An Integrated Next-Generation Sequencing System for Analyzing DNA Mutations, Gene Fusions, and RNA Expression in Lung Cancer[1]

Brian C. Haynes[*], Richard A. Blidner[*], Robyn D. Cardwell[*], Robert Zeigler[*], Shobha Gokul[*], Julie R. Thibert[*], Liangjing Chen[*], Junya Fujimoto[†], Vassiliki A. Papadimitrakopoulou[‡], Ignacio I. Wistuba[†] and Gary J. Latham[*]

[*]Asuragen, Inc., Austin, TX, USA; [†]Department of Translational Molecular Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; [‡]Department of Thoracic/Head and Neck Medical Oncology, Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Abstract

We developed and characterized a next-generation sequencing (NGS) technology for streamlined analysis of DNA and RNA using low-input, low-quality cancer specimens. A single-workflow, targeted NGS panel for non–small cell lung cancer (NSCLC) was designed covering 135 RNA and 55 DNA disease-relevant targets. This multiomic panel was used to assess 219 formalin-fixed paraffin-embedded NSCLC surgical resections and core needle biopsies. Mutations and expression phenotypes were identified consistent with previous large-scale genomic studies, including mutually exclusive DNA and RNA oncogenic driver events. Evaluation of a second cohort of low cell count fine-needle aspirate smears from the BATTLE-2 trial yielded 97% agreement with an independent, validated NGS panel that was used with matched surgical specimens. Collectively, our data indicate that broad, clinically actionable insights that previously required independent assays, workflows, and analyses to assess both DNA and RNA can be conjoined in a first-tier, highly multiplexed NGS test, thereby providing faster, simpler, and more economical results.

Translational Oncology (2019) 12, 836–845

Introduction

In the last decade, next-generation sequencing (NGS) has precipitated a paradigm shift in clinical molecular pathology from single-gene tests to multigene panels. As a technology, it has doubled as a basic research workhorse as well as a platform for routine clinical diagnostics. Research consortia such as The Cancer Genome Atlas (TCGA) have applied broad NGS profiling to catalog molecular variation in cancer, and these discoveries have been translated to clinically facing assays of prognostic and theranostic value. Routine NGS-based testing is enabling a model in which many therapeutically relevant molecular indications are simultaneously profiled and matched against an array of treatment options, thus overcoming the "one-gene/one-drug" serial testing model [1].

Clinical sequencing of tumor DNA has received the greatest attention with an emphasis on detection of hotspot single nucleotide variants (SNVs), small insertions and deletions (INDELs), and copy number variants (CNVs) that confer sensitivity to targeted therapies. For example somatic variation in exons 18-21 of *EGFR* occur in approximately 10%-15% of non–small cell lung cancer (NSCLC) tumors and are sensitizing to first-generation tyrosine kinase inhibitors (TKIs) erlotinib and gefitinib [2,3]. Tumors with innate or acquired resistance mutations are responsive to second- or third-generation inhibitors afatinib and osimertinib [4,5]. Routine profiling

of tumor DNA variation for established and emerging drug targets is now possible in clinical reference labs through validated NGS panels based on hybridization capture or targeted amplicon sequencing. While these targeted NGS technologies have largely addressed the challenge of clinical DNA-based testing, the analysis of other molecular modalities of diagnostic relevance remains unaddressed or requires disjointed workflows.

Gene fusions have emerged as an important class of markers for precision medicine in solid tumors. Transforming rearrangements of the anaplastic lymphoma kinase (ALK) gene are present in 3%-6% of lung adenocarcinomas (LUADs) [6], and these tumors are responsive to crizotinib [7]. Rearrangements of ROS1 and RET have also been found in LUADs at a prevalence of 1%-3% [8–10] and are responsive to crizotinib and multikinase inhibitors cabozantinib and vandetanib, respectively [8,11]. In addition to ALK, RET, and ROS1, fusions involving NTRK1, FGFR1/2/3, and NRG1 genes have been reported in NSCLC among other cancers and represent emerging therapeutic targets [6]. Gene fusions are detectable by immunohistochemistry (IHC) and fluorescence in situ hybridization (FISH) analysis of DNA, and this form of testing is routine in clinical reference labs.

Targeted RNA-Seq is an emerging form of testing for gene fusions with distinct advantages over IHC and FISH including sensitivity, specificity, and multiplexing density [12–15]. In contrast to NGS assays developed for SNVs, INDELs, and CNVs, targeted NGS assays developed for gene fusion detection are predominantly based on RNA-Seq. While NGS analysis of DNA can also detect chromosomal rearrangements and DNA mutations that lead to aberrant isoforms, RNA-based testing can be more sensitive, efficient, and functionally definitive considering that many DNA variants (e.g., multiple intronic breakpoints) give rise to the same oncogenic transcript. Unlike IHC, targeted RNA-Seq does not require overexpression of the 3′ fused gene, and unlike FISH, it verifies that the chimeric transcript is expressed and in-frame. In addition, targeted RNA-Seq is capable of detecting additional classes of clinically relevant RNA variation including aberrant splice variants such as the exon 14 skipped isoform of MET, which leads to a constitutively activated form of cMET that confers sensitivity to crizotinib [16]. Targeted RNA-Seq enables the quantification of gene expression markers of therapeutic value such as signatures of response to immune checkpoint blockade. While IHC testing of PD-L1 remains the entrenched patient selection tool for checkpoint inhibitors, evidence is emerging for RNA signatures with predictive accuracy superior to PD-L1 [17].

Despite the advantages of NGS in the analysis of gene fusions and other classes of RNA markers, NGS workflows for DNA and RNA markers remain largely segregated, and adoption of NGS testing for RNA has lagged DNA-based testing. One cause for this disparity is that the majority of targeted RNA-Seq and DNA-Seq assays are developed without consideration for harmonizing assay inputs, workflow, or analysis. Consequently, separate tissue slides and isolations are often required for analysis of DNA and RNA through discrete NGS procedures. While some NGS workflows have sought to address the interassay harmonization challenge [18], gaps remain with respect to separation of DNA and RNA during isolation, rigorous preanalytical sample characterization, applicability to real-world clinical specimens, and the accuracy of bioinformatics pipelines [19,20]. All the while, clinical testing guidelines are increasingly recommending broad, multicategorical molecular testing to inform treatment decisions and disease management in multiple cancers.

Lung cancer represents a disease in which broad yet streamlined molecular testing capabilities are acutely needed given the vast array of targeted and immunotherapeutic options available.

Molecular profiling of NSCLC is recommended by National Comprehensive Cancer Network (NCCN) guidelines, but many patients are not fully tested [21,22]. Barriers to access include cost, the number of tests required (≥3 separate tests are currently required), suitability of specimens, and regional access [23,24]. Many clinical labs do not offer comprehensive testing due to a lack of resources, expertise, or time required to develop these assays. This study addresses the need for a comprehensive, streamlined, sample-to-answer NGS technology and accommodates the spectrum of clinical specimen quality and molecular markers required to satisfy evolving NSCLC biomarker testing recommendations. Our results demonstrate that an NGS workflow that unifies the analysis of DNA and RNA markers in NSCLC can be used with challenging formalin-fixed paraffin-embedded (FFPE) specimens to recapitulate well-characterized molecular profiles.

Materials and Methods

Tumor Specimens
All specimens were obtained from MD Anderson Cancer Center (Houston, TX) under Institutional Review Board–approved study protocols. Core needle biopsies (CNBs) were obtained as a set of FFPE tissue slides in 5-μm or 10-μm sections with 10-30 μm available for each tumor (N = 109). Surgically resected specimens were obtained as 2 × 10–μm sections for each tumor (N = 110). CNBs and surgical resections are summarized in Table 1. Fine-needle aspirate (FNA) smears for 50 BATTLE-2 trial subjects were obtained as sets of one or three slides.

