Genomic profiling of multiple sequentially acquired tumor metastatic sites from an "exceptional responder" lung adenocarcinoma patient reveals extensive genomic heterogeneity and novel somatic variants driving treatment response

Romi Biswas,[1,10] Shaojian Gao,[1,10] Constance M. Cultraro,[1,10] Tapan K. Maity,[1] Abhilash Venugopalan,[1] Zied Abdullaev,[2] Alexey K. Shaytan,[3] Corey A. Carter,[4] Anish Thomas,[1] Arun Rajan,[1] Young Song,[5] Stephanie Pitts,[1] Kevin Chen,[1] Sara Bass,[6] Joseph Boland,[6] Ken-Ichi Hanada,[7] Jinqiu Chen,[8] Paul S. Meltzer,[5] Anna R. Panchenko,[3] James C. Yang,[7] Svetlana Pack,[2] Giuseppe Giaccone,[9] David S. Schrump,[1] Javed Khan,[5] and Udayan Guha[1]

[1]Thoracic and Gastrointestinal Oncology Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA; [2]Laboratory of Pathology, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA; [3]National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, Maryland 20894, USA; [4]Walter Reed National Military Medical Center, Bethesda, Maryland 20889, USA; [5]Genetics Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA; [6]Cancer Genomics Research Laboratory, Division of Cancer Epidemiology and Genetics, National Cancer Institute, Gaithersburg, Maryland 20848, USA; [7]Surgery Branch, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA; [8]Laboratory of Cell Biology, Center for Cancer Research, National Cancer Institute, Bethesda, Maryland 20892, USA; [9]Lombardi Comprehensive Cancer Center, Georgetown University, Washington, D.C. 20057, USA

Corresponding author: udayan.guha@nih.gov

Ontology terms: neoplasm of the lung

Abstract We used next-generation sequencing to identify somatic alterations in multiple metastatic sites from an "exceptional responder" lung adenocarcinoma patient during his 7-yr course of ERBB2-directed therapies. The degree of heterogeneity was unprecedented, with ∼1% similarity between somatic alterations of the lung and lymph nodes. One novel translocation, *PLAG1-ACTA2*, present in both sites, up-regulated ACTA2 expression. *ERBB2*, the predominant driver oncogene, was amplified in both sites, more pronounced in the lung, and harbored an L869R mutation in the lymph node. Functional studies showed increased proliferation, migration, metastasis, and resistance to ERBB2-directed therapy because of L869R mutation and increased migration because of ACTA2 overexpression. Within the lung, a nonfunctional CDK12, due to a novel G879V mutation, correlated with down-regulation of DNA damage response genes, causing genomic instability, and sensitivity to chemotherapy. We propose a model whereby a subclone metastasized early from the primary site and evolved independently in lymph nodes.

[Supplemental material is available for this article.]

INTRODUCTION

Lung cancer is the most common cause of cancer-related death globally (Ferlay et al. 2013). A majority of lung cancers are categorized as non-small-cell lung cancer (NSCLC), and adenocarcinoma is the most prevalent NSCLC histology. Smoking-associated NSCLC is one of the cancers with the highest mutation rate because of the mutagens involved (Alexandrov et al. 2013; Vogelstein et al. 2013). However, many lung cancers are diagnosed in nonsmokers (Subramanian and Govindan 2007). Recent studies have shown that epidermal growth factor receptor (*EGFR*) tyrosine kinase (TK) domain mutations and chromosomal rearrangements involving the echinoderm microtubule-associated protein-like 4 and anaplastic lymphoma kinase (*EML4-ALK*) are prevalent in never-smoker lung adenocarcinoma patients (Shigematsu et al. 2005; Soda et al. 2007). Tumors harboring these changes often show dramatic initial responses to treatment with EGFR tyrosine kinase inhibitors (EGFR-TKIs) (Lynch et al. 2004) or to ALK kinase inhibitors (Soda et al. 2007; Takeuchi et al. 2012). This has generated interest in using molecular profiling to provide NSCLC patients with personalized treatment (Thomas et al. 2013). The rapidly decreasing cost of next-generation sequencing technologies has facilitated several recent high-throughput sequencing studies (Vignot et al. 2013) aimed at unveiling the genomic alterations that may drive NSCLC tumor development (Ding et al. 2008; Lee et al. 2010; Govindan et al. 2012; Imielinski et al. 2012; The Cancer Genome Atlas Research Network 2014). Studies have validated gene mutations and rearrangements commonly found in lung adenocarcinoma and have also uncovered new alterations that may be potential therapeutic targets (Imielinski et al. 2012; The Cancer Genome Atlas Research Network 2014). Although these studies use large cohorts, they rarely find a single event consistently linked to a majority of the cancer patients, which clearly points to the complexity of lung cancer development.

We used next-generation DNA sequencing to analyze whole genomes, exomes, or specific genes from a patient with stage IV lung adenocarcinoma over his 7-yr course of treatment. We initially sequenced the metastatic lung tumor and lymph node genomes, then performed exome sequencing on the original formalin-fixed paraffin-embedded (FFPE) diagnostic lung tumor biopsy, and used Ion Torrent targeted deep resequencing to analyze several additional metastatic lymph nodes procured by excisional biopsy at different time points, including at autopsy. Only two somatic nonsilent single-nucleotide variants (SNVs) in *PHKG1* and *CTSG* and one novel translocation involving *PLAG1* and *ACTA2* were present in both lung tumors and the metastatic lymph nodes. However, a majority of the somatic SNVs in the original lymph node biopsy were also present in subsequent metastatic lymph node biopsies, even several years later. Here we report several novel somatic alterations in some key cancer-related genes, and copy number gain and loss of genes involved in cell cycle progression and inhibition, respectively. Although the extent of genomic heterogeneity between the two tumor sites was quite high, it appears that the same key pathways and cancer hallmarks were impacted.

