Low frequency KRAS mutations in colorectal cancer patients and the presence of multiple mutations in oncogenic drivers in non-small cell lung cancer patients

Liyan Jiang\textsuperscript{a,1}, Jiaqi Huang\textsuperscript{b,1}, Chris Morehouse\textsuperscript{b,1}, Wei Zhu\textsuperscript{b}, Susana Korolevich\textsuperscript{b}, Dan Sui\textsuperscript{c}, Xiaoxiao Ge\textsuperscript{c}, Kim Lehmann\textsuperscript{b}, Zheng Liu\textsuperscript{b}, Christine Kiefer\textsuperscript{b}, Meggan Czapiga\textsuperscript{b}, Xinying Su\textsuperscript{c}, Philip Brohawn\textsuperscript{b}, Yi Gu\textsuperscript{c}, Brandon W. Higgs\textsuperscript{b}, Yihong Yao\textsuperscript{b,*}

\textsuperscript{a}Department of Pulmonary Medicine, Shanghai Chest Hospital, Shanghai Jiao Tong University, Shanghai, China; \textsuperscript{b}MedImmune, LLC, Gaithersburg, MD, USA; \textsuperscript{c}Innovation Center China, AstraZeneca Global R&D, Shanghai, China

Intratumor heterogeneity can confound the results of mutation analyses in oncodriver genes using traditional methods thereby challenging the application of targeted cancer therapy strategies for patients. Ultradeep sequencing can detect low frequency and expanded clonal mutations in primary tumors to better inform treatment decisions. KRAS coding exons in 61 treatment-naive colorectal cancer (CRC) tumors and KRAS, EGFR, ALK, and MET in lung tumors from three Chinese non-small cell lung cancer (NSCLC) patients were sequenced using ultradeep sequencing methods. Forty-one percent of CRC patients (25/61) harbored mutations in the KRAS active domain, eight of which (13%) were not detected by Sanger sequencing. Three (of eight) had frequencies less than 10\% and one patient harbored more than one mutation. Low frequency KRAS active (G12R) and EGFR kinase domain mutations (G719A) were identified in one NSCLC patient. A second NSCLC patient showed an EML4-ALK fusion with ALK, EGFR, and MET mutations. A third NSCLC patient harbored multiple low frequency mutations in KRAS, EGFR, and MET as well as ALK gene copy number increases. Within the same patient, multiple low frequency mutations occurred within a gene. A complex pattern of intrinsic low frequency driver mutations in well-known tumor oncogenes may exist prior to treatment, resulting in resistance to targeted therapies. Ultradeep sequencing can characterize intratumor heterogeneity and identify such mutations to ultimately affect treatment decisions.

Keywords: Ultradeep sequencing, non-small cell lung cancer, colorectal cancer, oncogene, low frequency mutations

One of the major paradigm shifts over the past 15 years in anticancer therapy is the introduction of targeted therapy. This approach differs from nonspecific first-line treatments by focusing on activated protein targets related to cellular growth and proliferation and antiapoptotic processes. Such treatment strategies have proven to be highly effective, specifically when applied to aggressive cancers in certain patients, although, typically, clinical improvement is not sustained. Recent findings show that low frequency variants within the genes that code for these targets may
potentially explain this lack of sustained benefit in patients (1).

Epidermal growth factor receptor (EGFR) is one important target for cancer therapy. Monoclonal antibodies (mAb) such as cetuximab (Erbitux) (Bristol-Myers Squibb, New York City, NY and Eli Lilly, Indianapolis, IN) and panitumumab (Vectibix) (Amgen, Thousand Oaks, CA), which are directed against EGFR, have proven to be effective in the treatment of metastatic colorectal cancer (mCRC). In addition, EGFR tyrosine kinase inhibitors (TKIs) such as gefitinib (IRESS) (AstraZeneca, London, UK) and erlotinib (Tarceva) (Genentech, San Francisco, CA) have shown clinical efficacy by inhibiting the activating mutations in EGFR in patients with non-small cell lung cancer (NSCLC). More recently, the presence of a fusion gene, echinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase (EML4-ALK), has been identified as the driver mutation in a subgroup of NSCLC patients (2). The clinical features of lung cancer patients that harbor EML4-ALK include light- or never-smokers, younger age, the presence of adenocarcinomas with acinar pattern or signet ring adenocarcinoma, and a lack of EGFR or Kirsten rat sarcoma viral oncogene homologue (KRAS) gene mutations (3). Subsequent studies have led to the approval of crizotinib (XALKORI) (Pfizer, New York, NY) therapy in patients harboring this variant (4). The average frequency of this gene fusion across different ethnic populations is reported to be approximately 3.4% in unselected NSCLC patients and 4.5% in adenocarcinoma-enriched NSCLC patients (5), with a higher frequency in the Chinese population in both unselected NSCLC patients (4.9–11.7%) and in adenocarcinoma-enriched NSCLC patients (5.3–16.1%) (6,7).