Target Selection and Panel Design
The hybrid DNA/RNA panel consists of an RNA library pool which covers 107 gene fusions recurrent in NSCLC including ALK, RET, and ROS1. DNA hotspots were prioritized according to the mutation prevalence and level of evidence supporting targetable therapy options in NSCLC. The DNA NGS library covers 55 hotspot regions in 20 genes including EGFR, KRAS, STK11, PIK3CA, TP53, and others (Table 2) and is based on previously

Table 1. Clinicopathological Summary of NSCLC Specimens Profiled by Targeted NGS

		Core Needle Biopsies		Surgical Resections	
		Total (N = 109)	Percent	Total (N = 110)	Percent
Sex	Male	52	47.7	65	59.1
	Female	56	51.4	45	40.9
	Unknown	1	0.9	0	0.0
Age (years)	<60	24	22.0	21	19.1
	60-70	42	38.5	42	38.2
	>70	42	38.5	47	42.7
	Unknown	1	0.9	0	0.0
Stage	I	56	51.4	51	46.4
	II	30	27.5	36	32.7
	III	16	14.7	22	20.0
	IV	5	4.6	1	0.9
	Unknown	2	1.8	0	0.0
Histopathology	Adenocarcinoma	71	65.14	50	45.5
	Squamous cell carcinoma	38	34.86	60	54.5
Smoking history	Never	10	9.2	12	10.9
	Former	52	47.7	63	57.3
	Current	46	42.2	35	31.8
	Unknown	1	0.9	0	0.0

Table 2. Content of the Targeted NGS DNA-Seq and RNA-Seq Panel

RNA Pool Content					DNA Pool Content	
3' Fusion Partner	# Unique Breakpoints	3'/5' Imbalance	Additional RNA Targets	DNA Targets		
ALK	53	●	ABCB1	MSLN	*ALK*	*FGFR3*
ROS1	22	●	BRCA1	PDCD1	*BRAF*	*KRAS*
RET	12	●	CD274 (PD-L1)	PDCD1LG2 (PD-L2)	*CTNNB1*	*MAP2K1*
FGFR3	7		CDKN2A	PTEN	*DDR2*	*MET*
NTRK3	3		CTLA4	RRM1	*EGFR*	*NRAS*
NTRK1	4	●	ERCC1	TDP1	*ERBB2*	*PIK3CA*
NRG1	2		ESR1	TERT	*ERBB4*	*PTEN*
FGFR1	1		FGFR1, FGFR2	TLE3	*FBXW7*	*SMAD4*
FGFR2	1		IFNGR	TOP1	*FGFR1*	*STK11*
MBIP	1		ISG15	TUBB3	*FGFR2*	*TP53*
PDGFRA	1		MET / MET e14	TYMS		

The panel covers 107 recurrent gene fusions, 3'/5' imbalance targets, *MET* exon 14 skipping, 23 mRNA expression markers, and 55 DNA mutation region of interest regions in 20 genes relevant to NSCLC.

published methods [25–27]. Development and validation of the RNA NGS panel are described extensively in previously published work [28]. The panel includes coverage of all DNA and RNA markers recommended by the NCCN NSCLC guideline [29] as part of broad molecular testing, in addition to emerging markers of clinical research value [30]. RNA primer designs were selected to span exon-exon breakpoints specific to the target transcript. Additional designs enable detection of 3'/5' imbalance of recurrently 3' fused genes to support confirmation of known and detection of novel rearrangements. The RNA pool enables detection of *MET* exon 14 skipping and quantification of 23 other mRNA targets of published prognostic and theranostic value, including markers relevant to immune checkpoint inhibitor response such as PD-L1, PD-L2, IFNG, and CTLA4. Endogenous control targets utilized by the DNA- and RNA-specific real-time qPCR QC assays were selected based on an analysis of TCGA LUAD and lung squamous cell carcinoma (LUSC) cohorts to identify genes that are copy number neutral and stably expressed in diseased and normal lung tissue [31,32]. A set of three stably expressed reference genes, including the RT-qPCR controls, was included in the RNA pool to enable normalization of expression levels and provide quality control measures.

Specimen Preparation and Characterization

For each tissue specimen, multiple sections were scraped from the glass slides and combined into a single tube. The tissue was then incubated with xylene to remove residual paraffin followed by extraction using QIAamp® DNA FFPE Tissue Kit (QIAGEN, Germantown, MD) following manufacturer's protocol with the exclusion of RNase treatment in order to recover total nucleic acid (TNA). TNA was processed and eluted into 50 μl of ATE buffer (QIAGEN). FNA smears were isolated using the QIAamp® DNA FFPE Tissue Kit following manufacturer's protocol, excluding RNase digestion. TNA was eluted in 30 μl of nuclease-free water. Columns were eluted a second time with 50 μl into a fresh tube to capture any material that was not eluted in the first elution. TNA elutions were assessed for DNA and RNA yield using qPCR assays that quantify discrete populations of amplifiable DNA and RNA, QuantideX® qPCR DNA QC Assay (Asuragen, Inc., TX) and the qPCR RNA QC Assay (QuantideX® NGS RNA Lung Cancer Kit, Asuragen), respectively, according to manufacturer's protocol.

Library Preparation and Sequencing

Preanalytical QC of TNA was utilized to inform library input and downstream analysis of NGS results. TNA samples were stratified based on resulting amplifiable copies per μl. FNA smear TNA isolations followed the targeted DNA-Seq protocol for sample inputs described in previous work [27] to attain a target of 400 amplifiable DNA template copies per reaction. RNA libraries were not prepared from FNA smears as only one sample yielded sufficient RNA for evaluation. For surgical resections and CNBs, a minimum of 200 copies per reaction was used as input to gene-specific PCR for both DNA and RNA libraries. RNA and DNA libraries were either prepared in a shared 96-well PCR-plate and co-sequenced on a shared MiSeq run or prepared and sequenced as separate batches. RNA libraries were prepared utilizing the QuantideX® NGS RNA Lung Cancer Kit (Asuragen) reagents according to the manufacturer's protocol. The reverse transcription (RT) product resulting from the RNA QC was transferred to a multiplexed PCR for target-specific enrichment. Preparation of the DNA libraries followed processes, buffers, and cycling conditions identical to the QuantideX® NGS RNA Lung Cancer Kit with the exclusion of the reverse transcription step and utilization of DNA pool gene-specific primers. Following multiplexed target-specific enrichment PCR, libraries were transferred to a tagging PCR to simultaneously incorporate sample-specific index codes and sequencing adapters for MiSeq NGS compatibility. The resultant libraries were purified with Library Pure Prep Beads. Individual purified libraries were diluted in two serial 100-fold dilutions, and the library concentrations were measured using the included Library Quant qPCR assay (QuantideX® NGS RNA Lung Cancer Kit, Asuragen). The DNA and RNA purified libraries were normalized according to the relative coverage requirements of each library and pooled to 2.5 nM total concentration. The final library pool was denatured to allow for 15 pM loading concentration with the addition of PhiX (Illumina, San Diego, CA) at 1.2 pM. The denatured library pool and QuantideX® NGS custom sequencing primers were run at 2×201 cycles on the Illumina MiSeq with v3 MiSeq reagents (Illumina).

NGS Analysis

Postsequencing, data from DNA and RNA library pools were demultiplexed and analyzed through separately optimized informatics pipelines, and results were integrated for supplemental analyses. Outside of specifically referenced tools, all custom pipeline analysis code was developed in Python.

FASTQs generated from RNA libraries were adapter trimmed and filtered for dual index code purity. The I7 and I5 dual index code pairs were identified within the forward and reverse reads, and read pairs with unexpected code pairs were filtered. This index purity

filtering step improves the specificity of the final calls and addresses a known limitation of Illumina's chemistry [18,33,34]. A local, gapped alignment to a custom reference transcriptome (inclusive of targeted breakpoint junctions) was performed using bowtie2 v2.0.5 (using the option: –sensitive-local) [35]. Alignments that did not match the expected amplicon boundaries or contain large gaps were filtered. Fusions and splice variant detection employed an upper-tailed Poisson test statistic, and 3′/5′ expression imbalances were assessed on normalized 3′ and 5′ expression data described in previously published work [28]. Gene expression was normalized by the geometric mean of endogenous control genes (*TBP*, *RAB5C*, and *GGNBP2*). Library QC was determined based on preanalytical QC and sequencing coverage measures [28].