RESULTS

Case Description

A 50-yr-old African–American male never-smoker was diagnosed with stage IV lung adenocarcinoma in November 2007. Initially, he received standard first-line chemotherapy with limited success, followed by erlotinib, with no response. In December 2008, he enrolled in a National Cancer Institute (NCI) clinical trial for a second-generation irreversible panhuman EGFR tyrosine kinase inhibitor to which he responded, for a short duration. A case report was published (Kelly et al. 2010) and subsequently updated (Kelly et al. 2012). Here, we present a 7-yr treatment time line (Fig. 1A). Subsequent to progression on dacomitinib, he was
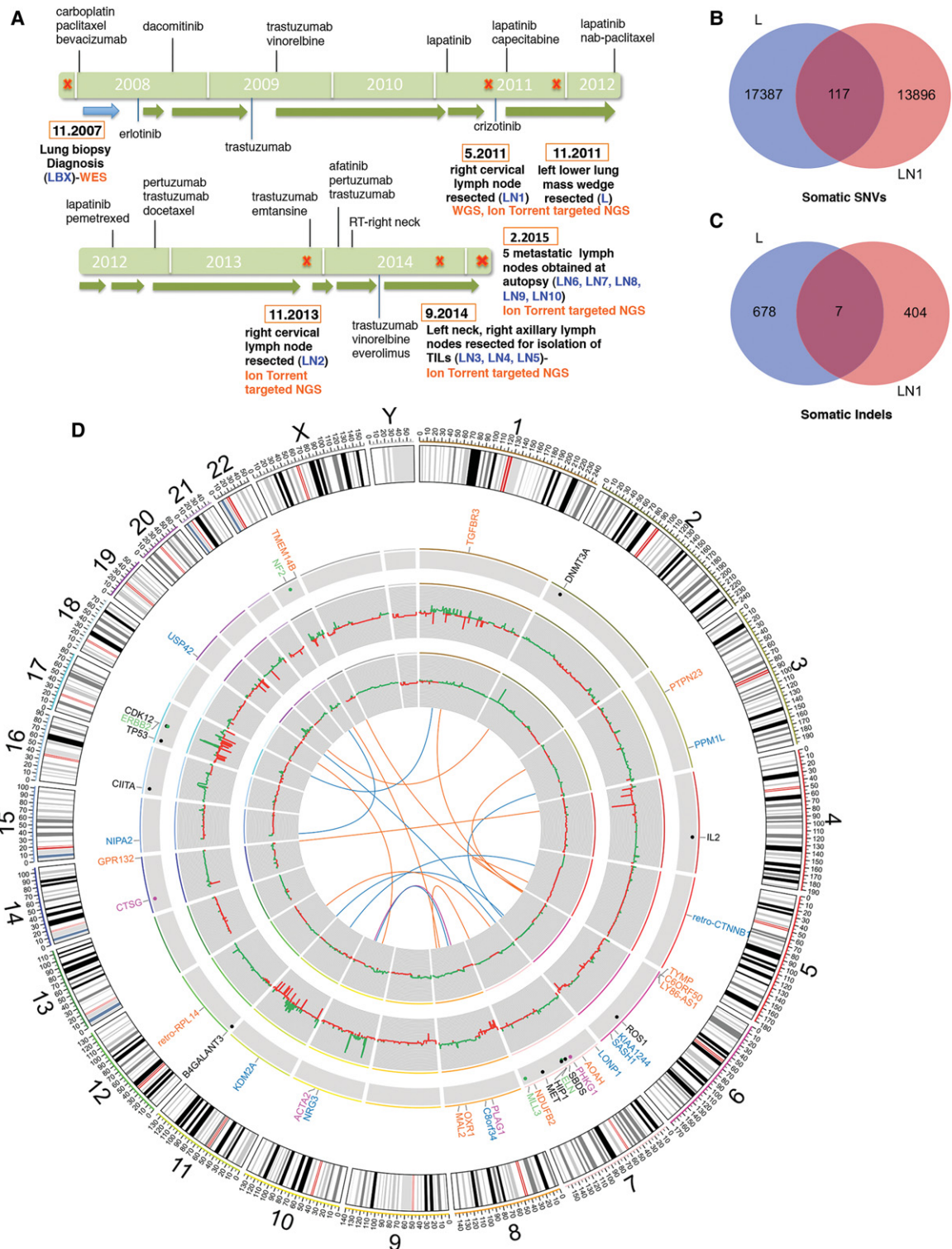
Figure 1. Treatment and tumor acquisition time line and summary of whole-genome sequencing (WGS) data analyses. (*A*) The time line of treatment history is indicated by the vertical lines along the green bar with names of the drug used at each approximate time point. (*Legend continues on following page.*)

treated for 6 wk with single agent trastuzumab and progressed slightly; then trastuzumab with vinorelbine for 18 mo; single agent lapatinib for 2 mo; a combination of lapatinib with successive use of capecitabine, nab-paclitaxel, pemetrexed for about a year; pertuzumab, combined with trastuzumab and docetaxel for another year; single agent trastuzumab emtansine with quick progression; trastuzumab combined with pertuzumab and afatinib with quick progression; and finally trastuzumab combined with vinorelbine and everolimus for about a year. As part of an NCI molecular profiling study, the patient underwent serial biopsies and had several metastatic lymph nodes removed along the course of his treatment (LN1-10; Fig. 1A). In addition, he underwent a lung wedge resection to eradicate a major progressive lesion in the left lower lobe (L). We obtained an FFPE specimen from the original computed tomography (CT)-guided lung tumor biopsy (LBX), which provided little tissue. Most importantly, the patient continued to have a good performance status and quality of life throughout his treatment. Unfortunately, in early 2015, he died of heart failure, a complication likely resulting from the preparative regimen for tumor-infiltrating lymphocyte (TIL) immunotherapy in another clinical study at the National Institutes of Health (NIH) Clinical Center. Interestingly, at autopsy no tumor tissue was found in his lung, suggesting his metastatic disease was eventually controlled in the lung. However, there was tumor infiltration in multiple lymph nodes, and five metastatic lymph nodes were obtained at autopsy for further studies (LN6-10; Fig. 1A). Contrary to all expectations, our patient survived >8 yr with metastatic lung adenocarcinoma and hence can be considered an "exceptional responder."

Summary of Whole-Genome and Whole-Exome Sequencing Data

We sought to identify temporal genomic alterations driving progression of specific metastatic sites during the course of treatment to determine how tumor heterogeneity impacts treatment response. First, we performed whole-genome sequencing (WGS) on two sequentially acquired metastatic sites; a right cervical lymph node biopsy from May 2011 (LN1) when his disease progressed on lapatinib and a lung wedge resection from November 2011 (L) after he again progressed on lapatinib combined with capecitabine (Fig. 1A). WGS was also performed on normal blood mononuclear cell DNA to exclude germline variants. We later performed whole-exome sequencing (WES) on DNA extracted from the original diagnostic FFPE LBX and whole blood as normal control to determine somatic alterations present before treatment was initiated.

Paired-end 100-bp reads were generated by Illumina's Genome Network Services through a contract with Knome Inc. Sequencing libraries from blood, lung tumor, and metastatic lymph node genomic DNA had mean estimated insert sizes of 326–343 bp with SD of 16–24 bp. A total of 126.7, 123.3, and 147.1 GB of sequencing data were generated for the normal, lung tumor, and metastatic lymph node samples, resulting in 91.2%, 95.1%, and

Figure 1. (*Continued*) The duration of the treatment and/or the response times are indicated by the lower horizontal arrows. Tumor biopsies were taken at each time point (red ✗) with the date (orange box); sample description below, sample names in parentheses. Venn diagrams show the common somatic single-nucleotide variants (SNVs) (*B*) and indels (*C*) identified by Strelka in lung tumor (L) and lymph node (LN1) metastasis genomes. (*D*) The Circos plot shows some of the validated key somatic alterations identified by WGS. The *outer* circle depicts each chromosome and its cytobands. Genes that carry key nonsynonymous mutations (dot in second circle; black, lung tumor only; green, lymph node only; and magenta, both) or the translocation breakpoints (no dot in the second circle; dark blue, lung tumor only; orange, lymph node only; and magenta, both) are marked with their respective chromosomal positions. Copy-number variations (CNVs) are shown as histogram plots from analysis of WGS data (two *inner* circles of lung tumor in the *outer* circle, and lymph node in the *inner* circle; green shows copy-number gains and red shows copy-number losses). Identified interchromosomal translocation events are depicted as links inside the plots (dark blue, lung tumor; orange, lymph node; and magenta, both). WES, whole-exome sequencing; TILs, tumor-infiltrating lymphocytes.

93.2% of the bases (≥Q30) aligned to the non-N human reference human genome hg19 (National Center for Biotechnology Information [NCBI] build 37), respectively. The aligned bases generated non-N reference mean coverage of 35.9×, 40.3×, and 41.4× for the normal, lung tumor, and metastatic lymph node samples, respectively. The quality of the sequencing data was further confirmed by the high concordance of single-nucleotide polymorphisms (SNPs) identified by genotyping SNP arrays (HumanOmni2.5-8v1), at 99.23%, 98.43%, and 99.24% for normal, lung tumor, and metastatic lymph node, respectively (Supplemental Table S1). We identified distinct genomic alterations between the lung tumor and metastatic lymph node genomes as SNVs, short insertions and deletions (indels), CNVs, and structural variations (SVs) (Fig. 1B–D). Only 117 of all somatic SNVs and seven of all somatic indels (<1%) were common between the lung and lymph node metastatic sites. We detected significantly more regions of CNVs in the lung tumor (L) when compared with the metastatic lymph node (LN1) using the matched normal genome as a reference (Supplemental Table S2). A majority of the CNV regions detected by WGS showed concordance with array-comparative genomic hybridization (aCGH) data (Supplemental Fig. S1).

The WES of DNA extracted from diagnostic lung biopsy FFPE tumor tissue generated a coverage of ~50× for the tumor biopsy sample, yet only ~60% of the targets reached a coverage of 30× because of the higher duplication rate (~56%) in the library constructed from the FFPE sample. WES revealed 323 highly confident somatic SNVs and 83 indels by Strelka (Saunders et al. 2012). One hundred and twenty-nine SNVs and 40 indels were presumed functionally important (missense, nonsense, and splicing sites). Thirty-five SNVs and two indels were common between the WES data on the original LBX and the WGS data on the lung tumor (L) (Supplemental Table S3). Furthermore, the total mutation burden was approximately similar among all three samples (LBX, L, and LN1) as assessed by the total number of somatic coding region nonsynonymous SNVs and indels (LBX = 169; L = 151; LN1 = 144).

