The Journal of Molecular Diagnostics, Vol. 19, No. 1, January 2017

ELSEVIER

the Journal of Molecular Diagnostics

jmd.amjpathol.org

Clinical Validation of a Next-Generation Sequencing Genomic Oncology Panel via Cross-Platform Benchmarking against Established Amplicon Sequencing Assays

Sabah Kadri,* Bradley C. Long,* Ibro Mujacic,* Chao J. Zhen,* Michelle N. Wurst,* Shruti Sharma,* Nadia McDonald,* Nifang Niu,* Sonia Benhamed,* Jigyasa H. Tuteja,† Tanguy Y. Seiwert,‡ Kevin P. White,† Megan E. McNerney,* Carrie Fitzpatrick,* Y. Lynn Wang,* Larissa V. Furtado,* and Jeremy P. Segal*

From the Division of Genomic and Molecular Pathology, Department of Pathology, the Institute for Genomics and Systems Biology,† and the Department of Medicine,‡ The University of Chicago, Chicago, Illinois*

Accepted for publication July 19, 2016.

Address correspondence to Jeremy P. Segal, M.D., Ph.D., Division of Genomic and Molecular Pathology, Department of Pathology, University of Chicago, 5841 S Maryland Ave, MC 1089, Room N-339, Chicago, IL 60637. E-mail: jsegal5@bsd.uchicago.edu.

Next-generation sequencing (NGS) genomic oncology profiling assays have emerged as key drivers of personalized cancer care and translational research. However, validation of these assays to meet strict clinical standards has been historically problematic because of both significant assay complexity and a scarcity of optimal validation samples. Herein, we present the clinical validation of 76 genes from a novel 1212-gene large-scale hybrid capture cancer sequencing assay (University of Chicago Medicine OncoPlus) using full-data comparisons against multiple clinical NGS amplicon-based assays to yield dramatic increases in per-sample data comparison efficiency compared with previously published validations. Using a sample set of 104 normal, solid tumor, and hematopoietic malignancy specimens, head-to-head NGS data analyses allowed for 6.8 million individual clinical base call comparisons, including 2729 previously confirmed variants, with 100% sensitivity and specificity. University of Chicago Medicine OncoPlus showed excellent performance for detection of single-nucleotide variants, insertions/deletions up to 52 bp, and *FLT3* internal tandem duplications of up to 102 bp or larger. Highly concordant copy number variant and *ALK/RET/ROS1* gene fusion detection were also observed. In addition to underlining the efficiency of NGS validation via full-data benchmarking against existing clinical NGS assays, this study also highlights the degree of performance similarity between hybrid capture and amplicon assays that is attainable with the application of strict quality control parameters and optimized computational analytics. *(J Mol Diagn 2017, 19: 43—56; http://dx.doi.org/10.1016/j.jmoldx.2016.07.012)*

Next-generation sequencing (NGS) has allowed for a dramatic expansion of diagnostic laboratory capabilities for surveillance of genomic alterations in cancer. Many clinical laboratories are now using NGS platforms for personalized oncology decision support or are in the process of developing laboratory-developed protocols (LDPs) of various types and scales to meet patient and oncologist demand.[1—7] However, the development of such assays represents a significant challenge, and demonstrating a level of performance sufficient to satisfy Clinical Laboratory Improvement Amendment (CLIA) requirements pertaining to the establishment of non—Food and Drug Administration—cleared LDPs can be a substantial hurdle.

The technical barriers to development of clinical oncology NGS assays are well documented, including specimen-related challenges (eg, small specimens, sample

Supported by the Department of Pathology at the University of Chicago.
S.K. and B.C.L. contributed equally to this work.
Disclosures: None declared.
Portions of this work were presented at the Association for Molecular Pathology Annual Meeting, November 5 to 7, 2015, Austin, TX.

Copyright © 2017 American Society for Investigative Pathology and the Association for Molecular Pathology. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jmoldx.2016.07.012

degradation, formalin fixation, and variable tumor cell purity) and factors related to the complexity of cancer genetics (eg, complex mutation types, rearrangements, copy number alterations, and subclonality).[5,8,9] More practical challenges related to the nature of desired validation strategies are less addressed. Ideally, LDP assay validations use samples previously analyzed in a CLIA laboratory via a validated assay targeting the same set of genomic features as the prospective assay. This provides high-confidence results using matched clinical specimens. However, this has not been a feasible option for most early adopter NGS laboratories conducting NGS LDP validations because of a shortage of available comparator laboratories able to share large numbers of NGS-tested samples along with complete lists of variant (and nonvariant) calls and mutant allele frequencies.

These shortcomings have led laboratories to pursue NGS oncology assay validations using alternative, less optimal sample sets. Common approaches include the use of oncology patient samples previously tested in CLIA laboratories via smaller, highly targeted assays (eg, Sanger sequencing and SNaPshot)[2–7] or the use of well-characterized cell lines previously sequenced at one or many research institutions.[1,2] The first option controls for sample type and allows for true clinical comparisons, but suffers from a lack of breadth, impeding comprehensive sensitivity and specificity analyses. The use of cell lines, on the other hand, allows for intercomparisons across broad genomic territory, but often lacks a true clinical laboratory gold standard, with data sets occasionally contaminated by sample-specific pseudogene signals and sequencing artifacts. Cell lines are also different sample types from those used in clinical service, and require alternative specimen handling protocols. Beyond the potential for discrepant prior sequencing results, cell lines may show genetic drift over serial passages,[10,11] leading to additional complications, including alterations of mutational burden and mutant allelic frequency (MAF). Thus, unlike a specific DNA sample tested in two laboratories, validation based on sequencing of publicly available cell lines involves additional caveats with the potential to inflate the observed false-positive and false-negative rates of the prospective LDP. Although some combination of the above methods may be used, given the troublesome and highly variable nature of clinical cancer samples [particularly formalin fixed, paraffin embedded (FFPE)], broad genomic assay intercomparisons of individual patient tumor samples would still be the most desirable approach from the standpoint of both cost and confidence.

One of the founding goals of our laboratory was to generate a large-scale genomic oncology panel suitable to help meet both the clinical and translational research needs of our institution. To accomplish this, we first validated separate smaller amplicon-based NGS panels for solid tumors and heme malignancies, accumulating a collection of patient samples tested via these assays during 1 to 2 years of clinical service. Subsequently, we have used a select group of these patient samples to validate a 76-gene clinical assay, University of Chicago Medicine (UCM) OncoPlus (UCM-OncoPlus), which represents a preliminary and expandable reportable subset of a larger comprehensive hybrid capture sequencing panel targeting 1212 cancer-associated genes.

In contrast to other published large-scale NGS validations,[1,2,5] the basis of this study is intensive comparison against multiple validated clinical NGS oncology assays using a variety of patient samples. UCM-OncoPlus was compared against a total of four NGS oncology platforms using three separate methods. Three comparator assays were internally validated in our laboratory (*Materials and Methods*), whereas the fourth assay is from an outside reference laboratory (Foundation One, Foundation Medicine).[1] Unlike samples tested via targeted assays or reference laboratory NGS assays (where only limited data points are returned), cross-validation of patient samples against full data sets from existing clinical NGS assays provides orders of magnitude more information per sample, allowing interrogation of the new assay across many thousands of base positions per sample. This enables extensive specificity and sensitivity analysis across large genomic regions, and also an intercomparison of detected MAFs for all variants. Other comparison strategies are much more limited. For example, comparison of an NGS LDP against a fragment length assay (eg, for *EGFR* exon 19) may yield only one comparable data point for each sample. Likewise, although commercial laboratory NGS assays may examine large areas of genomic territory, their results are always stripped of benign findings, including common inherited single-nucleotide and insertion/deletion polymorphisms and inherited copy number alterations. Variants of uncertain significance may be reported; however, the calling criteria for these anomalies are laboratory dependent, and thus the results are not helpful for specificity analysis. As a result, similar to targeted traditional PCR assays, these cross-comparisons may provide no more than a few comparable data points per sample for comparison, which is suboptimal for ascertainment of extensive sensitivity and specificity.

To our knowledge, this is the first reported oncology panel validation incorporating this efficient and advantageous strategy. Although other groups have raised questions about the potential level of concordance between variant calls from orthogonal amplicon and capture-based sequencing assay types,[12] our results highlight the performance similarity that is attainable between these methods with the application of optimized quality control parameters and informatics systems.

Materials and Methods

Sample DNA Preparation

DNA was prepared from FFPE tissue slides or from blood/bone marrow (Supplemental Table S1) using the QIAamp

DNA FFPE Tissue Kit or the QIAamp DNA Blood Mini Kit (Qiagen, Valencia, CA), respectively, per the manufacturer's instructions. After isolation, DNA was quantitated using a Qubit fluorometer and quantitation reagents (Life Technologies, Carlsbad, CA). DNA from FFPE specimens was also assessed for fragment size and amplifiable concentration using the KAPA hgDNA Quantification and QC Kit (Kapa Biosystems, Wilmington, MA).

CLIA-Validated Panels Used for Comparison Analysis

Three separate amplicon-based NGS oncology LDPs validated for patient care at the UCM Clinical Genomics Laboratory were used as comparators for the UCM-OncoPlus assay.