DNA library analysis followed a similar workflow wherein read pairs were adapter-trimmed and associated with expected amplicons through local sequencing alignment using bfast. Unassociated amplicons were filtered prior to a second pass alignment against the GRCh37 reference using bwa-mem v0.6.1 (using the following options: -O 5,5 -E 1,1) [36]; local realignment and QScore recalibration were performed using GATK v1.3-21 [37]. INDEL calling was performed based on the % variant allele frequency (VAF) and coverage depth with empirically determined thresholds [27]. A decision tree algorithm trained on an independent cohort of 400 FFPE samples was used to call SNVs [27]. The SNV classifier incorporates multiple features including functional input copies, sequence quality, sample specific error rates, local sequence complexity, and coverage depth. Evidence of copy number amplifications was assessed after normalizing each amplicon's coverage against the mean across all amplicons within a given library. SNPs and synonymous variants were filtered prior to supplementary analyses.

Gene Fusion and MET Exon 14 Splice Variant Confirmation by Orthogonal Assays

RT was performed on the original TNA isolations to generate a replicate cDNA product. Primers were designed (or reused from the panel) for each confirmation target to produce an amplicon of known length within a singleplex reaction with common sequence tags for second-stage tagging PCR. Singleplex PCR enrichments were tagged with FAM-labeled primers. PCRs were evaluated via capillary electrophoresis (CE) on a 3500XL Genetic Analyzer (Thermo Fisher Scientific, Waltham, MA) to confirm expected size. *MET* exon 14 skipping was confirmed by a digital PCR assay using custom primers on the QX200™ Droplet Digital™ PCR System (Bio-Rad Laborato-

ries, Inc., Hercules, CA). For a subset of targets, TaqMan Gene Expression Assays (Thermo Fisher Scientific) were used according to the manufacturer's instructions on an Applied Biosystems™ 7500 real-time PCR system. For fusions that did not have a commercial qPCR assay available, a hydrolysis probe was designed to span the specific breakpoint, and a qPCR assay was performed in conjunction with the gene-specific primers associated with the panel to confirm fusion presence (95°C for 5 minutes, 40 cycles: 95°C for 5 seconds, 60°C for 1 minute).

Results

An Integrated NGS Workflow Enabling Streamlined Analysis of DNA and RNA

An overview of the integrated targeted NGS workflow for DNA and RNA is shown in Figure 1. This approach enables the interrogation of both DNA and RNA through a single isolation of TNA from low-quality or low-quantity FFPE or FNA specimens. In order to determine the suitability of the TNA for NGS, a preanalytical QC assay is performed for both DNA and RNA with a readout of amplifiable template molecules for each analyte. Library preparation employs a two-pool strategy that utilizes a two-step PCR with shared buffers and cycling conditions across the DNA and RNA libraries to streamline the workflow. Postsequencing, DNA and RNA library pools are demultiplexed and analyzed through separately optimized informatics pipelines, the results of which are integrated together to assess implications across both categories of nucleic acid. The combined workflow detects and quantifies SNVs, INDELs, CNVs, fusions, splice variants, and expression of select genes to minimize cost and specimen consumption. To demonstrate the clinical research value of this integrated NGS method, we designed a panel (Table 2) for NSCLC that covers DNA mutation hotspots and RNA fusions, splice variants, and expression targets. Panel content was selected based on a review of the NCCN guidelines, COSMIC database, TCGA results, and active clinical trial research (see Methods).

Preanalytical QC Analysis of DNA and RNA

We applied the preanalytical QC assays to a cohort of 219 NSCLC patients (Table 1; Supplemental Table S1) to quantify the amplifiable DNA and RNA fractions from TNA isolations (Supplemental Table S2). The cohort consisted of 109 FFPE CNB specimens and 110 FFPE surgical resections representing 121 LUADs and 98 LUSCs.
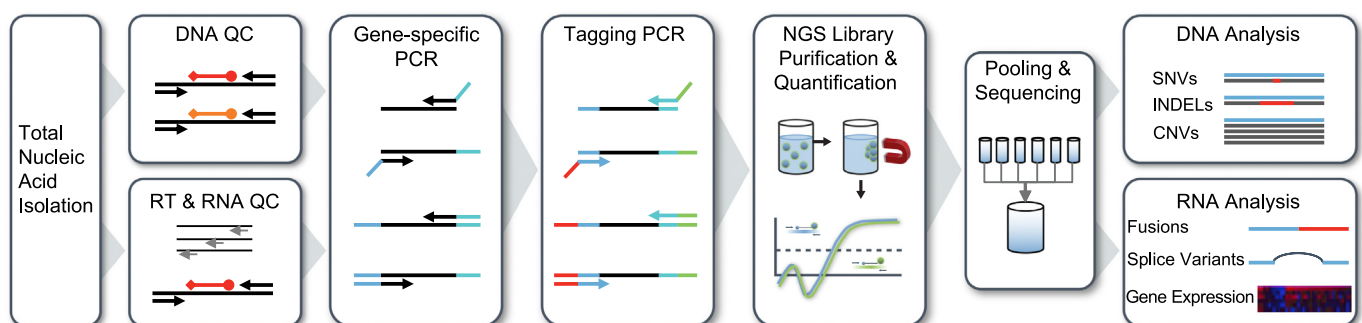

Figure 1. Overview of the NGS procedure targeting both DNA and RNA. An integrated NGS workflow enables parallel analysis of DNA and RNA through a real-time qPCR QC analysis and target enrichment through two-step PCR.
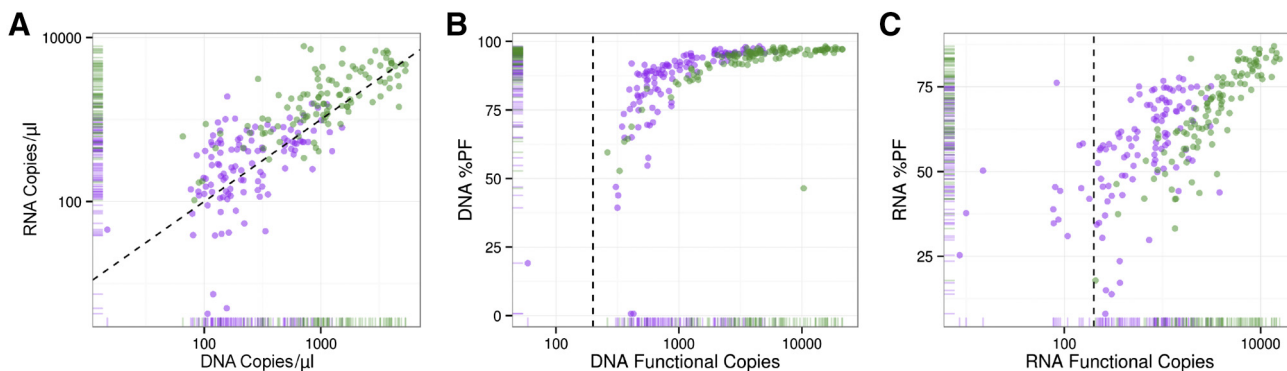
Figure 2. Amplifiable DNA and RNA yields are correlated and predictive of sequencing quality. (A) Assessment of CNBs (purple) and surgical resections (green) reveals a relationship between amplifiable DNA and RNA copies/μl as determined by RT-qPCR. (B) DNA-Seq and (C) RNA-Seq passing filter (PF) mapping rates to panel targets (y-axis) are predicted by amplifiable template molecules used for library prep (x-axis). Functional QC thresholds are shown as dashed vertical lines in panels B and C.