Identification and Validation of Somatic Chromosomal Rearrangements

We used the WGS data to identify somatic chromosomal rearrangements or structural variants (SV). To reduce false positives, we used both BreakDancer (Chen et al. 2009) and CREST (Clipping Reveals Structure; Wang et al. 2011) to identify putative large SVs (see Supplemental Methods). We identified eight interchromosomal translocations in the lung tumor and 10 in the metastatic lymph node and one intrachromosomal translocation in the lung tumor. A majority of these translocations have one or both breakpoints within a gene region (Fig. 1D). We used polymerase chain reaction (PCR) and Sanger sequencing to validate and further characterize the rearrangements. Using genomic DNA, PCR products of predicted size were obtained for all 18 interchromosomal rearrangements, and the breakpoint sequence was confirmed by Sanger sequencing (Supplemental Figs. S2, S3; Supplemental Table S4). The mutual exclusivity for all but one of the rearrangements was verified. The somatic rearrangement common to both sites was the translocation, Chr8:57120269–Chr10:90709142 (*PLAG1-ACTA2*). We also detected this translocation in the diagnostic LBX and additional metastatic lymph nodes (LN2–8) (Figs. 1A and 3). Thus, we identified and validated 18 translocations: one common to both the lung tumor and the metastatic lymph node, seven only in the lung tumor, and 10 unique to the lymph node.

We next examined the potential of these translocations to synthesize fusion transcripts. In most cases, the rearrangements had junctions within intergenic regions or had only one side in a gene region, and in a few cases where fusions were within introns, transcription was either head to head or tail to tail and not expected to produce a fusion transcript. However, three translocations had junctions within introns, and the RNAs for both genes were transcribed in tandem and capable of producing a fusion transcript and protein. These were

PLAG1-ACTA2, which was present in both the lung (L) tumor and lymph node (LN1), and two translocations exclusive to the lung tumor (L), *PPM1L-SASH1* (Chr3:160497007–Chr6:148755717), and *ARFGEF-B4GALNT3* (Chr6:138500192–Chr12:583642).

Structure, Expression, and Functional Analyses of the *PLAG1-ACTA2* Translocation

We used fluorescence in situ hybridization (FISH) together with cDNA PCR and cellular assays to better characterize *PLAG1-ACTA2*. This translocation juxtaposes 5′*PLAG1* within intron one to *ACTA2* intron two and downstream (Fig. 2A,B). We performed two-color FISH on tumor tissue sections from the resected left lower lung mass (L), two biopsied lymph nodes (LN1 and LN2), and two lymph nodes obtained at the autopsy (LN6 and LN7) using a spectrum gold-labeled 5′ *PLAG1* combined with a spectrum red-labeled 3′ *ACTA2* BAC probe. Colocalized *PLAG1*-5′ (gold)-*ACTA2*-3′ (red) signals were present in all samples (Fig. 2C). Moreover, *PLAG1*-5′, *ACTA2*-3′, and the *PLAG1-ACTA2* fusion were amplified in the lung and metastatic lymph nodes (three to six copies of the fusion per cell). *ACTA2* and *PLAG1* break-apart FISH further confirmed these results (Supplemental Fig. S4A).

   PLAG1-ACTA2 is expected to produce a transcript with *PLAG1* exon 1 joined to *ACTA2* exon 3 (Fig. 2A). We used cDNA PCR and quantitative real-time PCR (qPCR) to compare *PLAG1-ACTA2* expression to that of the wild-type alleles (Fig. 2D–F). Both assays showed that *PLAG1-ACTA2* levels in both the lung tumor and lymph node metastases were much higher than either *PLAG1* or *ACTA2* alone, suggesting expression of *PLAG1-ACTA2* was deregulated. Results of Sanger sequencing across the exon junction were consistent with the predicted structure of this transcript (Fig. 2E). We next confirmed expression of the entire *PLAG1-ACTA2* ORF (Supplemental Fig. S4B). Because translation is normally initiated within *ACTA2* exon 3 and within *PLAG1* exon 4, this transcript may not generate a PLAG1-ACTA2 fusion protein. However, within the fusion transcript, there is an ATG in *PLAG1* exon 1 that is in-frame with the canonical ATG within *ACTA2* exon 3, which has the potential to encode an additional 13 missense amino acids (MLKPRESCEAAPA) amino terminal to ACTA2. Both ATGs occur in an adequate translational context (RNNatgY), and because translation normally initiates at the first adequate ATG, it is possible that the upstream ATG is used (Kozak 1996). To gain insight into the functional consequences of the *PLAG1-ACTA2* translocation, we expressed ACTA2 and PLAG1-ACTA2 (encompassing the upstream in-frame PLAG1 ATG) in the immortalized lung epithelial cell line, HPL1D. Western immunoblots show that expression of both ACTA2 and PLAG1-ACTA2 were similar but considerably higher than endogenous ACTA2 (Fig. 2G), in agreement with qPCR results (Supplemental Fig. S4C). ACTA2 has been implicated in promoting migration and metastasis (Lee et al. 2013). HPL1D cells overexpressing either the *ACTA2* or *PLAG1-ACTA2* transgene displayed increased chemotaxis and invasion through Matrigel, and PLAG1-ACTA2 appeared to be functionally indistinguishable from ACTA2 (Fig. 2H,I; Supplemental Fig. S4D). Taken together, these results suggest that the impact of this fusion event may indeed be the transcriptional deregulation of *ACTA2* resulting in the overexpression of ACTA2 and increased metastatic potential.