The first, OncoScreen ST1.0, is a 45-gene hot-spot NGS test for somatic mutations in solid tumors using the Truseq Amplicon Cancer Panel[13,14] kit (Illumina, San Diego, CA); per the manufacturer's instructions and clinically, it is run with two replicate library preparations with 250 ng FFPE DNA input per replicate (as assessed by Qubit), using the manufacturer's laboratory protocol. It contains amplicons for hot-spot locations of the following genes: *ABL1*, *AKT1*, *ALK*, *APC*, *ATM*, *BRAF*, *CDH1*, *CSF1R*, *CTNNB1*, *EGFR*, *ERBB2*, *ERBB4*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *GNA11*, *GNAQ*, *GNAS*, *HNF1A*, *HRAS*, *IDH1*, *JAK2*, *JAK3*, *KDR*, *KIT*, *KRAS*, *MET*, *MLH1*, *MPL*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PIK3CA*, *PTEN*, *PTPN11*, *RB1*, *RET*, *SMAD4*, *SMARCB1*, *STK11*, *TP53*, and *VHL*.

OncoScreen ST2.0 is also a 50-gene hot-spot solid tumor panel, which uses the Ion Ampliseq Cancer Hotspot Panel V2 primer set[4] (Thermo Fisher Scientific, Waltham, MA) for amplification of 207 hot-spot targeted amplicons across 50 genes, including all genes in the ST1.0 assay plus five additional genes. The clinical assay can be run in singlet or replicate mode, with an assay input of only 5 ng FFPE DNA per genomic DNA real-time quantitative PCR assessment (Kapa Biosystems). In addition to the genes of OncoScreen ST2.0, it includes the following genes: *ABL1*, *AKT1*, *ALK*, *APC*, *ATM*, *BRAF*, *CDH1*, *CDKN2A*, *CSF1R*, *CTNNB1*, *EGFR*, *ERBB2*, *ERBB4*, *EZH2*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *GNA11*, *GNAQ*, *GNAS*, *HNF1A*, *HRAS*, *IDH1*, *IDH2*, *JAK2*, *JAK3*, *KDR*, *KIT*, *KRAS*, *MET*, *MLH1*, *MPL*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PIK3CA*, *PTEN*, *PTPN11*, *RB1*, *RET*, *SMAD4*, *SMARCB1*, *SMO*, *SRC*, *STK11*, *TP53*, and *VHL*.

OncoHeme is a 53-gene hematological malignancy panel, with custom-designed primers for amplification of 1346 amplicons from blood/bone marrow DNA in three pools targeting full or near-full coding territory. The genes in the panel are as follows: *ABL1*, *ASXL1*, *ATRX*, *BCOR*, *BCORL1*, *BRAF*, *CALR*, *CBL*, *CBLB*, *CBLC*, *CDKN2A*, *CSF3R*, *CUX1*, *DNMT3A*, *ETV6*, *EZH2*, *FBXW7*, *FLT3*, *GATA1*, *GATA2*, *GNAS*, *HRAS*, *IDH1*, *IDH2*, *IKZF1*, *JAK2*, *JAK3*, *KDM6A*, *KIT*, *KRAS*, *MLL*, *MPL*, *MYD88*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PHF6*, *PTEN*, *PTPN11*, *RAD21*, *RUNX1*, *SETBP1*, *SF3B1*, *SMC1A*, *SMC3*, *SRSF2*, *STAG2*, *TET2*, *TP53*, *U2AF1*, *WT1*, and *ZRSR2*.

For the ST2.0 and OncoHeme assays, multiplex PCR was performed using the Ion Ampliseq Library Kit 2.0 (Thermo Fisher Scientific). PCR products were analyzed for amplicon size using a TapeStation system (Agilent Technologies, Santa Clara, CA) and for yield using a Qubit fluorometer and quantitation reagents (Life Technologies). The amplicon products for these assays were subjected to library preparation with patient-indexed adapters using the KAPA HTP Library Preparation Kit (Kapa Biosystems). Final libraries were quantitated using the KAPA Library Quantification Kit (Kapa Biosystems). Sequencing was performed on MiSeq instruments with version 2 reagents (Illumina), using 2 × 152 bp sequencing (OncoScreen ST1.0 and ST2.0) or 2 × 255 bp sequencing (OncoHeme) with a single 8-bp index read.

Hybrid Capture–Based UCM-OncoPlus Assay

After quantitation, 100 ng Qubit-quantified DNA (blood or bone marrow) or Qubit/real-time quantitative PCR–quantified DNA (FFPE) was sheared using the Covaris S2 (Covaris, Woburn, MA) and subjected to library preparation and amplification using the KAPA HTP Library Preparation Kit (Kapa Biosystems), using adapters with patient-specific indexes (Roche, Indianapolis, IN). Libraries were quantified using Qubit reagents and the KAPA Library Quantification Kit (Kapa Biosystems). Pooled libraries were captured using a custom-designed SeqCap EZ capture panel targeting 1212 genes (Roche), supplemented with select xGen Lockdown Probes (IDT, Coralville, IA). After subsequent PCR amplification and real-time quantitative PCR library quantification (Kapa Biosystems), pooled captured libraries were sequenced on Illumina HiSeq instruments with rapid run version 2 reagents, using 2 × 101 bp sequencing with a single 7-bp index read (Illumina). Of the 81 genes covered by the UCM amplicon assays listed above, 76 were selected. These genes are as follows: *ABL1*, *AKT1*, *ALK*, *APC*, *ASXL1*, *ATM*, *ATRX*, *BCOR*, *BCORL1*, *BRAF*, *CALR*, *CBL*, *CBLB*, *CDH1*, *CDKN2A*, *CSF1R*, *CSF3R*, *CTNNB1*, *CUX1*, *DNMT3A*, *EGFR*, *ERBB2*, *ERBB4*, *ETV6*, *EZH2*, *FBXW7*, *FGFR1*, *FGFR2*, *FGFR3*, *FLT3*, *GATA1*, *GATA2*, *GNA11*, *GNAQ*, *GNAS*, *HRAS*, *IDH1*, *IDH2*, *IKZF1*, *JAK2*, *KDM6A*, *KDR*, *KIT*, *KMT2A*, *KRAS*, *MET*, *MPL*, *MYD88*, *NOTCH1*, *NPM1*, *NRAS*, *PDGFRA*, *PHF6*, *PIK3CA*, *PTEN*, *PTPN11*, *RAD21*, *RB1*, *RET*, *RUNX1*, *SETBP1*, *SF3B1*, *SMAD4*, *SMARCB1*, *SMC1A*, *SMC3*, *SMO*, *SRSF2*, *STAG2*, *STK11*, *TET2*, *TP53*, *U2AF1*, *VHL*, *WT1*, and *ZRSR2*.

Bioinformatics Pipelines for Data Analysis

After sequencing, data output directories were transferred to a Health Insurance Portability and Accountability Act–protected high-performance computing system within the University of Chicago Center for Research Informatics for

further processing. We developed custom pipelines to analyze sequence data from all four UCM panels, using a combination of publicly available software and internally developed tools. Schematics of the bioinformatics pipelines for the amplicon and hybrid capture assays are shown in Supplemental Figure S1.

Briefly, amplicon assay data were processed using custom quality filter and primer-matching scripts, after which the pipeline was split into two branches: alignment based and reference independent. In the first branch, the data were aligned to the hg19 reference human genome using Novoalign 3.02.07 (NovoCraft, Selangor, Malaysia) using automatic adapter and primer trimming options. Variant calling was then performed using a combination of Samtools 0.1.19[15] mpileup and Variant Inspector version 1.0, a UCM-developed variant calling software. Variants are filtered for quality [Phred quality score >30 (>Q30)] and depth (>50× for Oncoscreen and >100× for OncoHeme) with variant signal >5% retained for analysis. In the second branch, the filtered fastq files were analyzed for the presence of insertion and deletion mutations >5 bp using the reference-independent Amplicon Indel Hunter.[16] Variant calls were combined and annotated to Human Genome Variation Society nomenclature using Alamut Batch software version 1.3.0 (Interactive Biosoftware, Rouen, France). The flowchart is shown in Supplemental Figure S1A.

For UCM-OncoPlus analysis, fastq files were subjected to Illumina adapter sequence trimming using Trimmomatic 0.30,[17] followed by alignment to the hg19 human reference genome using BWA-MEM version 0.7.12.[18] After application of a custom mapping quality filter to remove ambiguously mapped reads, assembly-based insertion and deletion (indel) realignment was performed using Abra.[19] After PCR duplicates were removed using Picard tools 1.92, a combination of Samtools 0.1.19 mpileup and Variant Inspector version 1.0 were used to identify variants within the clinical territory. Variants were filtered based on depth (>100×), quality (>Q30), MAF (>5%), and location in clinical exonic territory for review. Variant calls were annotated and converted to Human Genome Variation Society nomenclature using Alamut Batch version 1.3.0 software. A separate module was designed to retain high sensitivity for *FLT3* internal tandem duplication (ITD) mutations to include separate analysis of the *FLT3* exon 13 to 15 region using Pindel version 0.2.4t[20] and a custom UCM algorithm (ITD Hunter). The flowchart is shown in Supplemental Figure S1B.

Both assay pipelines have quality control modules to calculate quality control statistics on raw fastq data, on-target amplification, alignments, and depth statistics. Pipelines in our laboratory are all subjected to preliminary testing with nontumor specimens to determine sites of background signal and potential pseudogene-affected regions. These problematic regions are either excluded or, if possible, addressed via a variety of pipeline filters and custom solutions intended to prevent false-positive variant calls. This process is conducted before subsequent validation testing.