KRAS is an important driver in the rat sarcoma/mitogen-activated protein kinase (RAS/MAPK) pathway that links EGFR activation to cell proliferation and survival. The associations among certain KRAS mutations and the effectiveness of anti-EGFR treatments in patients with CRC have been reported in numerous studies (8–12). Mutations affecting KRAS G12 or G13 residues cause over-activation of the RAS/MAPK pathway, and as signal transduction is activated at the level of KRAS proteins, upstream inhibition by EGFR-targeted agents becomes ineffective. Epidemiologic studies suggest that KRAS mutations may be a negative predictive biomarker for treatment with EGFR-TKIs in NSCLC as well as CRC (13), though due to the mutually exclusive relationship between KRAS and EGFR mutations in NSCLC and CRC, it is generally believed that KRAS mutations will have minimal impact on the effectiveness of EGFR-TKI treatment in NSCLC patients (13).

Currently, physicians depend on traditional mutation-detection strategies to guide treatment decisions with targeted therapies—the gold standard for assays being Sanger sequencing, allelic-specific polymerase chain reactions (PCRs) (e.g., amplification refractory mutation system [ARMS]-PCR), or fluorescence in situ hybridization (FISH). For example, KRAS mutation status is required in all mCRC patients who may receive anti-EGFR mAb therapy, and the detection of the activating mutations within the EGFR gene in NSCLC patients is a routine diagnostic determinant for gefitinib or erlotinib treatment. Additionally, the Vysis ALK Break Apart FISH Probe Kit (Abbott Molecular, Des Plaines, IL), which detects rearrangements of the ALK gene, was approved as a companion diagnostic test for crizotinib treatment. However, even within patient subgroups harboring such variants, not all respond to the targeted treatment. The first-line treatment response rates of EGFR-TKIs in patients with NSCLC containing EGFR mutations are 55–90% (14), and the responding patients eventually develop resistance to the targeted therapy. Either secondary mutations of EGFR or the amplification of c-MET (hepatocyte growth factor receptor) are responsible for only 50% of these EGFR-TKI resistance cases, leaving unknown molecular attribution for roughly half of the remaining patient population. Other biomarkers, such as phosphatidylinositol 4,5-bisphosphate 3-kinase, catalytic subunit alpha (PIK3CA) mutations, and loss of phosphatase and tensin homologue (PTEN), have been investigated, and their impacts on clinical efficacies of anti-EGFR therapies are not conclusive (8,15).

Traditional mutation-detection strategies reveal only mutations occurring in most cancer cells, namely, the expanded clonal population, while neglecting low frequency substitutions. Further, the technical limitations of detection sensitivity in traditional sequencing procedures have typically prevented the detection of these low frequency mutations. In principle, next-generation sequencing (NGS) technologies can provide a valid solution, as they rely on amplification and sequencing of distinct DNA molecules, and since sensitivity of NGS increases with depth and coverage, rare mutations are more readily detectable with ultradeep resequencing of a given DNA region. The detection of these low frequency mutations, in addition to expanded clonal mutations, is critical to understanding the mechanisms that resist cancer targeted therapy.

To further explore the presence and potential impact that low frequency somatic mutations in oncogenic driver genes may have in the resistance to targeted therapies, we sequenced the coding exons of KRAS in 61 treatment-naïve CRC patients using targeted ultradeep resequencing. With the extraordinary depth used here, we found that three patients (5%) harbored low frequency mutations (variant frequencies [Vfs] <10%) and eight patients had Vfs less than 27% in guanosine triphosphate (GTP)–ase nucleotide binding sites that were not detected using Sanger sequencing. We also sequenced the coding exons of EGFR, KRAS, ALK, and MET in tumor biopsies from three Chinese NSCLC patients. We found the coexistence of a low frequency KRAS active mutation (G12R) and EGFR kinase domain mutation (G719A) in tumor specimens in one NSCLC patient, an EML4-ALK fusion with ALK, EGFR, and MET mutations in the second patient, and multiple coexisting low frequency mutations in KRAS, EGFR, and MET with an ALK gene copy number increase in the third patient. We also found that multiple low frequency mutations were present in a single gene from the same CRC or NSCLC patients. These data suggest that intrinsic low frequency driver mutations, sometimes on different oncodrivers, in cancer tissues may exist prior to treatment, providing direct evidence for the cause of unsustainable clinical improvement in cancer patients undergoing monospecific targeted therapies.