The number of amplifiable DNA copies from these FFPE samples was significantly correlated with amplifiable RNA copies as defined by the RNA endogenous control target [Spearman's correlation coefficient (SCC): 0.81, P <2.2e-16, Figure 2A]. On average, surgically resected FFPEs yielded more DNA and RNA than CNBs (Figure 2A). By utilizing the matched DNA copy number data as a reference to determine the functional cell count equivalent for a given specimen (assuming diploidy of the DNA control), we calculated that the RT-qPCR endogenous control RNA target was expressed at approximately 2.8 copies per cell on average across all specimens, which are indicative of stable expression in the same copy number range as our DNA control.

We defined QC thresholds for DNA libraries through independent studies [27] and established 200 amplifiable copies as the minimum input required to reliably call DNA variants down to 5% VAF while avoiding false-negative calls. A similar approach was taken for RNA libraries [28], and a minimum functional cDNA copy input of 200 copies was established as the requirement to reliably call fusion variants down to 5% cell positivity. Based on these criteria, we found

that 108/109 (99%) and 94/109 (86%) of the CNB specimens met our preanalytical QC criteria for DNA and RNA, respectively. Preanalytical QC criteria were met by all 110 (100%) surgically resected specimens for both DNA and RNA.

Library preparation and analysis were performed on all specimens independent of preanalytical QC results. Consistent with other studies [25,26], the number of functional template molecules used in preparation of NGS libraries was significantly associated with the percentage of reads mapping to the intended targets for both targeted DNA-Seq and RNA-Seq targets (SCC: 0.87 and 0.75 for DNA and RNA, respectively; Fig. 2, B and C).

Targeted DNA- and RNA-Seq Cohort Characterization

The cohort was next characterized using the integrated NSCLC NGS assay (Figure 3; Supplemental Tables S3 and S4). Consistent with other studies such as TCGA, fusions in *ALK*, *RET* and *ROS1* were restricted to the LUAD subtype. Eight fusions were identified across all specimens: *EML4-ALK (N = 2)*, *KIF5B-RET*, *EZR-ROS1*, *CD74-ROS1*, *KIF5B-RET*, *CCDC6-RET*, and *CD74-NRG1*.
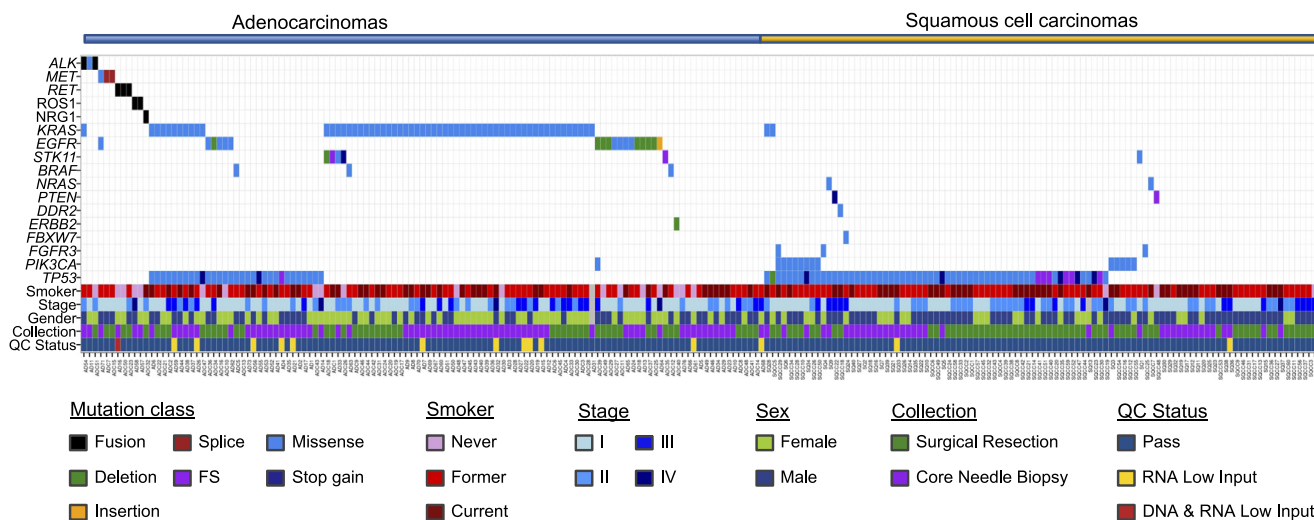

Figure 3. Multiomic characterization of NSCLC FFPE cohort reveals a spectra of mutations consistent with underlying histological subtype.
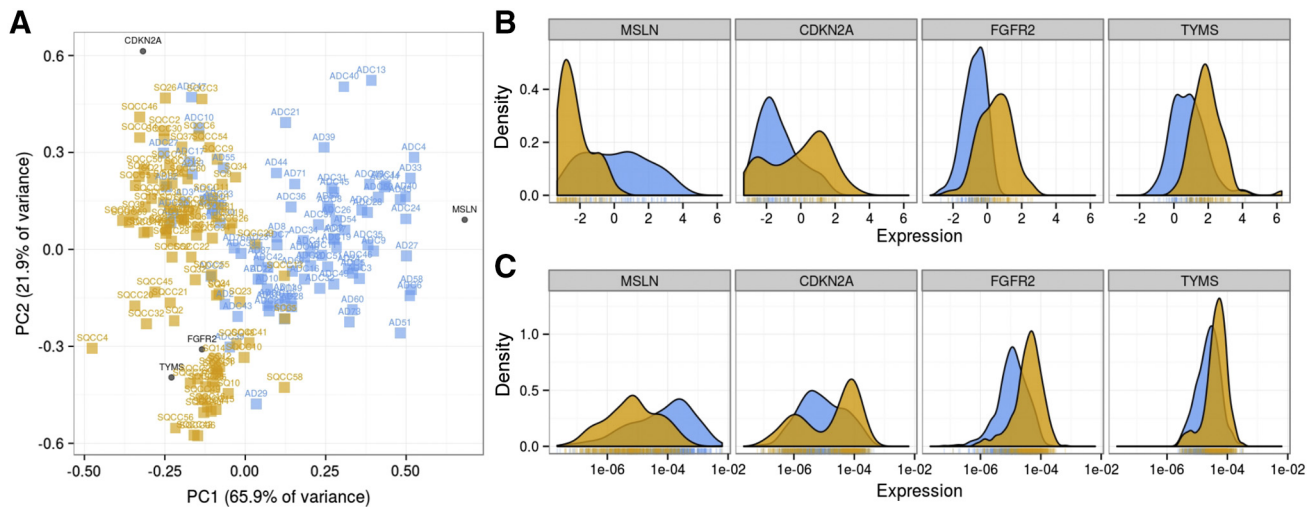
Figure 4. Select mRNA expression markers revealed by targeted RNA-Seq distinguish LUSCs from LUADs. (A) PCA analysis of differentially expressed mRNAs between LUADs (blue) and LUSCs (orange). Comparison of expression distributions of log2 transformed normalized expression (x-axis) for mRNAs differentially expressed between LUADs (blue) and LUSCs (orange) shows similar profiles when comparing (B) our data (targeted RNA-Seq) to C) TCGA (microarray).

Imbalance of 3′ expression relative to 5′ expression was detected for both *ALK* fusions and for two of three *RET* fusions. The one *RET* fusion for which an imbalance was not detected was in a specimen that fell below the minimum RNA functional copy input of 200 copies. Neither of the *ROS1* fusions showed evidence of 3′/5′ imbalance. Three imbalances were detected in samples that were not accompanied by a positive fusion call. Two of these samples, ADC48 (*ALK*) and SQCC17 (*ALK*), were otherwise mutation negative. However, the third sample (AD62), which had a *RET* imbalance, was also positive for *EGFR* and *TP53* mutations. In addition, *MET* exon 14 skipping was detected in two LUADs, one of which had ~30% skipped and the other with nearly 100% skipped. All gene fusions and *MET* exon 14 skipping calls were confirmed by an independent assay (see Methods).

