrei Krivstov, Garrett M. Frampton, Lauren E. Young, Shan Zhong, Mark Bailey, Jared R. White, Steven Roels, Jason Deffenbaugh, Alex Fichtenholtz, Timothy Brennan, Mark Rosenzweig, Kimberly Pelak, Kristina M. Knapp, Kristina W. Brennan, Amy L. Donahue, Geneva Young, Lazaro Garcia, Selmira T. Beckstrom, Mandy Zhao, Emily White, Vera Banning, Jamie Buell, Kiel Iwanik, Jeffrey S. Ross, Deborah Morosini, Anas Younes, Alan M. Hanash, Elisabeth Paietta, Kathryn Roberts, Charles Mullighan, Ahmet Dogan, Scott A. Armstrong, Tariq Mughal, Jo-Anne Vergilio, Elaine Labrecque, Rachel Erlich, Christine Vietz, Roman Yelensky, Philip J. Stephens, Vincent A. Miller, Marcel R. M. van den Brink, Geoff A. Otto, Doron Lipson and Ross L. Levine

Information about reproducing this article in parts or in its entirety may be found online at:
http://www.bloodjournal.org/site/misc/rights.xhtml#repub_requests

Information about ordering reprints may be found online at:
<http://www.bloodjournal.org/site/misc/rights.xhtml#reprints>

Information about subscriptions and ASH membership may be found online at:
<http://www.bloodjournal.org/site/subscriptions/index.xhtml>

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include digital object identifier (DOIs) and date of initial publication.