Materials and methods

CRC and NSCLC sample summary

Sixty-one fresh frozen colon rectum tumors and six normal colon specimens were purchased from ILSbio (Chesterstown, PA), AstraZeneca, San Francisco, CA, Amgen, Thousand Oaks, CA, and Genentech, San Francisco, CA. There were a total of 52 mCRC patients with a median age of 70 years. The Clinical Research Office at the University of Pennsylvania (Philadelphia, PA) performed the collection and grouping of CRC and NSCLC samples. These samples were used to confirm the diagnosis and to ensure that the samples were from patients with CRC or NSCLC. The final sample was purchased from the same patient who had previously undergone monospecific targeted therapy.
MD) and Asterand (Detroit, MI). All tumors were from treatment-naive patients diagnosed with stage I to IIIC adenocarcinoma of the colon or rectum. Tumor specimens were from white patients, with 28 females and 33 males and an age range of 32–84 years. Normal tissue specimens were from Vietnamese males with ages ranging from 10 to 42 years. Of the 61 CRC patients, the coding regions of KRAS were sequenced at least twice, inclusive of independent library preparations and sequencing runs in 21 patients (eight patients sequenced at least three times) on the Illumina HiSeq sequencing platform (San Diego, CA). In addition, samples from a subset of 15 CRC patients from the set of 61 who harbored variants in both pathogenic (i.e., residues 10–17) and nonpathogenic loci of KRAS (Vf range 1.01–51.12%) were sequenced again using a completely independent sequencing platform, Ion Personal Genome Machine (PGM) (Life Technologies, Carlsbad, CA), to verify variant calls made initially using the Illumina HiSeq platform.

The NSCLC specimens were collected from three treatment-naive patients from Shanghai Chest Hospital. For patients A and C, two independent biological specimens were procured, and each was replicated in two independent library preparations and sequencing runs. For patient B, there was only a single biological specimen procured, and this was not replicated by sequencing. The study protocol and all NSCLC specimens were approved by the ethics committee at Shanghai Chest Hospital.

DNA extraction, library preparation, and sequencing

DNA extractions from fresh frozen tissue were performed using the Qiagen DNeasy Blood and Tissue Kit (German-town, MD) following the manufacturer’s suggested protocol. DNA extractions from formalin-fixed paraffin-embedded (FFPE) tissue were performed using the Ambion RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies) following the manufacturer’s instructions.

Target sequencing libraries for Illumina HiSeq sequencing were prepared using the Fluidigm Access Array System (San Francisco, CA) using multiplexed sample barcoding and amplicon tagging following the manufacturer’s instruction. Targeted libraries were prepared for Ion PGM sequencing using Life Technologies’ AmpliSeq protocol for the KRAS gene.

The Illumina sequencing libraries were normalized and run on the Illumina HiSeq 2000 Instrument using a 100–base pair-paired-end sequencing read protocol. Targeted libraries were also prepared by AmpliSeq with the same DNA samples using the Ion OneTouch 2 System (Life Technologies) and were sequenced on the Ion PGM 318 Chip (Life Technologies).

Dilution series to determine limit of detection

A selection of four CRC tumor DNA samples (from the cohort of 61 patients) harboring known cancer-specific mutations in KRAS at pathogenic loci (protein codons 12 and 13; DNA: chr12: 25398284 and chr12: 25398281, respectively, GRCh37/hg19) and for normal healthy colon specimens were diluted serially in a normal human whole blood DNA sample background in a 1:2 fashion for seven iterations to establish a titration from the original Vf. This titration method brought the expected Vf to below 1% for each of the samples. Each of the stock and diluted samples were then prepared using the specific amplicons in KRAS, mothers against decapentaplegic homologue 4 (SMAD4), tumor protein p53 (TP53), mutL homologue 1 (MLH1), kinase insert domain receptor (KDR), fibroblast growth factor receptor 1 (FGFR1), and v-kinase Hardy-Zuckerman 4 feline sarcoma viral oncogene homologue (KIT) in the TruSeq Amplicon Cancer Panel (Illumina) following the manufacturer’s protocol. Samples were sequenced on the Illumina MiSeq instrument with 150–base pair paired-end sequencing. Vfs were detected and the expected KRAS VFs were calculated from the dilution series to establish limits of detection for the sequencing by synthesis chemistry.

FISH

The EML4–ALK fusion was detected by FISH assays on 4-micron FFPE sections. FISH probes were generated internally by directly labeling bacterial artificial chromosome (BAC) RP11-100C1 (Life Technologies) DNA with Spectrum Red (cat no. 02N34-050) (Enzo Life Sciences, Farmingdale, NY) for the ALK gene and BAC RP11-142M12 (Life Technologies) with Spectrum Green (cat no. 02N32-050) (Enzo Life Sciences) for the EML4 gene using a nick translation kit (cat no. 07J00-001) (Abbott Molecular) based on methods according to the manufacturer’s instructions. Tissue MicroArray (TMA) sections were deparaffinized and pretreated using the SpotLight Tissue Kit (cat no.00-8401) (Life Technologies) according to the manufacturer’s instructions. Tissue sections and ALK/EML4 probes were then codenatured at 80°C for 5 minutes and hybridized at 37°C for 48 hours. Excess probe was removed with posthybridization wash buffer (0.3% NP40/1 × saline-sodium citrate [SSC]) by washing the slides at 75°C for 5 minutes, followed by a wash with 2 × SSC at room temperature for 2 minutes. Sections were then counterstained with 0.3 µg/mL 4′,6-diamidino-2-phenylindole (DAPI, H-1200) (Vector Laboratories, Burlingame, CA), cover-slipped, and stored at 4°C until signal observation.