DNA variant calls were consistent with the underlying histopathology and mutation prevalence according to TCGA and COSMIC. For example, mutations in *KRAS*, *EGFR,* and *STK11* were present in LUADs whereas *PIK3CA* and *FGFR3* were detected in LUSC. *TP53* mutations were detected in both subtypes. Also consistent with other NSCLC cohorts, *KRAS* and *EGFR* mutations in LUAD specimens were mutually exclusive events [32,38]. When examining the spectrum of *KRAS* mutations in the LUAD specimens, codon 12 was the most frequently mutated (74% of all 62 *KRAS* mutations), a result that is consonant with other studies such as the TCGA LUAD cohort where codon 12 represented 96% of KRAS mutations [32]. Mutations in other codons of *KRAS* were largely represented by codons 13 (18%) and 61 (5%). Two unexpected *KRAS* mutations were detected in LUSC. One specimen, SQ28 (*KRAS* p.G13D) was identified as a poorly differentiated NSCLC with LUAD features (TTF1+, p40−) and the other, SQCC5, presented with a noncanonical *KRAS* variant p.N26I at 20% variant allele frequency.

Integrative analysis of DNA and RNA markers found 10 cases that were positive for targeted RNA-based variants (*MET* exon 14 skipping or gene fusions). Nine of those 10 cases were negative for any DNA mutation, suggesting that these oncogenic driver events are mutually exclusive. Interestingly, one case was positive for both

EML4-ALK and *KRAS* p.G12D. While *ALK* fusions are generally thought to be mutually exclusive with other oncogenic mutations such as *KRAS* and *EGFR*, these have been previously reported to co-occur in rare instances [39,40].

Comparative Expression Analysis of Adenocarcinomas and Squamous Cell Carcinomas

We compared the expression profiles between LUAD and LUSC subtypes to identify mRNAs within those targeted by the panel that were differentially expressed between the two histological subtypes. *MSLN, TYMS, CDKN2A,* and *FGFR2* showed significant differences in expression between subtypes (corrected $P < 0.05$ and more than two2-fold differences for all four genes, Figure 4), with elevated levels of *MSLN* in LUADs and elevated levels of *CDKN2A*, *FGFR2*, and *TYMS* in LUSCs. *MSLN* has been reported to be expressed at higher levels in LUADs relative to LUSCs [41]. Similarly, TCGA data indicate higher levels of *CDKN2A*, *FGFR2*, and *TYMS* in LUSCs [31,32]. Indeed, when comparing our results to TCGA LUAD and LUSC cohorts, we observed qualitatively similar gene expression distributions (Figure 4, *B* and *C*). While *CDKN2A* is on average expressed at higher levels in LUSCs relative to LUADs, the gene is also reported as recurrently downregulated in LUSCs via epigenetic silencing and whole gene deletions [31]. This is consistent with the apparent bimodality of the *CDKN2A* expression distribution within LUSCs, evident in our data (Figure 4*B*) and the TCGA LUSC cohorts (Figure 4*C*). The segregation of LUSCs, largely driven by *CDKN2A* expression, is also apparent by PCA analysis (Figure 4*A*). Further co-expression analysis revealed a strong positive correlative relationship between PD-L1 and PD-L2 mRNA expression in both LUAD (SCC: 0.76; $P < 2e-16$, Figure S1) and LUSC (SCC: 0.66; $P < 2e-16$) subtypes, consistent with previous reports [42]. Additional immune checkpoint related expression markers (CTLA4, IFNG, ISG15, PDCD1) also significantly correlated with PD-L1 and PD-L2 mRNA across both subtypes (data not shown).

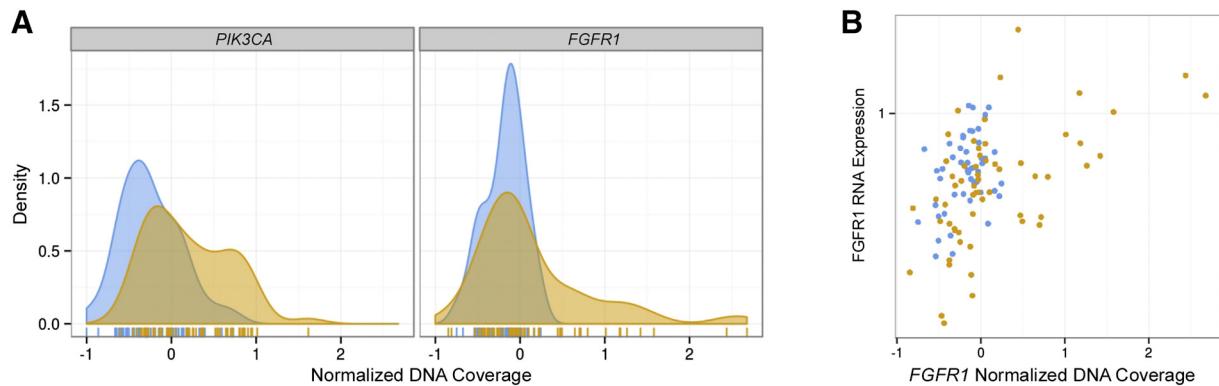### Integrated Copy Number Variation Analysis of DNA and RNA Targets

Figure 5. Evidence of copy number variation by integrated DNA and RNA analysis. (A) Increased levels of DNA coverage in LUSC (orange) cases relative to LUAD (blue) cases for *PIK3CA* and *FGFR1* in surgical specimens. (B) FGFR1 displays concomitantly elevated levels of normalized DNA coverage and RNA expression in surgical specimens specific to the LUSC subtype.

Analysis of normalized DNA coverage revealed copy number variation specific to histopathological subtype and consistent with previous molecular characterization studies. *FGFR1* and *PIK3CA* amplification was evident in LUSCs in contrast to LUADs for surgical specimens (Figure 5A). This observation was less pronounced for CNB specimens. Among the DNA loci that were targeted, *FGFR1* and *PIK3CA* were the two most frequently amplified genes in LUSC as reported by TCGA with frequencies of 18% and 46%, respectively [31]. Evidence of *FGFR1* amplification in the LUSC subtype was further supported by targeted RNA-Seq evidence of concomitant overexpression of *FGFR1* (SCC: 0.56; $P < 3.5e-6$; Figure 5B). Putative *MET* amplifications were also identified in two LUADs with concomitant MET mRNA overexpression (data not shown).

DNA Variant Analysis of the BATTLE-2 NSCLC Cohort

We tested the technical limits of the targeted DNA/RNA NGS technology by assessing a subset of 50 FNA smears from the BATTLE-2 trial. Smears contained only ~100 to a few hundred cells. Assessment by the preanalytical QC assays determined that only a single specimen yielded sufficient RNA for analysis, whereas 24/50 had sufficient yield for DNA analysis. Targeted DNA-Seq analysis was performed, and results were compared against matched data from the FoundationOne NGS assay [43], generated using much higher inputs of less challenging FFPE tissue specimens. When assessing DNA regions that were commonly covered by both assays, FoundationOne NGS analysis of tissue specimens was in 97% agreement with targeted DNA-Seq results from FNA smears (Table S5). Only one discordant variant was identified, *MET* p.T1010I, detected at 59% VAF in our assay but not reported by FoundationOne. This variant is annotated as both a COSMIC variant and an SNP and likely to be germline in origin.

Discussion

NGS analysis of DNA and RNA has increasingly become the primary enabling technology of precision medicine. Despite the centrality of this approach, NGS analyses of DNA and RNA have largely remained separate workflows. In this work, we developed an NSCLC panel using an NGS workflow that joins analysis of DNA and RNA through a single source TNA input and harmonized PCR conditions. We applied this workflow to the analysis of a cohort of FFPE NSCLC specimens, demonstrating its utility to accurately detect SNVs,

INDELs, CNVs, gene fusions, and gene expression signatures consistent with previously published molecular analyses of NSCLC. We then applied the approach to a challenging cohort of FNA smears with 97% concordance with an orthogonal analysis of matched surgical resections.