Fusion detection was performed using UCM internally developed software (UCM Fusion Detector version 1.0) that operates by using a combination of read pairs either aligned via BWA MEM coordinately or trimmed to 40 bp and aligned separately to detect read pairs that map on either side of a fusion junction within *ALK*, *RET*, and *ROS1*. The software considers gene direction compatibility and summarizes results across gene combinations that are unique and focused within narrow windows (500 bp) within prospective distal gene partners. Signal for each potential fusion is calculated as the maximum number of unique reads within any window in the distal partner gene divided by the total number of ALK, RET, and ROS1 mapped reads.

Copy number analysis involved evaluation of average exon interval depths recorded via the Genome Analysis Tool Kit DepthofCoverage module.[21] A historical normalized baseline for each interval in the panel was generated using the 24 nonmalignant clinical samples. Test sample data were subjected to a normalization algorithm to control for individual gene profile run-specific variability. To detect the potential copy number regions, fold change and Z-scores were calculated for each interval, and thresholds were set at >200% (gain) or <66% (loss) with Z-score >3 or <−2, respectively. Genes with more than half the intervals showing copy number changes in the same direction were then identified.

FLT3 ITD Mutation Detection

The UCM-OncoPlus data analysis pipeline has a separate module for *FLT3* ITD detection. All insertions from the alignment-based branch of the pipeline, which involves BWA-MEM, Abra, Samtools, and Variant Inspector, are first extracted for the *FLT3* region chromosome 13:28607774:28608774. Second, Pindel[20] is run for the same region on the Abra indel-realigned file, and any indels at >3% MAF are identified. Last, we use ITD Hunter, a UCM-developed tool (S. Kadri, J.P. Segal, unpublished data), which scans the region for reads with valid inserts (contained within the reads) or misaligned reads with trailing insertions. This is intended to boost sensitivity for large ITDs, especially when the aligner and indel-realigner fail to accurately align the ITD located at the end of the reads, leading to many alignments with trailing insertions of various sizes (Supplemental Figure S2).

Results

UCM-OncoPlus Tiered Assay Design

The foundation of the UCM-OncoPlus assay is Roche Nimblegen SeqCap EZ custom capture design targeting 6,064,966 bp of genomic sequence covering 1212 human genes relevant to cancer in both the somatic and inherited contexts. To facilitate testing of the more clinically relevant genes, the UCM-OncoPlus capture panel was designed in two tiers, a first tier with 316 genes deemed to be of higher

clinical relevance and a second tier with 896 genes considered more relevant for translational cancer research (Supplemental Table S2). The tiers cover 1.6 and 2.8 million coding bp, respectively, as well as padding bases to ensure depth across exons. However, the SeqCap EZ synthesis design for this panel includes tier 1 probes at a 3× concentration relative to tier 2 probes to produce higher depth for clinically important regions and allow increased multiplexing of samples for clinical analysis (Figure 1A). The panel includes full exon coverage for all genes, with coverage of 5′ and 3′ untranslated regions for the tier 1 gene set. The panel also includes intronic and/or upstream coverage for 18 genes for detection of select gene fusions (eg, *ALK*, *RET*, and *ROS1*), intragenic rearrangements (eg, *MET*, *EGFR*), and promoter mutations (eg, *TERT*), as well as probes to target select cancer-associated viruses. Certain clinically relevant regions that initially showed poor performance because of inefficient capture were supplemented by the addition of biotinylated xGen Lockdown probes (Integrated DNA Technologies, Coralville, IA) to boost capture efficiency.

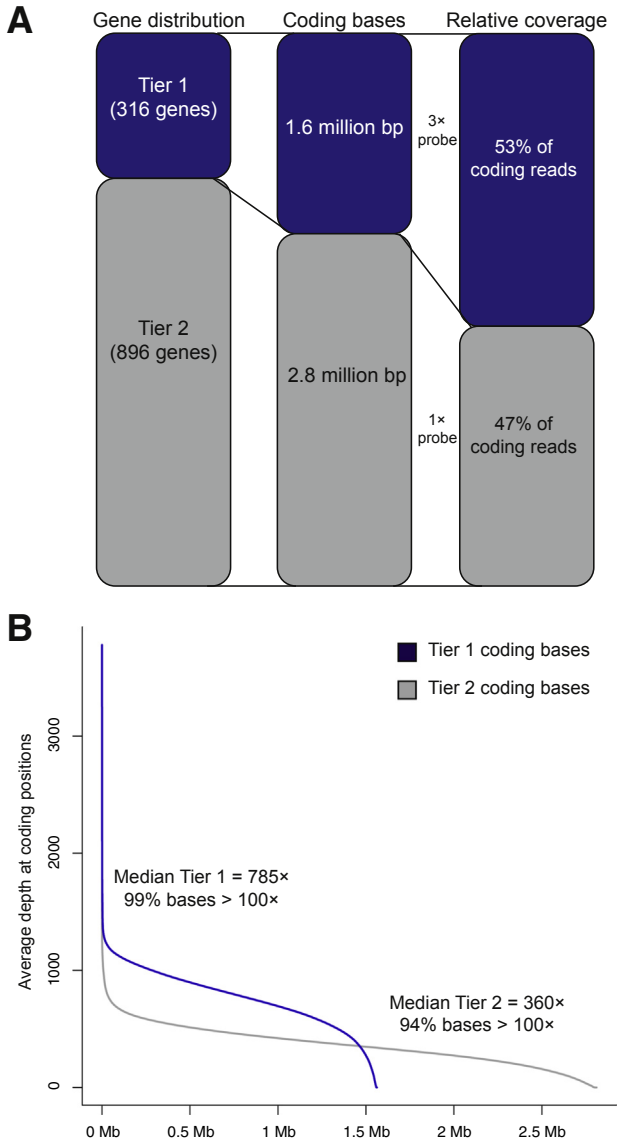


Figure 1 University of Chicago Medicine OncoPlus tiered design: **A:** The capture panel genes are split in two tiers: tier 1 has 316 genes of high clinical significance, 3× more probe than tier 2 (896 genes), and approximately 53% of the coding reads. **B:** The sorted average depth per position is shown for each tier.

Multiplexing at a level of nine samples per HiSeq 2500 rapid run flow cell yields an average of 25 to 35 million read pairs per sample. Approximately 99% of all reads map successfully to the genome, and an average of 80% are deemed on-target via mapping to within 200 bp of a probe target. As with all capture panels, a distribution of depth is noted. However, the tiered design proves to deliver a substantial boost to depth in the tier 1 target region, with approximately 53% of all coding bases belonging to tier 1 (Figure 1A), and with median depths of 785× and 360× in tier 1 and tier 2 coding regions, respectively, using nontumor (diploid control) FFPE specimens (Figure 1B). Overall, 99% of target coding bases in tier 1 genes were covered by >100× unique reads, compared with 94% of tier 2 coding bases.

Specimens for Clinical Validation

We assessed the performance of the assay using a total set of 114 validation samples, including 24 nonmalignant (8 blood/bone marrow and 16 FFPE spleen) and 80 malignant blood, bone marrow, and FFPE validation samples previously tested on multiple CLIA platforms (including 64 tested via NGS assays), and an additional 10 nonmalignant (7 blood and 3 FFPE) samples not previously run on any comparative assay (Supplemental Table S1). All malignant specimens were selected because of the presence of desired somatic mutations (Supplemental Table S3). The comparator assays included the UCM Clinical Genomics Laboratory OncoHeme and OncoScreen ST1.0 and ST2.0 assays, as well as Foundation One from Foundation Medicine (FM). The UCM assays are all amplicon-based NGS profiling panels, using either Illumina Custom Amplicon technology (OncoScreen ST1.0) or a customized Ion Torrent multiplex PCR preparation (OncoScreen ST2.0 and OncoHeme), with sequencing on the Illumina MiSeq system, whereas the FM assay is a capture-based assay more similar to UCM-OncoPlus. Of these 80 malignant samples, 64 (54 tested using UCM assays and 10 at FM) were used for testing variant concordance, and 13 were a separate set of lung cancer samples from an outside reference laboratory positive for a variety of ALK/RET/ROS1 fusions. The total set of malignant samples also included three *HER2*-amplified samples tested via fluorescence *in situ* hybridization for copy number analysis. The 24 nonmalignant samples (16 FFPE spleen and 8 blood/marrow) were all CLIA-tested via one of the UCM amplicon assays. Sample details, including diagnoses and prior analysis summaries, are shown in Supplemental Table S1.

Panel Overlaps for Validation Assessment

Of the 81 genes collectively analyzed via our solid tumor and heme malignancy amplicon panels, we selected 76 for clinical validation as part of the UCM-OncoPlus panel (*Materials and Methods*). Five genes were excluded because of inadequate capture efficiency and/or low clinical relevance. The overlap between the panels (genes, intervals, exon, and bp territory) is shown in Supplemental Table S4. Unlike the results for samples sent to FM, we retain all base call information for each sample analyzed with UCM amplicon NGS panels. Thus, each sample assayed by UCM-OncoPlus and OncoScreen ST1.0 or ST2.0 or OncoHeme allowed for the intercomparison of variant calling metrics at 16,452, 17,592, or 147,855 unique genomic positions, respectively. In combination, the total overlap region of the panels allowed comparison of assay performance across a total of 163,740 unique bases of genomic territory in 874 exons of the 76 genes.

Sensitivity Analysis

A total of 88 validation samples (78 samples CLIA-tested at UCM and 10 tumor samples sent to FM) CLIA-tested via clinical NGS were used to test the variant-calling sensitivity of the UCM-OncoPlus assay. Of the 78 samples, 54 were malignant, whereas the remaining were nonmalignant samples (Supplemental Table S1).