ALK and EML4 signals were observed under a fluorescence microscope (BX61) (Olympus, Tokyo, Japan) using a 100 × objective and proper filters. Co-localization of red ALK signal and green EML4 signal was defined as EML4–ALK fusion positive. ALK gene copy number increase was defined as an average copy number greater than four.

Sanger sequencing for KRAS exon 2

Sanger sequencing of KRAS exon 2 was performed to assess codon 12 and 13 mutations in 61 CRC samples and 2 cell line controls (Calu-1 and HCT116). Thirty nanograms of genomic DNA from each sample was amplified in duplicate 20 µL reactions containing 0.5 µmol/L forward and reverse primers (Life Technologies) and 10 µL HotStar Taq Plus 2 × Master Mix (Qiagen). The PCR program was as follows: initial denaturation at 95°C (5 min); 35 amplification cycles of 94°C (30 s), 59°C (30 s), 72°C (1 min); and a final extension at 72°C (10 min). Amplification products were purified using the QIAquick PCR purification kit (Qiagen) and eluted in
30 μL TE buffer (10 mM TrisCl, 1 mM EDTA, pH 8.0). Sequencing reactions were performed using 2 μL template with 0.3 μmol/L forward or reverse primer and 2 μL BigDye Terminator v3.1 Ready Reaction Mix in a 10 μL reaction (Life Technologies). The sequencing program was as follows: 25 cycles of 96°C (10 s), 50°C (5 s), 60°C (4 min). Extension products were purified using the BigDye XTerminator purification protocol and run on an Applied Biosystems 3730 DNA Analyzer (Life Technologies). All samples were sequenced in the forward and reverse directions (forward primer, 5’ CGATACAGCTCTGGAGTCAAC 3’; reverse primer, 5’ CCTGACATACTCCCCAGGAAAG 3’).

Data analysis

Following demultiplexing, FASTQ files generated from either the Illumina HiSeq or MiSeq sequencing runs were aligned to the human genome (UCSC hg19; Feb 2009 release: Genome Reference Consortium GRCh37) using Bowtie2 (version 2.0.0-beta7) (16). For Ion Torrent sequencing runs, the aligned Binary Alignment/Map (BAM) files from the Torrent Mapping Alignment Program (TMAP) were used in the downstream analysis pipeline. SAMtools (version 0.1.18) was utilized to convert the aligned sequence data from a Sequence Alignment/Map (SAM) to BAM format. The data was subsequently sorted and indexed with SAMTools software. The SAMtools mpileup function was used to summarize base calls at each locus. Single nucleotide polymorphism (SNP) and insertion and deletion (INDEL) calls were made with VarScan2 (version 2.3.2). SNPs and INDELs were characterized as being significantly different from the reference sequence if the variant to reference base frequency was greater than 1%, the calculated VarScan2 \( P \) value for variant calls (based on the Fisher exact test) was less than 0.001, total minimum depth was greater than 500, the variant minimum depth was greater than 5, and the base quality value was greater than 30. All identified variants within a particular sample were exported as a variant call format (VCF) (version 4.1) file. VCFs were examined with snpEff (version 3.0) for variant annotation and prediction of variant effects on genes. Further data analyses, including change-point regression for the titration experiment (segmented R library) and graphs, were all created in R (version 2.15.1) and Prism (version 5.10) (GraphPad Software, La Jolla, CA).

Results

Variant call limit of detection: titration experiment

To determine the level of detection for identifying rare somatic mutations with targeted ultradep sequencing, two different experiments were performed. The first was a dilution experiment designed to establish (i) where the break in linearity occurred between Vfs and a titrated concentration of tumor and (ii) the limit of detection for variant calling against a negative control background. Colon tissue specimens from four normal healthy donors were used as negative controls and four different CRC patient specimens were selected that harbored cancer specific mutations with high Vfs (>30%). These loci within the first nucleotide-binding domain were used, since variants in this region are well characterized as cancer hot spots with a large dynamic range of Vfs, while they also provide an adequate number of data points at the low Vf values after conducting serial dilutions. For each dilution factor within each specimen (four CRC and four healthy controls), healthy donor DNA from whole blood was added at 1:2 serial dilutions, providing a range of 0.0078–1 dilution factors. Then the loci for each specimen were sequenced using the Illumina MiSeq protocol. All samples had depth greater than 3,000 ×, and variants were called as described in the Materials and methods section.

The results for this experiment are provided in Figure 1A and Supplementary Figures S1 and S2. The difference between the upper range of Vfs in the normal specimens and closest CRC dilution data point occurs at a Vf of 1%. Additionally, change-point analysis calculated in a regression between the relative error \((|Vf_{\text{observed}}-Vf_{\text{expected}}|/Vf_{\text{expected}})\) and variant index for each patient specimen indicates that the break in linearity of the titration curve occurs at a Vf greater than 1% (data not shown).