The need for dual testing of DNA and RNA is primarily driven by the increasingly multiomic composition of the guidelines for molecular testing such as NCCN and European Society for Medical Oncology (ESMO), which now recommend testing of SNVs, INDELs, CNVs, gene fusions, and protein biomarkers within a single indication. NSCLC is perhaps one of the best examples of where this confluence of molecular marker categories is coming to the fore. Combined testing of DNA and RNA enables broad coverage of these categories, and RNA expression profiling may be an analytical substitute for certain protein expression biomarkers such as PD-L1 [44]. For some categories of markers, detection is possible with either DNA or RNA molecules, yet each offers distinct advantages. For example, RNA is better suited for the detection of gene fusions because a multitude of DNA-level intron-intron breakpoints can give rise to a single RNA exon-exon gene fusion. The average intron size is approximately 3.4 kb [45], which means that the DNA coverage requirements to resolve gene fusion events are at least an order of magnitude greater than the RNA coverage requirements. In addition, RNA fusion expression is often appreciably higher than the corresponding DNA, improving the sensitivity of detection. In contrast, DNA offers distinct advantages when profiling variants that lie outside of transcribed regions, such as promotor mutations. There is currently far more evidence associating DNA variant allele frequencies (VAFs) to therapeutic outcome than RNA VAFs. As such, the molecular pathology community is more accustomed to interpreting SNVs and INDELs from DNA-based assays. The advent of integrative analysis of DNA and RNA presents an opportunity to assess the relative merits of each analyte.

Single-workflow testing of DNA and RNA also offers advantages for intra-assay confirmation. For example, we identified *FGFR2* and *MET* amplification events by DNA analysis that exhibited concomitant RNA overexpression, thus providing further functional evidence for the amplification event. Detection of RNA 3′/5′ imbalances offers a means of affirming the presence of explicitly targeted gene fusions while also detecting the presence of fusions with rare, noncanonical breakpoints not represented by the set of RNA breakpoint-spanning amplicons. We

found that 4/7 gene fusions were supported by a corresponding 3′/5′ imbalance event. Of note, all of the *ALK* fusion events (*N* = 2) showed a supporting imbalance ratio, whereas none of the *ROS1* fusions did (*N* = 2). This is because *ROS1* is endogenously expressed in lung tissue, whereas *ALK* is not. Therefore, the *ROS1* imbalance has a greater intrinsic background, which must be overcome in order to be detectable relative to ALK, and thus the sensitivity of imbalance markers inherently depends upon a gene's level of endogenous expression. Interpretation of histopathology is also enabled through integrative molecular analysis. We identified DNA mutations that were able to lend further insights to ambiguous histopathologies and an RNA expression signature that distinguished adenocarcinomas from squamous cell carcinomas.

Combined, single-pass testing of DNA and RNA increases the likelihood of identifying a therapeutically relevant oncogenic driving event. This is due to the observation that DNA and RNA drivers are often mutually exclusive molecular events. Thus, DNA driver mutation-negative cases have a higher likelihood of bearing an RNA driver and vice-versa. In our own study, we found DNA and RNA variants to be largely exclusive events with 9 of the 10 subjects positive for an RNA driver lacking any of the common DNA mutations covered by the panel. In fact, when considering the theranostic yield of the assay (the number of subjects for which a guideline-recommended or emerging targeted therapy was identified), the inclusion of RNA markers achieved a 41% increase over DNA markers alone. Yet there were exceptions to this exclusivity. For example, one case revealed both an *EML-ALK* fusion and a *KRAS* p. G12D mutation. The co-occurrence of these two events is rare but not unprecedented [40], and the low VAF of *KRAS* p.G12D in this specimen (6.9%) is suggestive of a subclonal or potentially distinct population of tumor cells. Mutations in *KRAS* have been established as a mechanism of innate and acquired crizotinib resistance [46–48], which further highlights the clinical research value of integrative profiling of DNA and RNA to identify and further characterize resistance mechanisms. Despite the encouraging response rates of *ALK*, *RET*, and *ROS* positive cases to TKI therapies, resistance and progression ultimately occur within 1 year of treatment for the vast majority of patients. Joint analysis of DNA and RNA can support the identification of intrinsic and acquired resistance and enable the informed selection of second- and third-line TKI therapies.

Despite the advantages of unified DNA/RNA analysis, current commercial offerings and molecular testing solutions are largely fragmented and cumbersome. The vast majority of published and commercialized workflows address either DNA or RNA analysis exclusively but not both [49–52]. Some workflows have attempted to combine both into a single protocol such as the Oncomine Dx Target Test, Focus, and Comprehensive NGS assays [53], but these assays lack the advantages of our approach, such as support for gene expression quantification and a single-source input of TNA which enables a single isolation and eliminates purification steps (such as DNase treatment) that risk material loss. Our method enables streamlined single-plate library preparation through shared master mixes and harmonized reaction conditions. Workflows such as the TruSight Tumor 170 offer single-plate RNA/DNA library preparation but require sonication, ligation, hybridization, and clean-up steps that introduce workflow complexity and sample attrition. Our targeted amplicon sequencing workflow is free of these cumbersome steps and requires less material (20 ng TNA or 10 ng DNA and 10 ng RNA) than TruSight (minimum 40 ng DNA and 40 ng RNA). Finally, our incorporation of preanalytical QC measurements

to inform bioinformatics variant detection and interpretation enables the suppression of false-positive variant calls and flags samples at risk of false-negative calls, thus enabling accurate analysis of poorer-quality specimens. The analytical validity of our approach is supported by 100% confirmation rates of RNA fusions and *MET* exon 14 skipping events by qPCR or PCR/CE and high analytical concordance in our analysis of FNA smears from the BATTLE-2 trial with an independent NGS assay on matched surgical resections.

Joint preanalytical QC analysis of DNA and RNA offered additional insights into the quality distinctions between analytes and specimen types. As expected, FFPE surgical resections were of the highest quality, with 100% of specimens yielding sufficient DNA and RNA for NGS analyses. In contrast, 99% and 86% of CNB specimens offered sufficient material for DNA and RNA NGS analyses, respectively. At the extreme end of the quality spectrum were the FNA smears with 48% and 2% pass rates for DNA and RNA, respectively. Across this continuum of quality, we noticed a general trend with respect to the suitability of a specimen's DNA relative to RNA, namely, that specimens were more likely to offer sufficient DNA for analysis. A parallel and related observation was that analysis of RNA expression trends (analysis of immune checkpoint inhibitor expression markers and CNAs) tended to show more robust associations in the FFPE surgical resections relative to CNBs.

The longer-term rationale for a conjoined analysis of DNA and RNA is to provide a foundation for integrative, targeted NGS applications that pair genotype with molecular phenotype in order to interpret variants of unknown significance through orthogonal functional evidence. In this paradigm, RNA expression can be viewed as a molecular information bottleneck that can aid in the interpretation of the long tail of genetic and epigenetic variation that underlies tumor evolution. Functional pathway analysis will enable the resolution of cryptic and latent oncogenic driver mutations and shed light onto the "dark matter" that underlies cancer. The NGS workflow presented here offers a foundation for integrative analysis of DNA and RNA compatible with a range of clinical specimen types in a small-footprint format suitable for routine diagnostic and clinical research purposes to enable holistic interpretation of cancer specimens.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.tranon.2019.02.012.

References

[1] Pant S, Weiner R, and Marton MJ (2014). Navigating the rapids: the development of regulated next-generation sequencing-based clinical trial assays

and companion diagnostics. *Front Oncol* **4**, 1–20. http://dx.doi.org/10.3389/fonc.2014.00078.

[2] Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, and Haluska FG, et al (2004). Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* **350**, 2129–2139. http://dx.doi.org/10.1056/NEJMoa040938.

[3] Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, and Fulton L, et al (2004). EGF receptor gene mutations are common in lung cancers from "never smokers" and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci* **101**, 13306–13311. http://dx.doi.org/10.1073/pnas.0405220101.