Somatic Variant Analysis

We identified a total of 14,592 nonsynonymous SNVs and 695 indels in the three genomes. 94.6% of the SNVs and 85.3% of the indels are in the Database for Short Genetic Variations (dbSNP) release 137; whereas 12.2% of SNVs and 33.4% of indels are in the Catalogue of Somatic Mutations in Cancer (COSMIC) database (version 67). Using Strelka for somatic variant calling, we identified 17,504 somatic SNVs and 685 indels in the lung tumor, and 14,013 somatic SNVs and 411 indels in the metastatic lymph node. Surprisingly, there was little overlap in the somatic variants identified in the two genomes; only 117 SNVs and seven indels
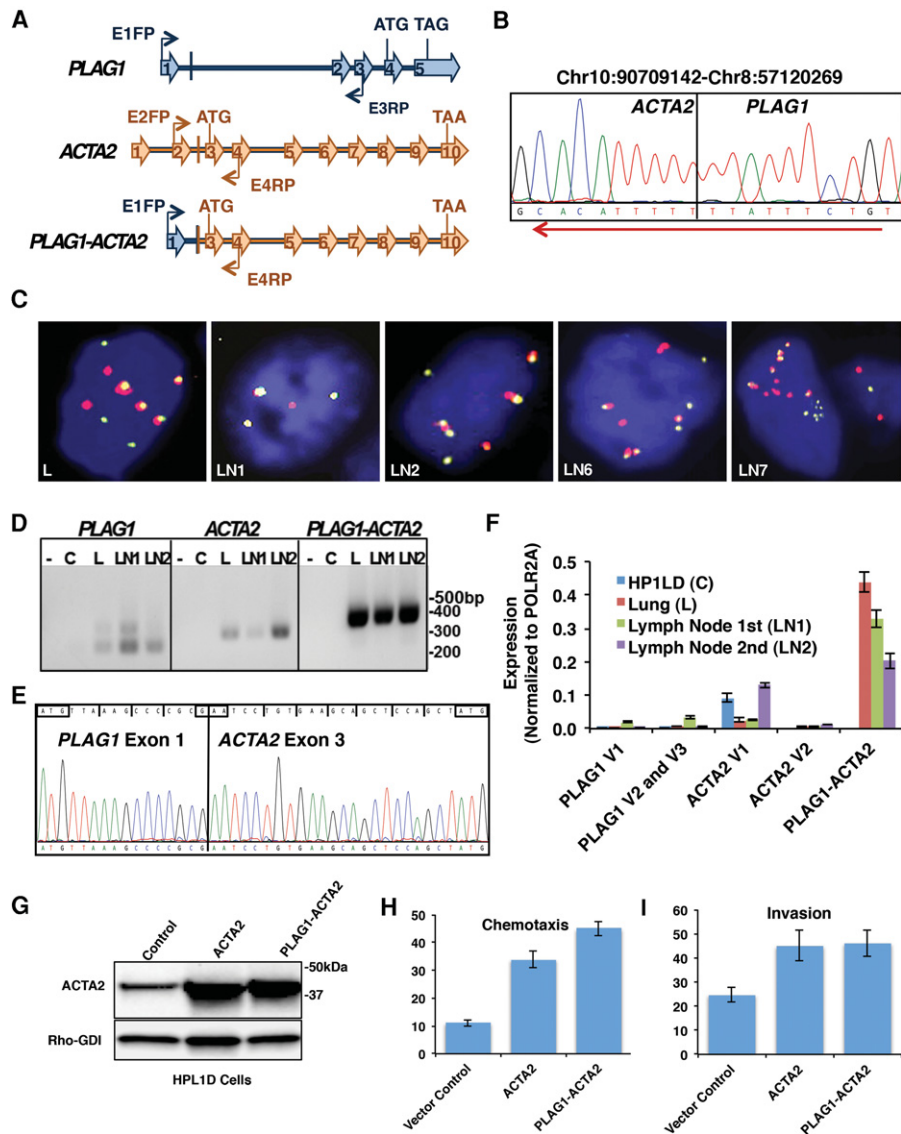
Figure 2. *PLAG1-ACTA2* translocation deregulates *ACTA2* expression in both the lung tumor and lymph node metastases and increases lung epithelial cellular migration and invasion. (*A*) Schematic representation of *PLAG1*, *ACTA2*, and *PLAG1-ACTA2* genomic structure with polymerase chain reaction (PCR) primer positions indicated. An exon 1 forward primer (E1FP) and an exon 3 reverse primer (E3RP) were used for *PLAG1*, an exon 2 forward primer (E2FP) and an exon 4 reverse primer (E4RP) were used for *ACTA2*, and a *PLAG1* E1FP and an *ACTA2* E3RP were used to PCR *PLAG1-ACTA2* from cDNA. (*B*) Chromatogram obtained from Sanger sequencing of genomic DNA (Supplementary Fig. S) showing breakpoint region and direction of transcription. (*C*) Dual-color fluorescence in situ hybridization (FISH) showing *PLAG1-ACTA2* gene fusion (5′*PLAG1*, gold; 3′*ACTA2*, red fluorescence signals) present in L, NSCLC (non-small-cell lung cancer), lung wedge, 2011; LN1, right cervical lymph node biopsy, 2011; LN2, right cervical lymph node biopsy, 2013; LN6 and LN7, autopsy material, 2015. (*D*) Gel image shows high level expression of *PLAG1-ACTA2* in both lung and lymph node (*right* panel) when compared with expression of untranslocated *PLAG2* (*left* panel) and *ACTA2 V2* (*middle* panel). *ACTA2 V2*, *PLAG1*, and *PLAG1-ACTA2* are not expressed in the HPL1D lung cell line control. (*E*) Chromatogram obtained from Sanger sequencing shows the fusion of *PLAG1* exon 1 to *ACTA2* exon 3 in the transcript and two potential translation initiation codons: canonical *ACTA2* ATG and upstream, in-frame *PLAG1* ATG. (*F*) Expression of *PLAG1*, *ACTA2*, and *PLAG1-ACTA2* in the lung tumor, metastatic lymph nodes, and HPL1D lung cell line control were determined by quantitative reverse transcription (RT)-PCR with *POLR2A* mRNA as an endogenous control. Samples were analyzed in quadruplicate, and values were expressed as the mean ± SD. (*G*) Western blot analysis of lysates of HPL1D cells with stable expression of ACTA2 or PLAG1-ACTA2. (*H*) Increased chemotaxis of HPL1D cells with stable expression of ACTA2 or PLAG1-ACTA2. (*I*) Increased Matrigel invasion of HPL1D cells with stable expression of ACTA2 or PLAG1-ACTA2.

(Fig. 1B,C; Supplemental Table S5A,B). Among the highly confident coding region somatic variants, 187 SNVs and seven indels were identified in the lung tumor, and 183 somatic SNVs and five indels were in the metastatic lymph node (Supplemental Table S5C,D). These numbers and the mutational spectrum in the lung tumor coding regions were similar to those found in other lifelong nonsmoker lung adenocarcinoma (LUAD) patients (Kandoth et al. 2013). However, the metastatic lymph node had a much different mutational spectrum, with higher A>G transition and A>T transversion frequencies but lower C>T transition frequency compared with the lung tumor (Supplemental Fig. S5). Among the functionally important somatic coding region SNVs (missense, nonsense, and splice sites) and indels identified by Strelka, 142 in lung tumor and 137 in metastatic lymph node, only three of the point mutations, *PHKG1* (NM_006213:c.1120C>T; p.A309T, COSMIC ID: COSM1090959), *FAM20C* (NM_020223:c.1181T>G; p.L317R), and *CTSG* (NM_001911:c.263C>T; p.A76T), and no indels were common to both genomes (Table 1).

Ion Torrent Targeted Resequencing of Identified Somatic Variants

To confirm the somatic SNVs and indels identified by WGS, we used the Ion Torrent Personal Genome Machine (PGM) to resequence, at higher depth, DNA from blood, the original tumor samples (L, LN1), and a lymph node metastasis collected in May 2011 (LN2), all at 2900×–4400× depth, and another series of lymph node excisional biopsies from later time points (LN3-LN10) at 1500×–2100×. Of the 277 sites validated on the Ion AmpliSeq Custom Panel, 142 were unique to the lung tumor, 137 were found only in the lymph node metastasis, and three were common to both. Of these, five were missed by the panel design (98% are on target). Of all the on-chip variant sites, using the criteria of ≥5% variant allele frequency and ≥100× coverage, 133 and 128 variant sites (~93%) were confirmed in the lung tumor (L) and lymph node metastasis, respectively (Fig. 3; Supplemental Table S6A,B). One from each of L and LN1 was confirmed to be germline mutations. Interestingly, of all the additional lymph node metastases we resequenced, the concordance among them ranged from 94% (LN10) to 43% (LN8). *PHKG1* and *CTSG*, two of the three common variants identified by WGS, were confirmed in all samples.

Expression of Gene Variants in the Lung Tumor and/or Metastatic Lymph Node

We used cDNA PCR and Sanger sequencing to further verify and characterize a subset of nonsynonymous coding region SNVs and indels originally identified by WGS. We focused on 20 mutations (three germline and 17 somatic). All 19 genes (one gene has two mutations) were expressed, although at different levels, in both the lung tumor and lymph node (Supplemental Fig. S6). Of the two somatic SNVs common to both tissues, *PHKG1*-A309T was the only gene expressing the mutated form (Supplemental Fig. S7A). Of the 10 somatic mutations found only in the lung tumor (L), we saw expression of the mutated allele for five genes: *MET*-Del S203, *CDK12*-G879V, *TP53*-Del E339-F341, *ROS1*-S1581F, and *HIP1*-S 866P (Figs. 4D; 6E; Supplemental Fig. S7B). Of the four somatically mutated genes found only in the lymph node (LN1), we saw expression of the mutated allele for three genes: *ERBB2*-L869R, *MLL3*-S2663*, and *NF2* -G43V (Fig. 4C; Supplemental Fig. S7C).