To show that this Vf threshold was not specific only to KRAS, regions of six well-characterized cancer-associated genes (SMAD4, TP53, MLH1, KDR, FGFR1, and KIT) were sequenced in the four CRC tumor patient specimens and four normal colon tissue specimens using the same dilution factors as for KRAS, and the Vfs of germline SNPs were calculated. The germline SNPs served as variants with high Vfs, akin to the pathogenic loci in KRAS used in the titration study described previously. If the CRC or normal subject specimen harbored the minor allele of the germline SNP within the gene, there was a clear linear association between the Vf and the dilution factor. If the CRC or normal subject did not harbor the minor allele of the germline SNP (i.e., it was homozygous to reference), the mean Vf plus one standard deviation (SD) of the Vfs were below a Vf threshold of 1% at all dilution levels and all genes, with the exception of the first dilution level in KDR, which was slightly higher than 1% but less than 2% (Supplementary Figure S3). This study demonstrates that the Vf threshold identified is not specific to KRAS and can be applied to other genes.

Variant call limit of detection: reproducibility experiment

The second experiment was based on replication of variants identified using two independent sequencing protocols—Illumina HiSeq and Ion PGM. This strategy was developed to show what variant calls exceed sequencing error rates (Illumina sequencing instruments range from 0.05 to 1%) and show reproducibility using an independent sequencing technology (17–20). We selected 15 CRC patient specimens with Vfs ranging from 1.01 to 51.12% from the Illumina HiSeq protocol to be rerun on the Ion PGM platform, from the pool of 61. The same variant caller as described in the Materials and methods section was used for both protocols. Of the 15 CRC patients, all KRAS coding variants identified with Vfs greater than 2% from the Illumina HiSeq protocol reproduced using the Ion PGM protocol (eight variants within seven patients). Figure 1B demonstrates the
Vfs for both sequencing platforms for just those variants called in codon 12 (six patients). Those variants originally identified using the Illumina HiSeq protocol with Vfs ranging from 1.01 to 1.89% did not reproduce using the Ion PGM protocol. Though there are inherent differences in sequencing sensitivity for both platforms, those variants called with Vfs greater than 2% on the Illumina HiSeq reproduced on the Ion PGM platform. This result, in combination with the dilution experiment, supports a conservative threshold of 2% or greater for a true positive variant call, which was implemented for all subsequent analyses reported in this study.

**Figure 1** (A) Observed Vfs for four normal healthy control participants (mean ± 1SD) and four CRC participants for two KRAS pathogenic loci (codons 13 and 12; DNA 25,398,281 and 25,398,284, GRCh37/hg19) across an eight-level 1:2 serial dilution series. The x- and y-axes have been transformed with a square root to better illustrate the lower Vf detection limit. The dashed line indicates a Vf = 1%, and lines for the four CRC participants are calculated using a least squares fit. (B) Reproducibility of variant frequencies (Vfs ≥ 2%) for six CRC participants on two independent detection platforms at KRAS DNA locus 25,398,284; HiSeq (blue bars) and Ion PGM (red bars). For a combined plot of CRC participant specimens for both loci in (A) on the original Vf scale, see Supplemental Figures 1 and 2.

**Error rates at variant calling threshold**

We next evaluated the error rates at a 2% or greater Vf threshold for variant calling. The false-positive rate is identified by summating the number of variants called at the Vf threshold within a negative control sample, or a sample that would not be expected to harbor variants. Six normal healthy donor colon specimens (two in addition to the four used in the titration experiment) were used for this determination, as beyond de novo mutations, normal colon tissue from healthy donors would not be expected to contain somatic variants. The most prevalent nucleotide alternative to the reference
Low frequency KRAS mutation was observed in tumors from treatment-naive CRC patients

A total of 39 (63%) CRC patients harbored five unique non-synonymous coding mutations, with 25 patients (41%) carrying the known active mutations in the first nucleotide-binding domain (residues 10–17) (Table 1). Eight CRC patients (13%) harbored mutations in the GTPase nucleotide-binding sites of KRAS that were not detected using Sanger sequencing, and three of eight had Vfs less than 10%, with one patient harboring more than one mutation (the distance between these two mutations prohibited determining a cis or trans association). This information is illustrated in Figure 2.

Coexistence of low frequency active KRAS and EGFR kinase domain mutations in a treatment-naive Chinese NSCLC patient

The coding exons within KRAS and EGFR were sequenced in a tumor biopsy from a treatment-naive Chinese NSCLC patient (patient A). After we excluded germline polymorphisms, four nonsynonymous single nucleotide changes were identified in EGFR (Supplemental Table 1). One of these mutations was located in the extracellular domain, one in the kinase domain, and two in the C-terminal cytoplasmic domain. The kinase domain mutation is a well-characterized active mutation of EGFR (G719A), which has been associated with sensitivity to EGFR-TKI drugs such as gefitinib and erlotinib in NSCLC patients. Three (of four) of these mutations were predicted to impact protein function, from the SIFT database (21) and the Vfs ranged from 3.23 to 4.34%, with the exception of that of the active mutation, which was 9.89%. EGFR variants in codons 1016 and 1034 were determined to be in trans. For this same NSCLC patient, we also observed a single G12R mutation in KRAS with a low Vf of 2.28%. The entire coding exons of ALK and MET were also sequenced, revealing one novel nonsynonymous mutation in the cytoplasmic domain of ALK with a Vf of 4.41%, while one mutation in the cytoplasmic domain of the MET gene was detected (Vf = 2.07%) (Figure 3).