The 10 samples compared against FM included a total of 17 reported pathogenic findings within the OncoPlus validation target region (in 9 of the 10 samples, the tenth sample harboring only an *ALK* fusion). All 17 were detected by UCM-OncoPlus at a range of MAFs from 5.4% to 95.3% (Table 1), although MAF data from FM were unavailable for comparison. Exact genomic mutation coordinates were not available to evaluate concordance (reports contained protein nomenclature only), but we were able to assess success via comparisons with the Human Genome Variation Society protein effect nomenclature output.

In contrast to the low number of variants available from the reference laboratory NGS comparisons, the 78 validation samples tested via UCM amplicon provided a set of 2713 expected variants for sensitivity studies, including all 136 pathogenic variant calls arising at a variety of MAFs. All other positions were documented to be negative for variation; thus, all other data points were usable for specificity analysis (below). The 2713 variants originally observed at MAF >10% via amplicon assays (the analytical sensitivity of the UCM-OncoPlus assay) across panel overlap territory included 2622 single-nucleotide variants and 91 indel mutations. Observed deletions ranged from 1 to 52 bp [the largest in exon 9 of *calreticulin* (*CALR*)], whereas insertions ranged from 1 to 102 bp, including 11 ITD type mutations of the *FLT3* gene of sizes 27 to 102 bp. *FLT3* ITD detection by NGS is regarded as challenging,[22] and for this reason the informatics pipeline included special modules design to detect them (*Large Indel Detection*). Of the 2713 expected genomic mutations, all except two were readily detected.

We investigated these two discordant mutation calls further. Both missed calls were within the location chromosome X:39,932,800 to 39,932,810 in one heme malignancy specimen (UC14). Interestingly, UCM-OncoPlus identified three different and unexpected mutations within the same small region. On further examination, this represented a complex *BCOR* mutation aligned in two different but equally correct ways by Novoalign (OncoHeme) versus BWA MEM (UCM-OncoPlus). Thus, the individual component variant

Table 1 Variant Concordance between 17 Pathogenic Findings across 9 of 10 Samples Sent to FM

Sample	FM gene	FM HGVS	OncoPlus gene	OncoPlus HGVS	OncoPlus MAF, %	Concordance
VAL1	*TP53*	R175H	*TP53*	R175H	67.70	Yes
VAL2	*FBXW7*	H379fs	*FBXW7*	H379fs	86.70	Yes
VAL2	*TP53*	G245S	*TP53*	G245S	95.30	Yes
VAL2	*APC*	T1556fs	*APC*	T1556fs	81.20	Yes
VAL2	*SMAD4*	D415fs	*SMAD4*	D415fs	90.70	Yes
VAL3	*TP53*	Q38*	*TP53*	Q38*	8.70	Yes
VAL4	*TP53*	H179Y	*TP53*	H179Y	59.40	Yes
VAL4	*DNMT3A*	R771*	*DNMT3A*	R771*	5.40	Yes
VAL4	*FLT3*	R773fs	*FLT3*	R773fs	25.50	Yes
VAL5	*TP53*	R248Q	*TP53*	R248Q	29.20	Yes
VAL5	*SMAD4*	Q534*	*SMAD4*	Q534*	21.70	Yes
VAL6	*TP53*	G245S	*TP53*	G245S	19.90	Yes
VAL7	*TP53*	S215R	*TP53*	S215R	63.90	Yes
VAL7	*PIK3CA*	H1047R	*PIK3CA*	H1047R	76.90	Yes
VAL8	*GNAS*	Q227K	*GNAS*	Q227K	35	Yes
VAL8	*KRAS*	Q61H	*KRAS*	Q61H	15.20	Yes
VAL9	*TP53*	A307fs	*TP53*	A307fs	54.80	Yes

Three mutations in *NFE2L2* and *BRCA2* are not included in the University of Chicago Medicine OncoPlus clinical territory and, thus, are not shown in the table.
FM, Foundation Medicine; HGVS, Human Genome Variation Society; MAF, mutant allele frequency.

calls of this complex variant were different, but the final Human Genome Variation Society nomenclature for the combined coding and protein changes was the same (Supplemental Figure S3), indicating overall concordance despite the initial discrepancy. This tendency for different aligners to produce different but equivalent results for the same complex mutations is underappreciated, and highlights the need to effectively detect, review, assemble, and annotate complex variants during clinical variant interpretation.[23,24]

In summary, after merging this complex variant into a composite complex call, UCM-OncoPlus identified 2729 of 2729 expected variant calls (100% sensitivity) across 88 malignant and nonmalignant FFPE and blood/bone marrow clinical specimens (including samples with at least 15 different malignant diagnoses), compared against four separate comparator NGS assays.

MAF Concordance Analysis

Unlike other validation approaches, retention of MAF information from UCM amplicon assays allowed for a detailed assessment of UCM-OncoPlus variant calling efficiency. Across the variants detected by UCM-OncoPlus >10% MAF, extremely high MAF concordance ($R^2 = 0.986$) was observed between the capture-based UCM-OncoPlus assay and the amplicon panels (Figure 2A). In addition to the complex *GNAS* deletion described above, a small number of points that were distant from the regression line were further investigated. Two variants showed reduced detection efficiency by UCM-OncoPlus (Figure 2A). For each variant, examination of the OncoHeme data showed an adjacent tiled amplicon positive for a nearby single-nucleotide polymorphism (SNP) present in *trans* that overlapped with the primer-binding site of the amplicon containing the variant of interest. Thus, the variance can be explained by biased OncoHeme primer binding and amplification favoring the mutant. In *SRSF2*, two somatic variants at the hot-spot position amino acid 95 showed higher MAFs in UCM-OncoPlus versus OncoHeme (Figure 2A). In this case, it is not readily evident which platform is responsible for the discrepancy. No surrounding SNPs were identified using UCM-OncoPlus that could potentially explain the higher performance of UCM-OncoPlus for these variants on the basis of primer binding site interference. The region is extremely GC rich, which could have an effect on amplification or sequencing via either panel.

Limit of Detection and Reproducibility

To assess low-level mutation detection in FFPE samples in a more comprehensive way across the 76 genes, we generated a mixture of three nonmalignant FFPE samples to produce a single sample with many low MAF variants. Individual analysis showed a total of 119 variants across the 76 genes in the component samples, all at approximately 50% and approximately 100% MAF and all noted previously as inherited variants in the 1000 Genomes Project,[25] Exome Aggregation Consortium (*http://exac.broadinstitute.org*, last accessed February 2016), or Exome Variant Server (*http://evs.gs.washington.edu*, last accessed February 2016) databases. Only this set of variants was expected to be observed after 1:1:1 mixture, with 49 variants anticipated at approximately 16.66% MAF, average MAFs observed between 11% and 20%, because of fluctuations in pooling. Overall, all 119 expected variants were detected, with average MAFs ranging from 11% to 100%. Data from all 20 replicates for all 119 variants are shown in Figure 2D.

Specificity Analysis

Unlike the specimens tested via narrow targeted analysis or by reference NGS send-out (FM), each specimen examined via UCM amplicon panels yielded between approximately 16,000 and approximately 147,000 points of data for genomic coordinates previously clinically documented by CLIA as absent for any mutation. However, because the 24 nonmalignant samples were used during initial assay development to identify regions with sequencing artifacts (eg, homopolymer regions and potential pseudogene sequences) to establish reportable territory and train the assay, they were not suitable for investigations of specificity. Instead, we evaluated the results of the 54 malignant samples, focusing on genomic coordinates for each sample where no variant was previously detected via the amplicon assay (approximately 16 K for Oncoscreen and approximately 167 K for OncoHeme). As these are tumor samples likely to harbor low percentage mutational load, this represents a stringent specificity examination. Across 5.37 million genomic samplings with expected results in these samples, only one positive variant call was made by OncoPlus (Figure 2, B and C). Further investigation revealed this to be a true-positive 27-bp complex deletion previously missed by OncoHeme because of involvement of multiple primer binding sites at an amplicon tiling point (Supplemental Figure S4). Repeat PCR-based NGS analysis of this locus using alternative primers distant from the mutation site also detected the presence of the same complex deletion (data not shown), confirming this finding. Thus, based on this and the presence of these mutations in the 1000 Genomes Project,[25] we conclude that this a legitimate finding previously missed because of amplicon primer binding interference. Overall, zero false-positive mutation calls were made across a total of 5.37 million base calls, including all 347 K base positions tested in 20 malignant FFPE specimens (Figure 2B).