CDK12 and *ERBB2* Mutation, Amplification, and Overexpression

CDK12 and *ERBB2* were amplified in the lung tumor (L), and to a lesser extent in the metastatic lymph node (LN1) (Figs. 1D, 4A). *ERBB2* with an L869R mutation in the tyrosine kinase domain was exclusive to and expressed in the lymph node (LN1), whereas *CDK12* with a G879V somatic mutation within the conserved DFG motif in the protein kinase domain was exclusive to and expressed in the lung tumor (L). We used Sanger sequencing to verify expression of the mutated genes and qPCR to more accurately assess their overall

Table 1. Summary of significant nonsynonymous mutations in the lung tumor and metastatic lymph node

Chr	Location	Gene	Allele change	RefSeq ID	cDNA pos	AA change	Mutation	SIFT score	PolyPhen2 pred[a]	Lung	Lymph node	Mutation type	Cancer gene[b]	#reads ref/alt[c]	Exp. by Sanger[d]	Ion Torrent[e]
Chr9	133760372	ABL1	C>T	NM_007313	3191	P918S	MISSENSE	0.62	D	Y	Y	Germline	Y	38/22, 14/37	Y, Y	NA
Chr17	37665984	CDK12	G>T	NM_016507	3222	G879V	MISSENSE	0	D	Y	N	Somatic	Y	3/941	Y	Y
Chr16	11000572	CIITA	G>A	NM_000246	1356	R408Q	MISSENSE	0.62	B	Y	N	Somatic	Y	102/18	N	Y
Chr14	25043994	CTSG	C>T	NM_001911	263	A76T	MISSENSE			Y	Y	Somatic	N	67/46, 36/15	N, N	Y, Y
Chr2	25523009	DNMT3A	G>A	NM_022552	443	P59L	MISSENSE	1	B	Y	N	Somatic	Y	79/9	N	Y
Chr7	73466269	ELN	C>T	NM_000501	996	A302V	MISSENSE	0.04	P	N	Y	Somatic	Y	40/10	N	Y
Chr17	37881414	ERBB2	T>G	NM_004448	2844	L869R	MISSENSE	0	D	N	Y	Somatic	Y	61/89	Y	Y
Chr7	286467	FAM20C	T>G	NM_020223	1181	L317R	MISSENSE	0	D	Y	Y	Somatic	N	76/6, 60/6	N, N	N, N
Chr7	75174450	HIP1	A>G	NM_005338	2641	S866P	MISSENSE	0.23	B	Y	N	Somatic	Y	28/6	Y	Y
Chr4	123377504	IL2	T>G	NM_000586	147	Q31P	MISSENSE	0.06	B	Y	N	Somatic	Y	46/9	N	Y
Chr7	116339743	MET	ATTC>A-	NM_000245	792–795	NS202N	CODON DELETION			Y	N	Somatic	Y	63/14	Y	Y
Chr7	151874550	MLL3	G>T	NM_170606	8207	S2663*	NONSENSE	1		N	Y	Somatic	Y	41/9	Y	Y
Chr22	30032753	NF2	G>T	NM_000268	571	G43V	MISSENSE		D	N	Y	Somatic	Y	27/5	Y	Y
Chr1	144854197	PDE4DIP	C>T	NM_014644	7258	D2323N	MISSENSE	0.07	B|D	Y	N	Somatic	Y	274/15	N	N
Chr7	56148986	PHKG1	C>T	NM_006213	1120	A309T	MISSENSE	0.06	B	Y	Y	Somatic	N	101/32, 60/35	Y, Y	Y, Y
Chr13	48878084	RB1	CGCCGCCGCT>C-	NM_000321	202–209	AAA13-	CODON DELETION	0.03	D	Y	Y	Germline	Y	25/6, 40/16	Y, Y	NA
Chr6	117662723	ROS1	G>A	NM_002944	4941	S1581F	MISSENSE	0	P	Y	N	Somatic	Y	29/24	Y	Y
Chr6	117704607	ROS1	T>C	NM_002944	2568	N790S	MISSENSE	0.05	P	Y	Y	Germline	Y	21/24, 11/29	Y, Y	NA
Chr7	66459318	SBDS	G>T	NM_016038	323	L47I	MISSENSE	0.86	B	Y	N	Somatic	Y	27/14	N	Y
Chr17	7574003	TP53	GGAACATCTC>G-	NM_000546	1225–1234	EMF339-	CODON DELETION			Y	N	Somatic	Y	8/27	Y	Y