Coexistence of EML4-ALK fusion or ALK copy number increase and EGFR, MET, and KRAS mutations in tumor tissue from two treatment-naive Chinese NSCLC patients

Two treatment-naive Chinese NSCLC patients were tested for an EML4-ALK fusion using FISH analyses. Patient B was positive for the EML4-ALK fusion (Figure 4A), and patient C was identified with an ALK gene copy number increase, with an average copy number of approximately 5 (Figure 4B). To evaluate the mutations related to crizotinib resistance, all coding exons of KRAS, EGFR, ALK, and MET were sequenced in the tumors from these two NSCLC patients. The summary of the mutation status for patient B is summarized in Supplemental Table 2. Four novel nonsynonymous mutations in ALK were identified in this patient; each of them was located in the extracellular domain of this gene. One of these mutations, L170P, was predicted to have damaging effects on protein function, according to the SIFT database.

Eight novel nonsynonymous mutations were identified in EGFR in this patient, with three of them predicted to have damaging effects on protein function, according to the SIFT database. Among these eight, the homozygous stop code mutation E66* was located in the N-terminus of EGFR and the homozygous deleterious mutation R832C was found in the kinase domain of EGFR. One novel heterozygous mutation was also identified in the extracellular domain of MET. The Vfs of the mutations identified in patient B ranged from 10.35 to 99.85%. Variants in close proximity were assessed for cis or trans relationships, or occurrence of shared variants within the same clone. It was determined that the variants in codons 170 and 173 in EGFR were in cis, while the variants in codons 173 and 182 were in trans.

The summary of the mutation status for patient C is provided in Supplemental Table 3. One nonsynonymous mutation was identified in ALK that was predicted to have damaging effects on protein function (SIFT database). This variant at codon 170 and the other variant in this gene at codon 198 were determined to be in trans. Two nonsynonymous mutations were detected in EGFR. Among them, one is located in the extracellular domain and is predicted to have damaging effects on protein function (SIFT database), while the other is in a serine-rich domain and predicted to have damaging effects on protein function. One nonsynonymous mutation was found in the C-terminal region of KRAS, R164*, which is a stop codon mutation and results in a truncated KRAS protein. One nonsynonymous mutation was found in MET, located in the extracellular domain. The Vfs of the five nonsynonymous mutations in patient C ranged from 2.08 to 3.84%. Although four of these mutations each had an entry in the Single Nucleotide Polymorphism database (dbSNP), we classified these SNPs as cancer-specific mutations with the following considerations: (i) the Vfs of
### Table 1  Summary of mutations identified within the first nucleotide-binding domain (residues 10–17) in KRAS in tumor specimens from 25 CRC patients