[4] Heigener DF, Schumann C, Sebastian M, Sadjadian P, Stehle I, Marten A, Luers A, Griesinger F, and Scheffler M (2015). Afatinib in non–small cell lung cancer harboring uncommon egfr mutations pretreated with reversible EGFR inhibitors. *Oncologist* **20**, 1167–1174. http://dx.doi.org/10.1634/theoncologist.2015-0073.

[5] Wang S, Cang S, and Liu D (2016). Third-generation inhibitors targeting EGFR T790M mutation in advanced non-small cell lung cancer. *J Hematol Oncol* **9**, 1–7. http://dx.doi.org/10.1186/s13045-016-0268-z.

[6] Kohno T, Nakaoku T, Tsuta K, Tsuchihara K, Matsumoto S, Yoh K, and Goto K (2015). Beyond ALK-RET, ROS1 and other oncogene fusions in lung cancer. *Transl Lung Cancer Res* **4**, 156–164. http://dx.doi.org/10.3978/j.issn.2218-6751.2014.11.11.

[7] Kwak EL, Bang Y-J, Camidge DR, Shaw AT, Solomon B, Maki RG, Ou S-HI, Dezube BJ, Jänne PA, and Costa DB, et al (2010). Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* **363**, 1693–1703. http://dx.doi.org/10.1056/NEJMoa1006448.

[8] Bergethon K, Shaw AT, Ou SHI, Katayama R, Lovly CM, McDonald NT, Massion PP, Siwak-Tapp C, Gonzalez A, and Fang R, et al (2012). ROS1 rearrangements define a unique molecular class of lung cancers. *J Clin Oncol* **30**, 863–870. http://dx.doi.org/10.1200/JCO.2011.35.6345.

[9] Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, and Shimada Y, et al (2012). KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* **18**, 375–377. http://dx.doi.org/10.1038/nm.2644.

[10] Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, and Uehara H, et al (2012). RET, ROS1 and ALK fusions in lung cancer. *Nat Med* **18**, 378–381. http://dx.doi.org/10.1038/nm.2658.

[11] Rosell R and Karachaliou N (2016). RET inhibitors for patients with RET fusion-positive and RET wild-type non-small-cell lung cancer. *Lancet Oncol* **17**, 1623–1625. http://dx.doi.org/10.1016/S1470-2045(16)30557-5.

[12] Abel HJ, Al-Kateb H, Cottrell CE, Bredemeyer AJ, Pritchard CC, Grossmann AH, Wallander ML, Pfeifer JD, Lockwood CM, and Duncavage EJ (2014). Detection of gene rearrangements in targeted clinical next-generation sequencing. *J Mol Diagn* **16**, 405–417. http://dx.doi.org/10.1016/j.jmoldx.2014.03.006.

[13] Moskalev EA, Frohnauer J, Merkelbach-Bruse S, Schildhaus H-U, Dimmler A, Schubert T, Boltze C, König H, Fuchs F, and Sirbu H, et al (2014). Sensitive and specific detection of EML4-ALK rearrangements in non-small cell lung cancer (NSCLC) specimens by multiplex amplicon RNA massive parallel sequencing. *Lung Cancer* **84**, 215–221. http://dx.doi.org/10.1016/j.lungcan.2014.03.002.

[14] Pekar-Zlotin M, Hirsch FR, Soussan-Gutman L, Ilouze M, Dvir A, Boyle T, Wynes M, Miller VA, Lipson D, and Palmer GA, et al (2015). Fluorescence in situ hybridization, immunohistochemistry, and next-generation sequencing for detection of EML4-ALK rearrangement in lung cancer. *Oncologist* **20**, 316–322. http://dx.doi.org/10.1634/theoncologist.2014-0389.

[15] Drilon A, Wang L, Arcila ME, Balasubramanian S, Greenbowe JR, Ross JS, Stephens P, Lipson D, Miller VA, and Kris MG, et al (2015). Broad, hybrid capture-based next-generation sequencing identifies actionable genomic alterations in lung adenocarcinomas otherwise negative for such alterations by other genomic testing approaches. *Clin Cancer Res* **21**, 3631–3639. http://dx.doi.org/10.1158/1078-0432.CCR-14-2683.

[16] Frampton GM, Ali SM, Rosenzweig M, Chmielecki J, Lu X, Bauer TM, Akimov M, Bufill JA, Lee C, and Jentz D, et al (2015). Activation of MET via diverse exon 14 splicing alterations occurs in multiple tumor types and confers clinical sensitivity to MET inhibitors. *Cancer Discov* **0**. http://dx.doi.org/10.1158/2159-8290.CD-15-0285.

[17] Karachaliou N, Gonzalez-Cao M, Crespo G, Drozdowskyj A, Aldeguer E, Gimenez-Capitan A, Teixido C, Molina-Vila MA, Viteri S, and De Los Llanos Gil M, et al (2018). Interferon gamma, an important marker of response to immune checkpoint blockade in non-small cell lung cancer and melanoma patients. *Ther Adv Med Oncol* **10**. http://dx.doi.org/10.1177/1758834017749748 [1758834017749748].

[18] Conroy JM, Pabla S, Glenn ST, Burgher B, Nesline M, Papanicolau-Sengos A, Andreas J, Giamo V, Lenzo FL, and Hyland FCL, et al (2018). Analytical validation of a next-generation sequencing assay to monitor immune responses in solid tumors. *J Mol Diagn* **20**, 95–109. http://dx.doi.org/10.1016/j.jmoldx.2017.10.001.

[19] Poirot B, Doucet L, Benhenda S, Champ J, Meignin V, and Lehmann-Che J (2017). MET exon 14 alterations and new resistance mutations to tyrosine kinase inhibitors: risk of inadequate detection with current amplicon-based NGS panels. *J Thorac Oncol* **12**, 1582–1587. http://dx.doi.org/10.1016/j.jtho.2017.07.026.

[20] Oliveira DM, Mirante T, Mignogna C, Scrima M, Migliozzi S, Rocco G, Franco R, Corcione F, Viglietto G, and Malanga D, et al (2018). Simultaneous identification of clinically relevant single nucleotide variants, copy number alterations and gene fusions in solid tumors by targeted next-generation sequencing. *Oncotarget* **9**, 22749–22768. http://dx.doi.org/10.18632/oncotarget.25229.

[21] Ruggiero JE, Rughani J, Neiman J, Swanson S, Revol C, and Green RJ (2017). Real-world concordance of clinical practice with ASCO and NCCN guidelines for EGFR/ALK testing in aNSCLC. *J Clin Oncol* **35**, 212. http://dx.doi.org/10.1200/JCO.2017.35.8_suppl.212.

[22] Lynch JA, Berse B, Rabb M, Mosquin P, Chew R, West SL, Coomer N, Becker D, and Kautter J (2018). Underutilization and disparities in access to EGFR testing among Medicare patients with lung cancer from 2010–2013. *BMC Cancer* **18**, 306. http://dx.doi.org/10.1186/s12885-018-4190-3.

[23] Levy BP, Chioda MD, Herndon D, Longshore JW, Mohamed M, Ou S-HI, Reynolds C, Singh J, Wistuba II, and Bunn PA, et al (2015). Molecular testing for treatment of metastatic non–small cell lung cancer: how to implement evidence-based recommendations. *Oncologist* **20**, 1175–1181. http://dx.doi.org/10.1634/theoncologist.2015-0114.

[24] Gonzalo MB, House L, Santiago K, Buzaglo JS, Zaleta AK, and Gupta NK (2017). Access to care in cancer: barriers and challenges. *J Clin Oncol* **35**, 33. http://dx.doi.org/10.1200/JCO.2017.35.8_suppl.33.

[25] Hadd AG, Houghton J, Choudhary A, Sah S, Chen L, Marko AC, Sanford T, Buddavarapu K, Krosting J, and Garmire L, et al (2013). Targeted, high-depth, next-generation sequencing of cancer genes in formalin-fixed, paraffin-embedded and fine-needle aspiration tumor specimens. *J Mol Diagn* **15**, 234–247. http://dx.doi.org/10.1016/j.jmoldx.2012.11.006.