[a] The PolyPhen2 pred column contains PolyPhen2 predictions: D, probably damaging; P, possibly damaging; and B, benign.
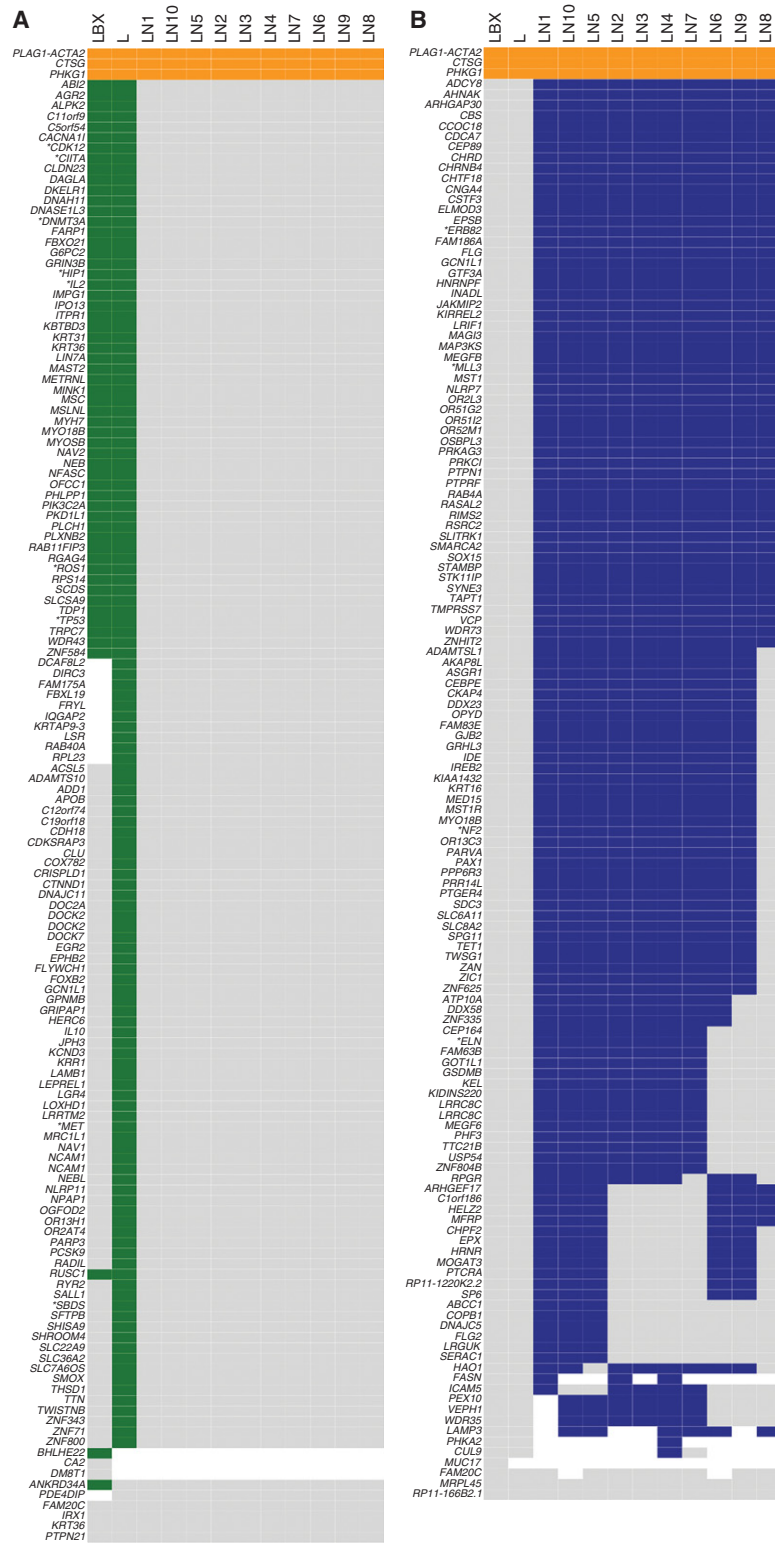[b] The Cancer gene column indicates whether the gene is in the cancer gene census list from a December 18, 2013 COSMIC download.
[c] For #reads ref/alt column, if the mutant occurred in both tumors, then the first set of numbers is from the lung tumor and the second is from metastatic lymph node.
[d] The Exp. by Sanger column contains data from Sanger sequencing of cDNA.
[e] The Ion Torrent column contains validation data from Ion Torrent deep sequencing.
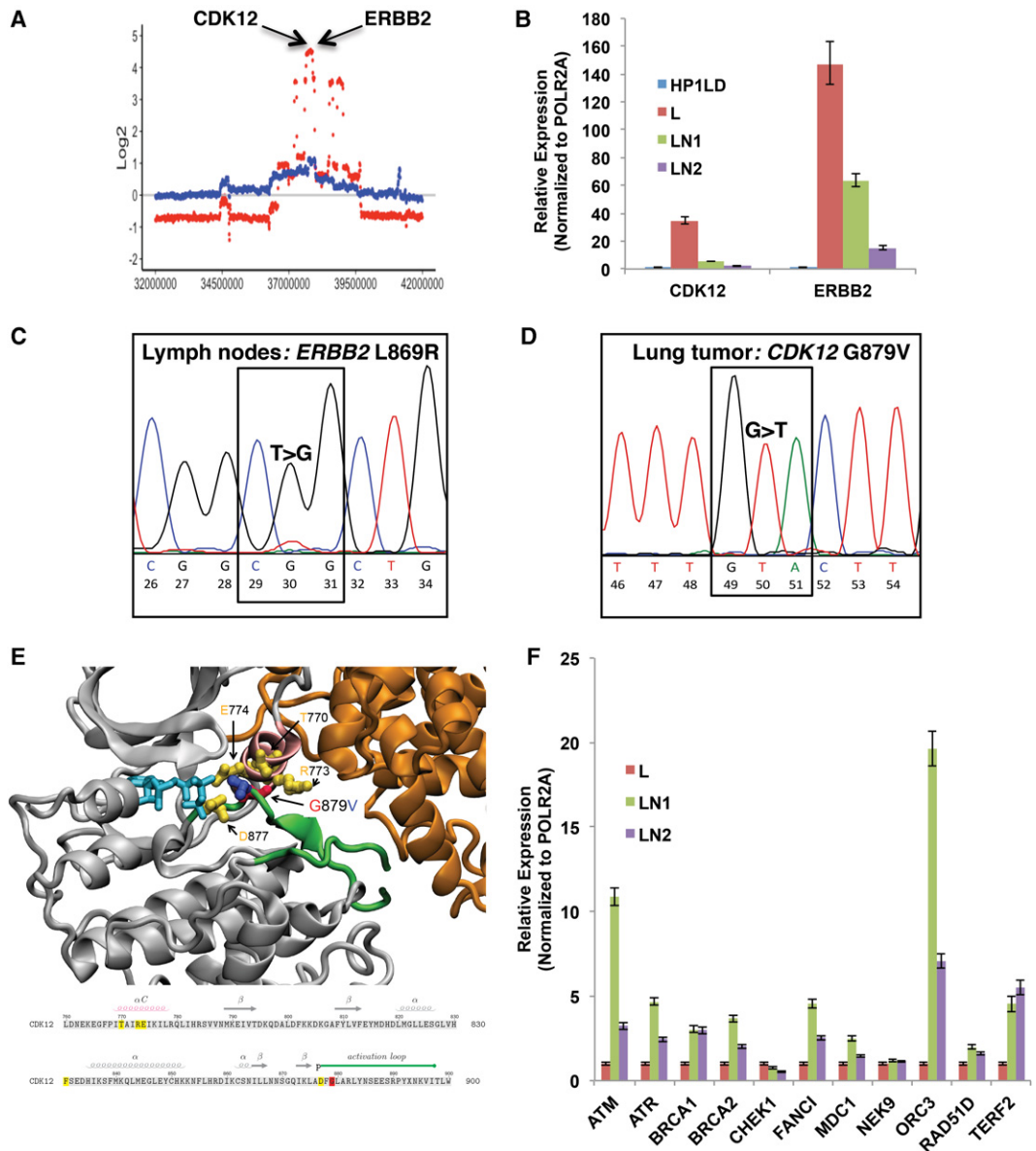Y, yes; N, no; NA, not available.

Figure 3. Validation of all nonsilent somatic variants identified by whole-genome sequencing (WGS) with exome and targeted sequencing and by polymerase chain reaction (PCR)/Sanger sequencing for *PLAG1-ACTA2* translocation. Heatmaps show the presence of variants (green, lung tumor only; blue, lymph node only; orange, both), the absence of variants (light gray), and the presence/absence cannot be determined (white) because of low or lack of coverage at each somatic variant site in lung tumor (*A*) and lymph node metastases (*B*) as assessed by targeted Ion Torrent deep resequencing or by exome sequencing (lung tumor biopsy [LBX] only), and by PCR/Sanger sequencing (for *PLAG1-ACTA2* translocation in orange). Genes carrying mutations are labeled on the *left* edges, and the cancer genes are marked with an asterisk (*) on the side.

Figure 4. *CDK12* and *ERBB2* are coamplified and overexpressed but display a different mutation pattern in both metastatic sites. (*A*) Region with high-copy-number amplification in the lung tumor (L; red) and metastatic lymph node (LN1; blue) as assessed by copy-number variation sequencing (CNV-seq) analysis of WGS data. Arrows indicate the approximate locations for *ERBB2* and *CDK12*. (*B*) *CDK12* and *ERBB2* expression levels in the lung tumor, lymph nodes, and HPL1D lung cell line control were determined by quantitative reverse transcription–polymerase chain reaction (RT-PCR) with *POLR2A* mRNA as an endogenous control. Samples were analyzed in quadruplicate, and values expressed as the mean ± SD. (*C*) Chromatogram obtained from Sanger sequencing of cDNA shows that *ERBB2* with a somatic T>G (CtG to CgG: leucine to arginine) mutation is expressed in the metastatic lymph node. (*D*) Chromatogram obtained from Sanger sequencing of cDNA shows that *CDK12* with a somatic G>T (GgA to GtA: glycine to valine) mutation is expressed in the lung tumor. (*E*) Effects of G879V mutation on the active conformation of CDK12. (*Top*) Active conformation of CDK12 with cyclin K (PDB ID 4CXA chains *A*, *B*) and mutated structure are superimposed in the same orientation and are colored as follows: αC-helix, activation loop, and ATP-like compound are highlighted in pink, green, and cyan, respectively; the residue G879 and its mutated version V879 are shown in red and blue, respectively; the side chains of amino acids that interact with the G879V site are shown in yellow. Sequence representation is shown *below*. Region of CDK12 sequence (760–900): Residues interacting with the point mutation site are highlighted by yellow; the mutation site is highlighted in red. (*F*) Expression of CDK12-regulated DNA repair genes in the lung tumor, metastatic lymph nodes, and HPL1D lung cell line control were determined by quantitative RT-PCR with *POLR2A* mRNA as an endogenous control. Samples were analyzed in quadruplicate, and values expressed as the mean ± SD.

expression levels in both tumors. Consistent with their amplification profiles, *CDK12* and *ERBB2* were more highly overexpressed in the lung tumor (L) when compared with the metastatic lymph nodes (LN1 and LN2). Moreover, expression of these genes was higher in LN1, which was resected 2.5 yr before LN2 and 6 mo before the lung tumor (Fig. 4B–D). Thus, the lung tumor overexpressed a potentially nonfunctional *CDK12* variant, whereas the metastatic lymph nodes expressed lower levels of wild-type *CDK12*. On the other hand, the lung tumor expressed higher levels of wild-type *ERBB2*, whereas the metastatic lymph nodes expressed lower levels of an *ERBB2* variant harboring a kinase domain mutation, predicted to be constitutively active.