To further interrogate the level of background seen in FFPE samples outside of the 16 K bases common between UCM-OncoPlus and Oncoscreen, we examined the data from 20 replicates of our 1:1:1 nonmalignant FFPE mixed control sample. All positive results seen were the previously identified 119 SNPs that showed approximately the expected MAF alterations predicted from admixture. Across all replicates ($n = 20$, 3.27 million base samplings), zero

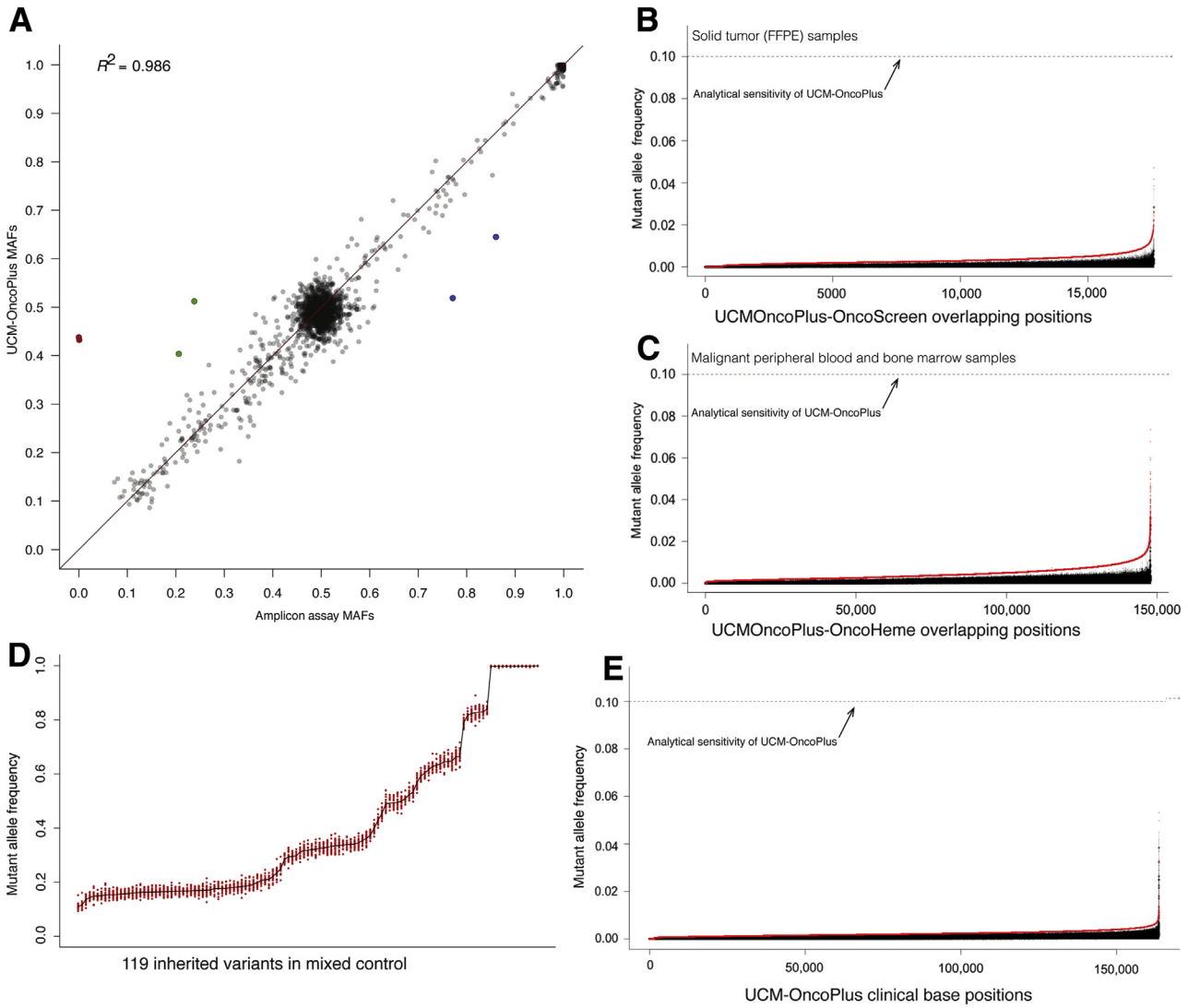


Figure 2 **A:** Mutant allele frequency (MAF) concordance: Scatter plot of University of Chicago Medicine (UCM) OncoPlus versus UCM amplicon assay MAFs for all variants across 78 clinical (54 malignant and 24 nonmalignant) samples, showing >10% signal via either assay. The **red points** highlight two additional variants detected by UCM-OncoPlus not previously detected by OncoHeme. The **blue points** represent two mutations with higher amplicon MAFs, whereas the **green points** represent two *SRSF2* mutations with higher MAFs in UCM-OncoPlus. **B** and **C:** Specificity of UCM-OncoPlus in malignant samples: the average MAFs of all non–single-nucleotide polymorphism (SNP) positions across all solid tumor (**B**) and malignant peripheral blood or bone marrow (**C**) samples. The **vertical lines** show the SDs, whereas the **red points** show the maximum MAF at each position across all compared samples. **D:** MAFs of 119 SNP variants in the mixed control sample across 20 sequencing runs. Each MAF is plotted as a **red point**. The **black line** represents the average across replicates. **E:** Average and SDs of MAFs at each non-SNP position in the mixed normal control across replicates. The **red points** show the maximum signal at each position across all replicates. $n = 20$ (**B** and **D** and **E**, replicates); $n = 34$ (**C**). FFPE, formalin fixed, paraffin embedded.

unexpected variants were detected at or above 10% MAF. Across the entire assay, the mean identified variant signal was 0.05%. Means, SDs, and maximum signals identified at every position are shown in Figure 2E. Overall, specificity was 100% across 8.64 million assessed genomic samplings (Figure 2, B, C, and E). A similar analysis of nontumor FFPE ($n = 3$) and nonmalignant blood sequenced independently using UCM-OncoPlus was performed to investigate the specificity at all non-SNP positions. SNP positions were determined as before using a combination of 1000 Genome, Exome Aggregation Consortium, and Exome Variant Server databases. This analysis further allowed us to ascertain the level of background noise inherent to the assay. Across 163,717 base positions, we observed a low mean variant frequency of 0.047% (Supplemental Figure S5).

Large Indel Detection

Highly sensitive detection of indel mutations remains a challenge in clinical NGS diagnostics. Dramatic divergence from the reference genome sequence presents a challenge for alignment-based software, and without special care many larger indels may end up unmapped or improperly mapped, preventing accurate variant calling. We addressed this issue

in depth during the validation of our amplicon-based panels, and have previously reported a novel tool, Amplicon Indel Hunter,[20] specific for indel detection in amplicon NGS data that does not rely on reference genome alignment. As a result of using this highly sensitive large indel calling system as part of our amplicon panel informatics, the inter-NGS validation sample set was able to include 93 clinically validated indel calls in 48 specimens of the 102 specimens tested, with insertions seen up to 102 bp and deletions up to 52 bp.

Matching this performance represented a substantial challenge from the standpoint of capture panel informatics pipeline development. Random fragmentation-based capture panel data require alternative tools, of which many have been described. We used Abra, a localized de novo realignment tool, to act as the main indel detection platform for the pipeline. In comparison studies, Abra produced similar results to the indel-detection algorithm, Pindel,[18] for most indels, but is advantageous because after de novo realignment and establishment of contigs, it maps individual reads back to the reference genome, producing modified alignment files incorporating realigned reads. This allowed for unified variant calling of all anomalies (ie, single-nucleotide variants, small indels, and realigned indels) in parallel using the variant calling module of the pipeline (*Materials and Methods*). However, such realignments require that reads contain sequence footprints on either side of an indel to produce accurate variant calls, which is a limitation specifically for long insertions, especially >60 bp. In contrast, Pindel is less affected by this problem. We selected all of our most challenging clinically identified indels for this 76-gene validation set; however, in clinical practice, we have only observed *FLT3* among these genes to show insertions >60 bp. Activating ITD mutations in acute myeloid leukemia patients have been associated with a poor prognosis in many studies.[26–28] To detect these clinically important and historically difficult mutations, we supplemented the informatics pipeline with two modules specific for *FLT3* ITD detection. The first, Pindel, proved effective for most of our tested set of 11 ITD mutations, but depending on the specific ITD sequence, Pindel either aligned all junctional overlapping reads (resulting in high sensitivity) or evaluated only reads containing the entire ITD in a similar manner to Abra (lower sensitivity) (Supplemental Figure S6). The performance was dependent on Pindel's assessment of the insert (ie, short insert versus duplication), although Pindel's basis for such categorization was opaque. To maintain high sensitivity in the latter case, we developed an algorithm (ITD Hunter) to detect all insertional reads in the *FLT3* region of the Abra-realigned BAM file, regardless of whether the reads included the full insert sequence (Supplemental Figure S2). Thus, the module uses an integrated approach combining the results from Abra, Pindel, and ITD Hunter for ITD detection. This resulted in high detection efficiency/sensitivity within the *FLT3* ITD region without a significant impact on specificity. The maximum observed MAF across the three methods for all ITDs was comparable to prior amplicon-based detection with Amplicon Indel Hunter (including all *FLT3* ITDs up to 102 bp), including concordance with two additional low MAF ITD calls not included in our original sensitivity analysis, which included only variants with >10% amplicon MAF (Figure 3).