<table>
<thead>
<tr>
<th>Patient</th>
<th>Chr</th>
<th>Physical position</th>
<th>Genotype</th>
<th>Codon</th>
<th>Amino acid change</th>
<th>Sanger result</th>
<th>Ref Reads</th>
<th>Var reads</th>
<th>Variant frequency (%)</th>
<th>dbSNP ID</th>
<th>Confirmed by Ion PGM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>249584</td>
<td>176771</td>
<td>41.45</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>25398281</td>
<td>C/T</td>
<td>13</td>
<td>G/D</td>
<td>G/D</td>
<td>294912</td>
<td>137593</td>
<td>31.72</td>
<td>rs112445441</td>
<td>Not examined</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>G/V</td>
<td>167032</td>
<td>146537</td>
<td>46.62</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
<tr>
<td>8</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>Negative</td>
<td>323608</td>
<td>107178</td>
<td>24.85</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>10</td>
<td>12</td>
<td>25398279</td>
<td>C/T</td>
<td>14</td>
<td>V/I</td>
<td>Negative</td>
<td>466547</td>
<td>43473</td>
<td>8.52</td>
<td>rs104894365</td>
<td>Yes</td>
</tr>
<tr>
<td>13</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>225479</td>
<td>265428</td>
<td>54.05</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td>25398281</td>
<td>C/T</td>
<td>13</td>
<td>G/D</td>
<td>G/D</td>
<td>303499</td>
<td>154506</td>
<td>33.72</td>
<td>rs112445441</td>
<td>Not examined</td>
</tr>
<tr>
<td>16</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>280184</td>
<td>112988</td>
<td>28.73</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>17</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>268793</td>
<td>210304</td>
<td>43.88</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
<tr>
<td>18</td>
<td>12</td>
<td>25398285</td>
<td>C/T</td>
<td>12</td>
<td>G/S</td>
<td>G/S</td>
<td>185069</td>
<td>82814</td>
<td>30.89</td>
<td>rs121913530</td>
<td>Not examined</td>
</tr>
<tr>
<td>19</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>Negative</td>
<td>330901</td>
<td>10386</td>
<td>3.04</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>387693</td>
<td>139363</td>
<td>26.44</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>21</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>G/V</td>
<td>142621</td>
<td>317999</td>
<td>68.79</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>22</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>Negative</td>
<td>286307</td>
<td>89578</td>
<td>23.79</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>23</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>267279</td>
<td>116448</td>
<td>30.33</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
<tr>
<td>24</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>Negative</td>
<td>381445</td>
<td>10224</td>
<td>2.61</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
<tr>
<td>25</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>269119</td>
<td>214402</td>
<td>44.33</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>30</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>G/V</td>
<td>312311</td>
<td>232101</td>
<td>42.51</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>31</td>
<td>12</td>
<td>25398285</td>
<td>C/A</td>
<td>12</td>
<td>G/C</td>
<td>G/C</td>
<td>187782</td>
<td>234737</td>
<td>55.06</td>
<td>rs121913530</td>
<td>Not examined</td>
</tr>
<tr>
<td>32</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>G/V</td>
<td>263896</td>
<td>141116</td>
<td>33.13</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>33</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>Negative</td>
<td>412155</td>
<td>64179</td>
<td>13.45</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>35</td>
<td>12</td>
<td>25398281</td>
<td>C/T</td>
<td>13</td>
<td>G/D</td>
<td>G/D</td>
<td>168767</td>
<td>344810</td>
<td>67.09</td>
<td>rs112445441</td>
<td>Not examined</td>
</tr>
<tr>
<td>36</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>Negative</td>
<td>276673</td>
<td>100422</td>
<td>26.56</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>37</td>
<td>12</td>
<td>25398284</td>
<td>C/A</td>
<td>12</td>
<td>G/V</td>
<td>Negative</td>
<td>378649</td>
<td>76508</td>
<td>16.78</td>
<td>rs121913529</td>
<td>Not examined</td>
</tr>
<tr>
<td>38</td>
<td>12</td>
<td>25398284</td>
<td>C/T</td>
<td>12</td>
<td>G/D</td>
<td>G/D</td>
<td>180748</td>
<td>189179</td>
<td>51.12</td>
<td>rs121913529</td>
<td>Yes</td>
</tr>
</tbody>
</table>
these mutations in the tumor specimens differed by at least 10-fold from the true germline SNPs in these same tumor specimens with allele frequencies in the range of 32.37–99.98%, and (ii) the allele frequencies of these SNPs are unknown in the dbSNP, and most of these SNPs were generated from the Cancer Genome Project, while the true germline SNPs in these tumor specimens have population data reported in dbSNP.

Discussion

Using ultradeep NGS, we were able to detect low frequency mutations in key oncodriver genes associated with treatment response to anti-EGFR therapy in patients with CRC or NSCLC. The sensitivity of NGS, its extreme depth, and its coverage allow such low frequency mutations to become detectable, with more than 10,000 × read depth supporting the variant calls (22). The tumor specimen is a heterogeneous, indistinguishable mixture of nonclonal somatic mutations and base call errors, the latter of which have been introduced during DNA amplification and sequencing, so ultradeep sequencing depth for each targeted gene is necessary to accurately call variants with frequencies that exceed the error level (17–20) with confidence. Additionally, here we show the importance of validating variant calls using a titration experiment as well as an independent sequencing platform to reproduce low frequency variant calls. Nevertheless, single cell PCR and sequencing may be the ultimate way to make the unambiguous validation of low frequency mutations, but this technology is still in its infancy and its utilization is beyond the scope of this study.

Figure 2 The distribution of mutations detected in 25 CRC patients across the coding region codons of the KRAS gene. Each point corresponds to a patient and each black tick mark in the orange protein coding block indicates the presence of at least one mutation identified in at least a single patient.

Figure 3 The VFs for mutations identified in patient A (blue), patient B (red), and patient C (green) tumor biopsy specimens in (A) MET, (B) KRAS, (C) EGFR, and (D) ALK. For each plot, the amino acid residues are on the x-axis, the VFs on the y-axis, and the patient ID on the z-axis.
A limitation of this study is the lack of a germline control for each CRC or NSCLC patient specimen, although at least two independent sequencing runs (including new library preparations) were conducted for 21 CRC patient specimens, and for two of the three the NSCLC patients, two independent sequencing runs were conducted on two different biological specimens (providing confidence in those variants that were consistently identified). In addition, most of the variants identified here are present in the Catalogue of Somatic Mutations in Cancer (COSMIC) database and have been associated with specific changes in tumors. De Grassi et al. compared low frequency mutations in nonneoplastic tissues from hereditary nonpolyposis colorectal cancer (HNPCC) patients and healthy donors and found that low frequency mutations were primarily associated with genomic instability. More important, these mutations were present only in the HNPCC patients and not in the healthy donors (22), suggesting that the novel low frequency mutations identified in this study may be associated with genomic instability in the tumor specimens of these patients even if such low frequency mutations are present in the germline.