Molecular modeling shows that the G879V mutation in the active conformation of CDK12 perturbs interactions with residues D877, E774, R773, and T770. Because this mutation introduces a bulky amino acid side chain into the structure, it is likely that the G879V mutation will cause local rearrangement of the structure (Fig. 4E). The estimations of changes in free energy of unfolding upon G879V mutation (1.6 kcal/mol) suggest that the mutation may destabilize the known active structure of CDK12.

CDK12 is a transcriptional cyclin-dependent kinase that controls elongation of long complex genes, many of which are involved in DNA damage repair (DDR) (Blazek et al. 2011; Joshi et al. 2014). Analysis of The Cancer Genome Atlas (TCGA) high-grade serous ovarian carcinoma revealed a coordinated down-regulation of DDR genes in patients with *CDK12* mutations (Ekumi et al. 2015). Because our patient's lung tumor expressed high levels of a potentially nonfunctional *CDK12*, we wanted to determine if expression of these DDR genes whose transcription is regulated by CDK12 differed between these two tissues. Results of qPCR analyses (Fig. 4F) showed that indeed, in comparison to the metastatic lymph node, the lung tumor expressed significantly lower levels of most of these DDR genes (*BRCA1*, *BRCA2*, *ATM*, *ATR*, *FANCI*, *TERF2*, *RAD51D*, *MDC1*, *ORC3*).

Functional Characterization of ERBB2-L869R and Its Association with Metastasis and Lapatinib Resistance

We first analyzed the published crystal structures of the active and inactive conformations of the ERBB2 and EGFR complex with or without lapatinib (Supplemental Fig. S8A,B). The L869R mutation in the inactive conformation of ERBB2 would perturb interactions with E766, I767, and E770. The potential for R869 to form salt bridges with E766 and E770 would likely further stabilize this inactive conformation. The estimations of changes in free energy upon mutation suggests that inactive conformations of EGFR1 and ERBB2 with lapatinib are destabilized, which may correlate with increased lapatinib resistance observed in this study (Supplemental Fig. S8C).

We sought to determine how this mutation affects ERBB2 function, particularly with respect to growth, survival, migration, and sensitivity to the ERBB2 inhibitor, lapatinib. We used the immortalized human mammary epithelial cell line, MCF10A, to overexpress either ERBB2-WT or ERBB2-L869R. Both were overexpressed in MCF10A cells; however, the level of mutant ERBB2 was consistently lower than ERBB2-WT (Fig. 5A). Interestingly, the ratio of phospho-ERBB2 (pERBB2) to total ERBB2 was higher in cells expressing mutant ERBB2 (Fig. 5B). Cells expressing either ERBB2-WT or ERBB2-L869R had a significantly accelerated growth rate in complete medium (Fig. 5C). However, in serum-free medium, ERBB2-WT-expressing cells were growth-arrested, whereas the ERBB2-L869R-expressing cells continued to grow (Fig. 5D). ERBB2-L869R-expressing cells also formed significantly more colonies on tissue culture plates (Fig. 5E,F). When grown in dense cultures, cells expressing ERBB2 L869R formed foci of proliferating cells. When grown as colonies, a significant number of colonies expressing EGFR L869R showed elongated migratory cells in the periphery of the colonies (Supplemental Fig. S9A). We next studied the effects of this mutated protein on
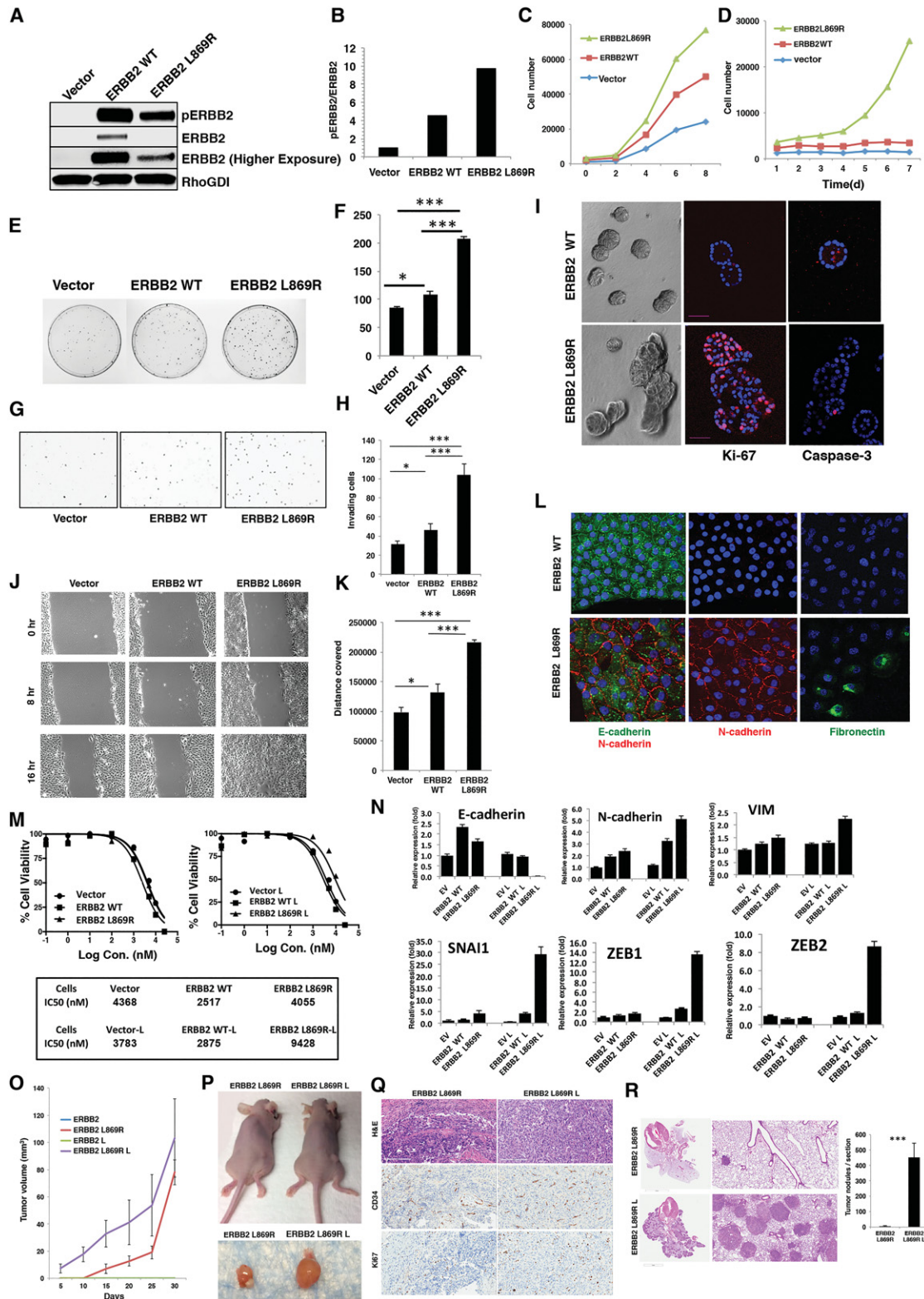
Figure 5. Functional consequences of ERBB2 L869R mutation. (*Legend continues on following page.*)

chemotaxis and migration. Using a 3D transwell migration assay, we showed greater invasion of ERBB2-L869R-expressing cells in comparison to ERBB2-WT (Fig. 5G,H). Similarly, 3D culture in Matrigel-containing basement membrane extract showed greater invasion and formation of complex acini in mutant ERBB2-expressing cells. These acinar cells in 3D culture expressed more Ki67, suggesting increased proliferation and showed no significant change in the apoptosis marker, caspase-3 (Fig. 5I). The mutant ERBB2-expressing cells were also more migratory on plastic than cells expressing ERBB2-WT as evidenced by a wound healing scratch assay (Fig. 5J,K). The migratory potential of ERBB2 L869R-expressing elongated cells was also evident in the periphery of a subset of colonies formed in tissue culture plates (Supplemental Fig. S9A). Because these migratory cells had characteristics of mesenchymal cells, we examined epithelial and mesenchymal marker expression. Compared with ERBB2-WT, ERBB2-L869R cells expressed decreased levels of E-cadherin and increased levels of N-cadherin and fibronectin (Fig. 5L).