Copy Number Alterations and Gene Fusions

The ability for NGS assays to simultaneously assess both sequence and structural rearrangements represents a major advantage of the technology compared with traditional sequencing methods. Our previous amplicon assays were not validated for copy number variant detection; thus, we evaluated the UCM-OncoPlus assay using a set of 21 samples, including 7 nonmalignant blood samples, 3 samples previously tested positive for *HER2* amplification by fluorescence *in situ* hybridization, and 9 of 10 samples with a total of 36 amplification/deletion findings from FM. No false-positive gains or losses were noted in the nontumor samples (Figure 4A), and the findings from the copy number variant–containing samples are summarized in Table 2. An example of a lung cancer specimen (Val5) showing FM-validated EGFR, MET, and other amplifications is shown in Figure 4B. For each sample, all expected copy number changes were either detected via UCM-OncoPlus above the set thresholds or trended in the appropriate direction without being prominent enough to be called using our applied cutoffs. Different copy number algorithms incorporate

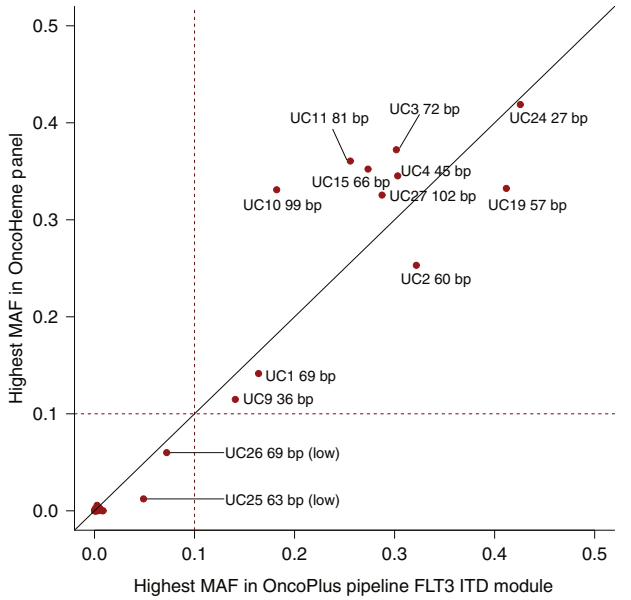


Figure 3 *FLT3* internal tandem duplication (ITD) mutant allele frequency (MAF) comparison: University of Chicago Medicine (UCM) OncoPlus versus OncoHeme *FLT3* ITD MAFs for 42 samples (34 malignant and 8 nonmalignant), including 11 positives and 2 low positives (UC25 and UC26). The *x* axis shows the highest MAF in the UCM-OncoPlus *FLT3* ITD module across BWA MEM/Abra, Pindel, and ITD Hunter, whereas the *y* axis shows the highest OncoHeme MAFs among alignment-based and Amplicon Indel Hunter results for each sample. **Dotted lines** indicate analytical sensitivity of UCM-OncoPlus.

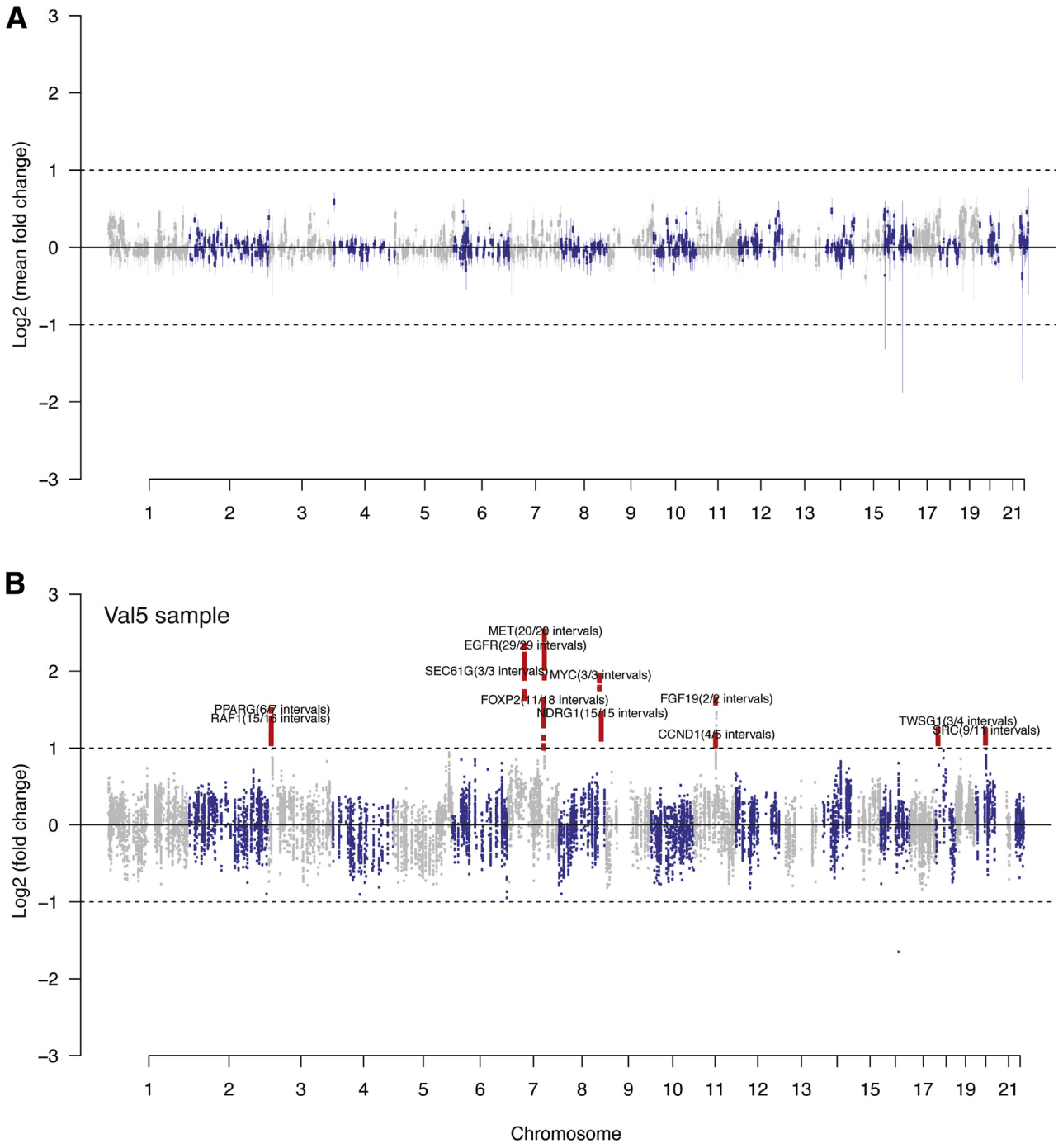


Figure 4 Copy number alterations: **A:** Log2 average fold change across seven independent nonmalignant blood samples using the baseline historical normal. No false-positive changes were detected as expected. **B:** Log2 of fold changes in a Foundation Medicine sample (VAL5) using a baseline historical normal. The analysis shows copy number changes in the expected genes, *EGFR*, *MET*, *RAF1*, *CCND1*, *MYC*, *FGF3*, and *FGF19*, in addition to few other genes on our panel. **Dotted lines** indicate the fold change cutoffs.

different normalization methods and cutoff values, making assessments of interlaboratory CNV concordance more problematic than for sequence variants. UCM-OncoPlus uses similar thresholds (greater than twofold change) to other published assays.[5]

The UCM-OncoPlus panel is constructed to include tiling of introns in a total of 16 genes showing recurrent somatic rearrangements. Our initial efforts toward progressive validation of these anomalies involved testing a panel of 15 lung cancer samples harboring a variety of *ALK*, *RET*, and *ROS1* gene fusions. These were previously confirmed either in our laboratory by fluorescence *in situ* hybridization ($n = 2$) or by an outside reference laboratory using a combination of fluorescence *in situ* hybridization and NGS

Table 2 Copy Number Changes Reported in FM Samples and Their Concordance with the Copy Number Analysis Using UCM-OncoPlus

Sample	Gene	Loss/gain	Comparator assay	Concordance	Average fold change	In clinical territory
UC48	*ERBB2* (*HER2*)*	Gain	FM	Yes	5.7612	Yes
VAL3	*CCND1*	Gain	FM	Yes	2.306521	No
	FGF3	Gain		Yes	2.216868	No
	FGF4	Gain		No	1.746214	No
	FGF19	Gain		No	2.506074	No
VAL1	*ERBB2* (*HER2*)*	Gain	FM	Yes	18.20375	Yes
	CCND3	Gain		Yes	9.287972	No
	MYC	Gain		Yes	6.506277	No
VAL4	*CCND1*	Gain	FM	Yes	8.389682	No
	FGF19	Gain		Yes	11.00565	No
	FGF3	Gain		Yes	10.90079	No
	FGF4	Gain		Yes	8.357692	No
VAL9	*CDKN2A**	Loss	FM	No†	0.7754569	Yes
	CDKN2B	Loss		Yes	0.6198064	No
	CCND1	Gain		Yes	3.74403	No
	FGF19	Gain		Yes	4.370285	No
	FGF4	Gain		Yes	3.383176	No
	FGF3	Gain		Yes	4.132068	No
	NKX2-1	Gain		Yes	11.40722	No
VAL2	*SRC*	Gain	FM	Yes	6.113183	No
	TSC1	Loss		Yes	0.0502757	No
VAL7	*FGFR1**	Gain	FM	Yes	8.527815	Yes
	RICTOR	Gain		Yes	2.515251	No
	FGF10	Gain		Yes	2.823522	No
	ZNF703	Gain		Yes	11.86782136	No
VAL8	*CDK4*	Gain	FM	Yes	6.338542	No
	GLI1	Gain		NA	Not in panel	No
	MDM2	Gain		No	1.884392	No
	ZNF217	Gain		Yes	6.963416	No
VAL5	*EGFR**	Gain	FM	Yes	4.281734	Yes
	*MET**	Gain		Yes	4.757686	Yes
	RAF1	Gain		Yes	2.37489	No
	CCND1	Gain		Yes	2.180732	No
	MYC	Gain		Yes	3.587087	No
	FGF3	Gain		Yes	2.567123	No
	FGF4	Gain		No	2.150498	No
	FGF19	Gain		Yes	2.615929	No
VAL10	*CDKN2A**	Loss	FM	Yes	0.472469475	Yes
	CDKN2B	Loss		Yes	0.54852575	No
UC55	*ERBB2* (*HER2*)*	Gain	FISH	Yes	4.692491	Yes
UC56	*ERBB2* (*HER2*)*	Gain	FISH	Yes	9.324157	Yes
UC57	*ERBB2* (*HER2*)*	Gain	FISH	Yes	23.09234	Yes

*Included in the UCM-OncoPlus clinical territory.