[26] Sah S, Chen L, Houghton J, Kemppainen J, Marko AC, Zeigler R, and Latham GJ (2013). Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Med* **5**, 1. http://dx.doi.org/10.1186/gm481.

[27] Houghton J, Hadd AG, Zeigler R, Haynes BC, and Latham GJ (2016). Integration of wet and dry bench processes optimizes targeted next-generation sequencing of low-quality and low-quantity tumor biopsies. *J Vis Exp* , 1–13. http://dx.doi.org/10.3791/53836.

[28] Blidner RA, Haynes BC, Hyter S, Schmitt S, Pessetto ZY, Godwin AK, Su D, Hurban P, van Kempen LC, and Aguirre ML, et al (2019). Design, optimization, and multisite evaluation of a targeted next-generation sequencing assay system for chimeric RNAs from gene fusions and exon-skipping events in non-small cell lung cancer. *J Mol Diagn* **21**, 352–365. http://dx.doi.org/10.1016/j.jmoldx.2018.10.003.

[29] Shah P and Sands J (2018). Consensus on molecular testing in lung cancer. *Curr Pulmonol Reports* **7**, 49–55.

[30] Gallant J and Lovly CM (2018). Established, emerging and elusive molecular targets in the treatment of lung cancer; 2018 565–577. http://dx.doi.org/10.1002/path.5038.

[31] Cancer T and Atlas G (2012). Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525. http://dx.doi.org/10.1038/nature11404.

[32] Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, Beer DG, Cope L, Creighton CJ, and Danilova L, et al (2014). Comprehensive molecular profiling of lung adenocarcinoma. *Nature* . http://dx.doi.org/10.1038/nature13385.

[33] Kircher M, Sawyer S, and Meyer M (2012). Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. *Nucleic Acids Res* **40**, 1–8. http://dx.doi.org/10.1093/nar/gkr771.

[34] Esling P, Lejzerowicz F, and Pawlowski J (2015). Accurate multiplexing and filtering for high-throughput amplicon-sequencing. *Nucleic Acids Res* , 1–12. http://dx.doi.org/10.1093/nar/gkv107.

[35] Langmead B and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359. http://dx.doi.org/10.1038/nmeth.1923.

[36] Li H and Durbin R (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760. http://dx.doi.org/10.1093/bioinformatics/btp324.

[37] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, and Daly M, et al (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297–1303. http://dx.doi.org/10.1101/gr.107524.110.

[38] Unni AM, Lockwood WW, Zejnullahu K, Lee-Lin SQ, and Varmus H (2015). Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. *Elife* **4**, 1–23. http://dx.doi.org/10.7554/eLife.06907.

[39] Baldi L, Mengoli MC, Bisagni A, Banzi MC, Boni C, and Rossi G (2014). Concomitant EGFR mutation and ALK rearrangement in lung adenocarcinoma is more frequent than expected: report of a case and review of the literature with demonstration of genes alteration into the same tumor cells. *Lung Cancer* **86**, 291–295. http://dx.doi.org/10.1016/j.lungcan.2014.09.011.

[40] Campos-Gomez S, Lara-Guerra H, Routbort MJ, Lu X, and Simon GR (2015). Lung adenocarcinoma with concurrent KRAS mutation and ALK rearrangement responding to crizotinib: Case report. *Int J Biol Markers* **30**, e254–e257. http://dx.doi.org/10.5301/jbm.5000127.

[41] Ho M, Bera TK, Willingham MC, Onda M, Hassan R, FitzGerald D, and Pastan I (2007). Mesothelin expression in human lung cancer. *Clin Cancer Res* **13**, 1571–1575. http://dx.doi.org/10.1158/1078-0432.CCR-06-2161.

[42] Yearley JH, Gibson C, Yu N, Moon C, Murphy E, Juco J, Lunceford J, Cheng J, Chow LQM, and Seiwert TY, et al (2017). PD-L2 expression in human tumors: Relevance to anti-PD-1 therapy in cancer. *Clin Cancer Res* **23**, 3158–3167. http://dx.doi.org/10.1158/1078-0432.CCR-16-1761.

[43] Frampton GM, Fichtenholtz A, Otto GA, Wang K, Downing SR, He J, Schnall-Levin M, White J, Sanford EM, and An P, et al (2013). Development and validation of a clinical cancer genomic profiling test based on massively parallel DNA sequencing. *Nat Biotechnol* **31**, 1023–1031. http://dx.doi.org/10.1038/nbt.2696.

[44] Velcheti V, Schalper KA, Carvajal DE, Anagnostou VK, Syrigos KN, Sznol M, Herbst RS, Gettinger SN, Chen L, and Rimm DL (2014). Programmed death ligand-1 expression in non-small cell lung cancer. *Lab Invest* **94**, 107–116. http://dx.doi.org/10.1038/labinvest.2013.130.

[45] Hnilicová J and Staněk D (2011). Where splicing joins chromatin, 2; 2011 182–188. http://dx.doi.org/10.4161/nucl.2.3.15876.

[46] Doebele R, Pilling A, Aisner DL, Kutateladze TG, Le AT, Weickhardt AJ, Kondo KL, Linderman DJ, Heasley LE, and Franklin WA, et al (2012). Mechanisms of resistance to crizotinib in patients with ALK gene reaaranged NON-Small Cell lung cancer. *Clin Cancer Res* **18**, 1472–1482. http://dx.doi.org/10.1158/1078-0432.CCR-11-2906.Mechanisms.

[47] Mengoli MC, Barbieri F, Bertolini F, Tiseo M, and Rossi G (2016). K-RAS mutations indicating primary resistance to crizotinib in ALK-rearranged adenocarcinomas of the lung: report of two cases and review of the literature. *Lung Cancer* **93**, 55–58. http://dx.doi.org/10.1016/j.lungcan.2016.01.002.

[48] Bordi P, Tiseo M, Rofi E, Petrini I, Restante G, Danesi R, and Del Re M (2017). Detection of ALK and KRAS mutations in circulating tumor DNA of patients with advanced ALK-positive NSCLC with disease progression during crizotinib treatment. *Clin Lung Cancer* **18**, 692–697. http://dx.doi.org/10.1016/j.cllc.2017.04.013.

[49] Nikiforova MN, Wald AI, Roy S, Durso MB, and Nikiforov YE (2013). Targeted next-generation sequencing panel (ThyroSeq) for detection of mutations in thyroid cancer. *J Clin Endocrinol Metab* **98**, E1852-1860. http://dx.doi.org/10.1210/jc.2013-2292.

[50] Choudhary A, Mambo E, Sanford T, Boedigheimer M, Twomey B, Califano J, Hadd A, Oliner KS, Beaudenon S, and Latham GJ, et al (2014). Evaluation of an integrated clinical workflow for targeted next-generation sequencing of low-quality tumor DNA using a 51-gene enrichment panel. *BMC Med Genomics* **7**, 62. http://dx.doi.org/10.1186/s12920-014-0062-0.

[51] Beadling C, Wald AI, Warrick A, Neff TL, Zhong S, Nikiforov YE, Corless CL, and Nikiforova MN (2015). A multiplexed amplicon approach for detecting gene fusions by next-generation sequencing. *J Mol Diagnostics* **18**, 1–12. http://dx.doi.org/10.1016/j.jmoldx.2015.10.002.

[52] Vendrell JA, Taviaux S, Béganton B, Godreuil S, Audran P, Grand D, Clermont E, Serre I, Szablewski V, and Coopman P, et al (2017). Detection of known and novel ALK fusion transcripts in lung cancer patients using next-generation sequencing approaches. *Sci Rep* **7**, 1–11. http://dx.doi.org/10.1038/s41598-017-12679-8.

[53] Hovelson DH, McDaniel AS, Cani AK, Johnson B, Rhodes K, Williams PD, Bandla S, Bien G, Choppa P, and Hyland F, et al (2015). Development and validation of a scalable next-generation sequencing system for assessing relevant somatic variants in solid tumors. *Neoplasia* **17**, 385–399. http://dx.doi.org/10.1016/j.neo.2015.03.004.