It has been speculated that intratumor heterogeneity is at least partially responsible for the discordance in mutation status between different sites of a tumor, as well as varied response to certain cancer treatments. The traditional opinion that explains the manifestation and progression of human cancers starts with a DNA mutation in a single cell, followed by malignant cell clonal expansion and potential additional genetic aberrations. This continuing acquisition of genetic alterations can result in the emergence of tumor subclones with varying phenotypic advantages, such as invasion, proliferation, or the ability to colonize to different organs (23). The presence of more than one clone of cancer cells within a given tumor mass as well as the presence of different genetic alterations in different metastatic tumors from a single patient have been identified in several tumor types (24–26).

If resistant clones are present at low frequency in the primary tumor, a key question to address is whether they can be detected at an early stage, thus influencing the choice of primary therapy. Diaz et al. identified DNA mutations in the KRAS gene in the circulation of 28 patients receiving monotherapy with panitumumab and suggested that the mutations were preexisting in the tumors before the initiation of treatment (1). Our data indicate that approximately 13% of pretreatment CRC tumors harbor low frequency mutations (2.61–26.56%) in the GTPase nucleotide-biding site domain of KRAS that are not detected by traditional Sanger sequencing. In NSCLC patients, we also observed a low frequency KRAS active mutation (G12R) and an EGFR kinase domain mutation (G719A) in the same tumor specimen. These findings, along with those from Diaz et al., suggest that the resistance to EGFR inhibitor therapy in NSCLC patients could be the result of the expansion of cancer cells harboring low frequency KRAS mutations, which effectively inhibit EGFR-driven cancer clones.

Yang et al. reported that approximately 1.0% of NSCLC patients (4/398) has concomitant EGFR mutations and EML4-ALK fusions. Shaozhang et al. reported a concurrent KRAS mutation and EML4-ALK fusion in 1 of 8 Asian NSCLC patients (12.5%) using Sanger sequencing and reverse transcription PCR (RT-PCR), respectively (27). Here we identified an EML4-ALK fusion presenting with multiple mutations in ALK, EGFR, and MET in tumor specimens from a Chinese NSCLC patient. Further, we found multiple low frequency mutations in KRAS, EGFR, and MET that coexist with an ALK copy number increase in another Chinese NSCLC patient. However, most of the mutations in our study were present at very low frequencies—below those capable of being detected by these previous studies that used traditional Sanger sequencing. These results clearly indicate not only that mutations can present in multiple oncogenes but also that one gene from a patient can harbor several different mutations and the frequency differences among these mutations indicate that they likely arose from different clones of cells.

Current targeted cancer therapies usually lack durability and demonstrate limited overall efficacy in patients. The types of low frequency concurrent mutations in candidate oncogenes presented here suggest necessary modifications both to methods for detection of these variants and to general treatment strategies. To date, Sanger sequencing has been effectively used for detection of treatment-relevant somatic mutations. However, in a heterogeneous mixture of cancerous and normal tissue, Sanger sequencing will likely

---

**Figure 4** FISH detection of the EML4-ALK fusion using a Spectrum Green–labeled EML4 probe (green signal) and a Spectrum Red–labeled ALK probe (red signal). (A) The EML4-ALK fusion gene is indicated with a white arrow in patient B; (B) both EML4 and ALK gene copy number increases were increased in patient C.
fail to detect low frequency mutations. Beyond the evidence presented here, where VFs less than approximately 26% are not detected using Sanger sequencing, another study demonstrated that Sanger sequencing failed to identify EGFR mutations in primary lung tumor samples with approximately 10% VFs (28). ARMS-PCR assays, on the other hand, are superior to Sanger sequencing in both sensitivity and robustness on a large and diverse set of clinical tumor specimens, although the assays are effective in detecting known, well-characterized mutations. Since the coexistence of several driver mutations within different, or even the same, key oncogenes has such an impact on the success of targeted cancer therapy, standard mutation-detection strategies are not sufficient. More sensitive and cost-effective sequencing methods are required to systematically assess the mutation status within cancer pathway genes or at the whole genome level. Furthermore, because patients often develop resistance to targeted therapy over time that is due to the preexistence of low frequency mutations in oncogenes, treatment strategies based on combination therapy might prove to be the most optimal treatment approach for cancer patients. These hypotheses need to be rigorously evaluated in future cancer trials.

Acknowledgments

The work described in this article was supported by MedImmune, LLC. All authors with the MedImmune, LLC, affiliation are full-time employees and have stock in AstraZeneca.

Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cancergen.2013.09.004.

References