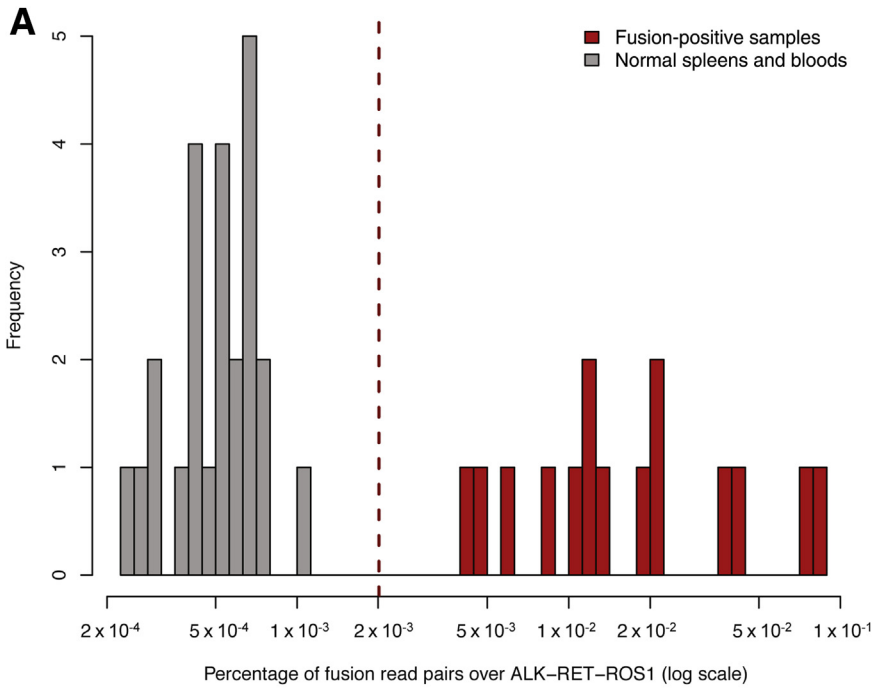
†Not detected by set thresholds but showed loss across all exons.

FISH, fluorescence *in situ* hybridization; FM, Foundation Medicine; NA, not available; UCM, University of Chicago Medicine.

methods ($n = 13$). For fusion detection, we applied a novel algorithm incorporating a combination of paired-end read alignments as well as alignments of individual trimmed read mates to boost sensitivity for unambiguous intergenic read pair mapping. Using this algorithm to look for fusions specifically in *ALK*, *RET*, and *ROS1*, we identified all 15 fusions, including the correct partner genes, with low fusion signal detected in the 24 nonmalignant samples (Figure 5A). Figure 5B shows an example of a *KIF5B-RET* fusion where the highlighted reads were identified by the fusion algorithm to detect the fusion. The *RET* alignments show the reads resulting from the tiling design in the intron of the gene.

Discussion

Genomics technologies are continuing to transform the field of oncology diagnostics, with many laboratories working to develop NGS LDPs to support personalized medicine and institutional translational research. However, validating these



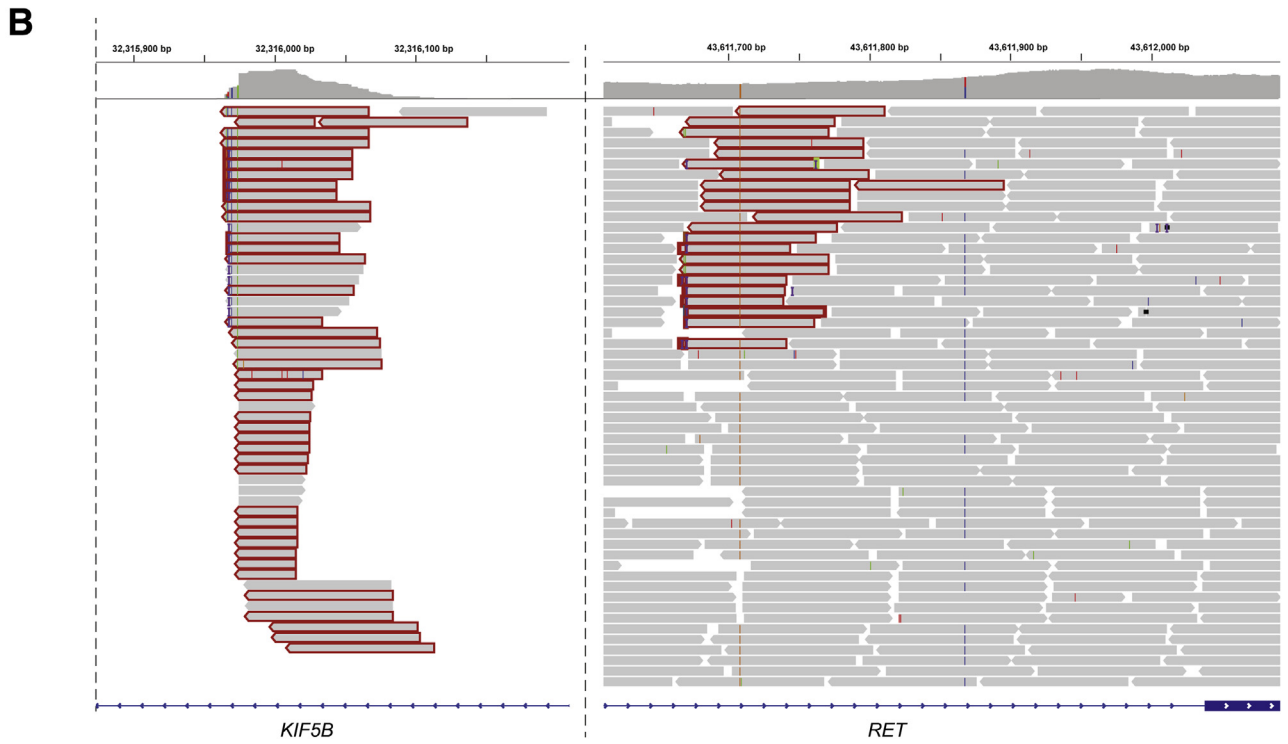


Figure 5 Gene fusions: **A:** Histogram of percentage of fusion read pairs identified via the University of Chicago Medicine fusion algorithm across all read pairs mapped to *ALK*, *RET*, and *ROS1*. There is a clear distinction between the 16 fusion-positive and the 24 nonmalignant clinical samples. The *x* axis is in log10 scale. **B:** An example of *KIF5B* and *RET* fusion reads shown in Integrated Genomics Browser.[29] The highlighted reads are paired-end mates that map across the fusion junction.

assays to satisfy CLIA regulations pertaining to non−Food and Drug Administration−cleared LDPs remains a significant challenge. In this validation, we focused on the use of true patient-derived oncology specimens previously tested with orthogonal clinical NGS assays to maximize efficiency. Using fewer samples than other large-scale validations, we were able to interrogate UCM-OncoPlus via >8.6 million comparisons with CLIA-certified results, including 2729 confirmed variant calls. Although it is true that many of these variant calls represented inherited SNPs, the most important thing during technical validation is variant concordance, whether the variants are inherited or somatic. In cancer samples, inherited SNPs and somatic mutations can both be found at a wide range of MAFs because of copy number abnormalities (and in heme samples because of donor/recipient chimerism). To augment our sample set for low MAF variants, we also conducted 20 replicates of a mixed normal FFPE specimen, confirming

the reproducible detection of 52 variant calls <20% MAF. This mixed control specimen is now routinely used in our laboratory as an in-assay control, assessed for the proper list of 119 clinical variants on each assay run. We were also simultaneously able to test the limits of the assay's detection efficiency for large indels across many genes and perform a detailed evaluation of variant calling efficiency via MAF intercomparison, analyses that would be difficult or impossible using cell lines. As a result of this validation study, the UCM-OncoPlus assay is currently live in our laboratory for clinical reporting of mutations at MAF >10% in the 76 genes analyzed, including single-nucleotide variants, generic indels <52 bp, and *FLT3* ITDs <102 bp. Data from the nonreported 1136 genes are accessible to our clinicians and investigators with patient consent and University of Chicago institutional review board approval.

It is our intention to constantly work to expand the clinically reported region of the panel to meet current and future clinical needs. Ongoing validation studies are focused on expanding the validated gene set and finalizing additional features, including copy number changes and common lung cancer gene fusions. Ideally, future genes of interest will already be present within the panel, such that ongoing validation will be facilitated by prior sequencing experience. If not, they can be added to future versions of the design or included via spike-in of custom-manufactured probes.

The extremely high concordance observed between these LDPs stands in contrast to previous reports documenting a degree of discordance between amplicon and hybrid capture assays, or suggesting the general superiority of one or the other of these methods.[12] In our experience, both assay types can be made robust and reproducible with strict attention to library preparation methods, informatics algorithms, and above all quality control parameters. Regardless of the method, the most critical experimental consideration in NGS oncology sequencing assays is ensuring sufficient sampling of original DNA molecules to prevent false-positive polymerase-induced errors or stochastic false-negative errors (typically affecting low MAF variants). This can be a substantial challenge because of the highly variable quantity and quality of DNA yielded by cancer specimens. We have found careful assay loading of FFPE genomic DNA based on amplifiable concentration (determined by real-time quantitative PCR) to be an invaluable process step for improving the robustness of all of our oncology sequencing assays.

In its first iteration, the UCM-OncoPlus clinical assay involves sequencing of only tumor tissue, with reporting of variants in the context of the tumor specimen. For a small set of reported genes, tumor-only sequencing remains the standard of care, and has a variety of advantages compared with default tumor-normal sequencing. Although it is true that tumor-normal sequencing does allow for discrimination of somatic/inherited alterations, this is not typically the most important consideration during the workup of an individual cancer patient in clinical pathology practice. For example, in a breast cancer patient, a BRCA2 mutation might raise the consideration of a poly (ADP-ribose) polymerase inhibitor trial regardless of the origin of the mutation. The determination of biological pathogenicity is of greater importance, and this is the area of greatest technical difficulty regardless of sequencing strategy. In addition to additional consenting, genetic counseling, and costs, there are practical considerations as well in that for many patients it can be difficult to find a suitable normal tissue. A blood draw may be acceptable for a solid tumor patient, but for hematopoietic malignancy patients, a skin biopsy and/or time-consuming fibroblast culture may be necessary. Informatics systems may also be a pitfall for tumor-normal assays, as computational subtraction of normal from tumor will remove any pathogenic inherited variants, and so the normal sequencing results must be separately analyzed and reported.

In clinical practice at UCM, UCM-OncoPlus allows for timely assessment of tumor mutations in genes of known relevance while at the same time producing data from >1100 additional genes for use in a variety of institutional translational research projects. Because of the tiered assay design and the incorporation of larger-scale sequencers compared with our amplicon assays, we perform this assay at a cost only 30% higher than that of our 53 gene heme panel test. The additional data can then be mined for presumptively somatic or inherited mutations and correlated with health status or patient performance. Additional sequencing of normal tissue may be conducted if clinically appropriate based on tumor findings, both in clinical practice and as a part of specific research studies. The emergence of assays like UCM-OncoPlus at multiple institutions highlights the need for improved methods for understanding the wealth of genomic data and clinical associations that they produce. This will depend on increased options for interinstitutional data sharing and aggregation. It will also require more institutions to engage in this scale of data gathering, which has historically been cost-prohibitive. We hope that the efficient assay design and validation strategies presented herein, and the emergence of additional laboratories like ours willing to share validation samples and full data sets, will be helpful in convincing other institutions to pursue similar multipurpose panels intended to support both clinical service and more basic cancer discovery.

Acknowledgments

We thank Dr. Dara Aisner (University of Colorado Molecular Correlates Laboratory) for providing ALK, RET, and ROS1 fusion-positive samples. This work would not have been possible without the vision and support of Dr. Vinay Kumar.

Supplemental Data

Supplemental material for this article can be found at *http://dx.doi.org/10.1016/j.jmoldx.2016.07.012*.

References

1. Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, et al: Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. Nat Biotechnol 2013, 31:1023−1031
2. Cottrell CE, Al-Kateb H, Bredemeyer AJ, Duncavage EJ, Spencer DH, Abel HJ, Lockwood CM, Hagemann IS, O'Guin SM, Burcea LC: Validation of a next-generation sequencing assay for clinical molecular oncology. J Mol Diagn 2014, 16:89−105
3. Pritchard CC, Salipante SJ, Koehler K, Smith C, Scroggins S, Wood B, Wu D, Lee MK, Dintzis S, Adey A: Validation and implementation of targeted capture and sequencing for the detection of actionable mutation, copy number variation, and gene rearrangement in clinical cancer specimens. J Mol Diagn 2014, 16:56−67
4. Tsongalis GJ, Peterson JD, De A, Tunkey CD, Gallagher TL, Strausbaugh LD, Wells WA, Amos CI: Routine use of the Ion Torrent AmpliSeq™ Cancer Hotspot Panel for identification of clinically actionable somatic mutations. Clin Chem Lab Med 2014, 52:707−714
5. Cheng DT, Mitchell TN, Zehir A, Shah RH, Benayed R, Syed A, Chandramohan R, Liu ZY, Won HH, Scott SN, Brannon AR, O'Reilly C, Sadowska J, Casanova J, Yannes A, Hechtman JF, Yao J, Song W, Ross DS, Oultache A, Dogan S, Borsu L, Hameed M, Nafa K, Arcila ME, Ladanyi M, Berger MF: Memorial Sloan Kettering-Integrated Mutation Profiling of Actionable Cancer Targets (MSK-IMPACT): a hybridization capture-based next-generation sequencing clinical assay for solid tumor molecular oncology. J Mol Diagn 2015, 17:251−264
6. de Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, Koudijs MJ, Nijman I, Hinrichs JWJ, Cuppen E, van Lieshout S, Loberg RD, de Jonge M, Voest EE, de Weger RA, Steeghs N, Langenberg MHG, Sleijfer S, Willems SM, Lolkema MP: Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumours using minimal DNA amounts from formalin fixed paraffin embedded material. PLoS One 2016, 11:e0149405
7. Fisher KE, Zhang L, Wang J, Smith GH, Newman S, Schneider TM, Pillai RN, Kudchadkar RR, Owonikoko TK, Ramalingam SS, Lawson DH, Delman KA, El-Rayes BF, Wilson MM, Sullivan HC, Morrison AS, Balci S, Adsay NV, Gal AA, Sica GL, Saxe DF, Mann KP, Hill CE, Khuri FR, Rossi MR: Clinical validation and implementation of a targeted next-generation sequencing assay to detect somatic variants in non-small cell lung, melanoma, and gastrointestinal malignancies. J Mol Diagn 2016, 18:299−315
8. Daber R, Sukhadia S, Morrissette JJ: Understanding the limitations of next generation sequencing informatics, an approach to clinical pipeline validation using artificial data sets. Cancer Genet 2013, 206:441−448
9. Bennett NC, Farah CS: Next-generation sequencing in clinical oncology: next steps towards clinical validation. Cancers (Basel) 2014, 6:2296−2312
10. Torsvik A, Stieber D, Enger PØ, Golebiewska A, Molven A, Svendsen A, Westermark B, Niclou SP, Olsen TK, Chekenya Enger M, Bjerkvig R: U-251 revisited: genetic drift and phenotypic consequences of long-term cultures of glioblastoma cells. Cancer Med 2014, 3:812−824
11. Hughes P, Marshall D, Reid Y, Parkes H, Gelber C: The costs of using unauthenticated, over-passaged cell lines: how much more data do we need? Biotechniques 2007, 43:575, 577−578, 581−582 passim
12. Samorodnitsky E, Jewell BM, Hagopian R, Miya J, Wing MR, Lyon E, Damodaran S, Bhatt D, Reeser JW, Datta J: Evaluation of hybridization capture versus amplicon-based methods for whole-exome sequencing. Hum Mutat 2015, 36:903−914
13. Sie D, Snijders PJF, Meijer GA, Doeleman MW, van Moorsel MIH, van Essen HF, Eijk PP, Grünberg K, van Grieken NCT, Thunnissen E, Verheul HM, Smit EF, Ylstra B, Heideman DAM: Performance of amplicon-based next generation DNA sequencing for diagnostic gene mutation profiling in oncopathology. Cell Oncol (Dordr) 2014, 37:353−361
14. Simen BB, Yin L, Goswami CP, Davis KO, Bajaj R, Gong JZ, Peiper SC, Johnson ES, Wang ZX: Validation of a next-generation-sequencing cancer panel for use in the clinical laboratory. Arch Pathol Lab Med 2015, 139:508−517
15. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009, 25:2078−2079
16. Kadri S, Zhen CJ, Wurst MN, Long BC, Jiang ZF, Wang YL, Furtado LV, Segal JP: Amplicon indel hunter is a novel bioinformatics tool to detect large somatic insertion/deletion mutations in amplicon-based next-generation sequencing data. J Mol Diagn 2015, 17:635−643
17. Bolger AM, Lohse M, Usadel B: Trimmomatic: a flexible trimmer for Illumina Sequence Data. Bioinformatics 2014, 30:2114−2120
18. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. ArXiv Prepr 2013, [Epub ahead of print] doi:arXiv:1303.3997v2
19. Mose LE, Wilkerson MD, Hayes DN, Perou CM, Parker JS: ABRA: improved coding indel detection via assembly-based realignment. Bioinformatics 2014, 30:2813−2815
20. 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
21. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M: The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010, 20:1297−1303
22. 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
23. Zook JM, Chapman B, Wang J, Mittelman D, Hofmann O, Hide W, Salit M: Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls. Nat Biotechnol 2014, 32:246−251
24. Ye K, Wang J, Jayasinghe R, Lameijer EW, McMichael JF, Ning J, McLellan MD, Xie M, Cao S, Yellapantula V, Huang KL, Scott A, Foltz S, Niu B, Johnson KJ, Moed M, Slagboom PE, Chen F, Wendl MC, Ding L: Systematic discovery of complex insertions and deletions in human cancers. Nat Med 2016, 22:97−104
25. 1000 Genomes Project Consortium, Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, Kang HM, Marth GT, McVean GA: An integrated map of genetic variation from 1,092 human genomes. Nature 2012, 491:56−65
26. 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
27. 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
28. Smith CC, Wang Q, Chin C-S, Salerno S, Damon LE, Levis MJ, Perl AE, Travers KJ, Wang S, Hunt JP, Zarrinkar PP, Schadt EE, Kasarskis A, Kuriyan J, Shah NP: Validation of ITD mutations in FLT3 as a therapeutic target in human acute myeloid leukaemia. Nature 2012, 485:260−263
29. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP: Integrative genomics viewer. Nat Biotechnol 2011, 29:24−26