Our patient's lymph nodes repeatedly progressed on ERBB2-directed therapies. Hence, we interrogated the effects of overexpressed wild-type or mutant ERBB2 on erlotinib and lapatinib sensitivity in 3D cultures. MCF10A vector control, ERBB2-WT, and mutant ERBB2-overexpressing MCF10A cells were grown on 3D Matrigel for 20 d in the presence of dimethyl sulfoxide (DMSO) (0.5%), erlotinib (3 μM), or lapatinib (5 μM), drugs which our patient received during his course of treatment. Lapatinib inhibited the growth of both ERBB2-WT- and ERBB2-L869R-expressing cells in 3D Matrigel (Supplemental Fig. S9B). However, after 20 d in 3D culture, a fraction of ERBB2-L869R cells continued to grow. These cells were maintained in lapatinib and allowed to grow for several more days to select

Figure 5. (*Continued*) (*A,B*) Expression of ERBB2 WT and ERBB2 L869R mutant in MCF10A cells results in increased relative phosphorylation of ERBB2 L869R. Cell lysates of MCF10A cells expressing vector control, ERBB2 WT, and ERBB2 L869R were resolved in sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS-PAGE). The immunoblots were probed with indicated antibodies (*A*). Quantitation of the western blots is represented by the bar graphs (*B*). (*C,D*) Increased proliferation of cells expressing ERBB2 L869R mutant. MCF10A cells expressing the vector control, ERBB2 WT, and ERBB2 L869R were grown in complete growth medium (*C*) or serum-free medium (*D*) in 96-well plates and counted every 2 d by a cell-titer glow assay. The data represents ±SD of five wells for each data point. (*E,F*) ERBB2 L869R mutant increases colony formation. MCF10A cells expressing vector control, ERBB2 WT, and ERBB2 L869R were grown in six-well plates (500 cells per well) for a colony formation assay. The bar graph (*F*) shows the number of colonies obtained. The data are representative of three independent experiments. (*G,H*) ERBB2 L869R mutant increases Matrigel invasion. MCF10A cells expressing vector control, ERBB2 WT, and ERBB2 L869R were grown on Matrigel-coated transwell chambers. In each well, 10^5 cells were seeded. The bar graph (*H*) shows the number of cells that invaded through the Matrigel. The data are an average of triplicate wells. (*I*) The ERBB2 L869R mutant increases proliferation and reduces apoptosis in 3D cultures of MCF10A cells. MCF10 cells expressing ERBB2 WT and ERBB2 L869R were grown in basement membrane extract for 12 d. 3D colonies were fixed and stained for immunofluorescence with Ki67 and caspase-3-specific antibodies as indicated. DAPI was used for nuclear staining. (*J,K*) ERBB2 L869R mutant increases migration in a wound-healing assay. MCF10A cells grown in monolayer expressing vector control, ERBB2 WT, and ERBB2 L869R were subjected to a scratch-induced wound healing assay. The bar graph (*K*) represents quantitation of migration from three independent experiments. (*L*) Expression of ERBB2 L869R mutant results in increased expression of mesenchymal markers, N-cadherin and fibronectin and reduced expression of epithelial marker, E-cadherin. Immunofluorescence stainings with E-cadherin (green)-, N-cadherin (red)-, and fibronectin (green)-specific antibodies were performed, and DAPI was used for nuclear counterstaining. (*M*) GI$_{50}$ determination of lapatinib-treated MCF10A cells expressing ERBB2 WT or ERBB2 L869R first grown in 3D culture in the presence of dimethyl sulfoxide (DMSO) (*left*) and lapatinib (*right*). (*N*) Expression of EMT markers in MCF10A cells expressing vector control, ERBB2 WT, and ERBB2 L869R treated with DMSO or rendered lapatinib resistant in 3D culture. (*O–R*) MCF10A cells expressing ERBB2 L869R that are sensitive or resistant to lapatinib form xenograft tumors in nude mice. The lapatinib-resistant ERBB2 L869R-expressing cells form larger tumors (*O,P*). There are increases in Ki67-positive proliferating cells and angiogenesis as evidenced by increased CD34 staining in lapatinib-resistant ERBB2 mutant tumors (*Q*). Lapatinib-resistant ERBB2 mutant cells form significantly increased metastases in a tail-vein injection metastases assay (*R*). (***) $P < 0.001$, (*) $P < 0.05$.

for resistant cells. ERBB2-WT-expressing cells had fewer cells growing in 3D culture, in the presence of lapatinib. The lapatinib-resistant ERBB2-WT and ERBB2-L869R cells were then isolated from the 3D culture and grown in 2D tissue culture plates in presence of lapatinib. The lapatinib-resistant cells were designated as ERBB2-WT-L and ERBB2-L869R-L. These 3D colonies were dissociated and plated for a GI_{50} analysis for lapatinib sensitivity. The GI_{50} of mutant ERBB2-expressing MCF10A cells was ~1.5× that of ERBB2-WT-expressing cells and similar to vector control cells. We next determined the GI_{50} values of lapatinib-resistant ERBB2-expressing cells. The GI_{50}, in the presence of lapatinib, for ERBB2-L869R-L cells was about three times that of ERBB2-WT-L cells, confirming a greater resistance of mutant ERBB2-expressing cells to lapatinib (Fig. 5M). This demonstrates that the ERBB2-L869R cells were more resistant to lapatinib initially and their resistance increased even further after selection in the presence of lapatinib in 3D followed by 2D culture system. Western blot analysis showed that lapatinib-resistant cells had increased expression of both ERBB2-WT and ERBB2-L869R compared with cells grown in DMSO (Supplemental Fig. S10). Interestingly, expression of the corresponding ERBB2 protein was higher in ERBB2-WT-L cells compared with ERBB2-L869R-L cells. EGFR:ERBB2 heterodimers more potently (Hirsch et al. 2009) activate EGFR tyrosine kinase than do EGFR homodimers. Consequently, there was increased EGFR phosphorylation in ERBB2-L869R-expressing cells. We found increased phospholipase-Cγ (PLCγ) phosphorylation consistent with ERBB2 overexpression increasing EGFR phosphorylation at Y1173, which forms a docking site for PLCγ (Hartman et al. 2013). In addition, there was increased ERK, AKT, and SRC phosphorylation, suggesting activation of specific growth and survival signaling downstream from ERBB2/EGFR.

To further characterize these laptinib-resistant ERBB2 mutant cells, we analyzed the expression of epithelial–mesenchymal transition (EMT) markers at the transcript level. We saw a significant reduction of E-cadherin expression and enhanced expression of mesenchymal markers such as N-cadherin, vimentin, *SNAI1*, *ZEB1*, and *ZEB2*, demonstrating a strong mesenchymal phenotype in ERBB2-L869R-L cells (Fig. 5N). These findings further substantiate that upon lapatinib treatment, the ERBB2 mutant cells acquired an enhanced EMT phenotype. The lapatinib-resistant ERBB2 mutant cells showing the mesenchymal phenotype may have increased metastatic potential and may form tumors in distant organs. MCF10A cells overexpressing ERBB2-WT and ERBB2-L869R that were either sensitive or resistant to lapatinib were injected in the flanks of nude mice to generate xenografts. Whereas the ERBB2-WT cells failed to form tumors in these mice, the ERBB2-L869R-expressing cells formed palpable tumors in 16 d. The mice were observed for a period of 30 d, and the ERBB2-WT cells failed to form tumors. Interestingly, the ERBB2-L869R-L cells formed tumors more rapidly and were larger in size than the tumors formed by ERBB2-L869R cells (Fig. 5O, P). Hematoxylin and eosin (H&E) staining showed that the ERBB2-L869R-L xenograft tumors had more dense, hypercellular masses with spindle cells. Immunohistochemistry of the ERBB2-L869R-L xenograft tumors showed an increase in CD34, a vascularization marker. The ERBB2-L869R-L tumors also showed higher Ki67 staining, demonstrating increased tumor cell proliferation (Fig. 5Q). We next interrogated the metastatic potential of the ERBB2-L869R-L cells. MCF10A cells (5×10^6) overexpressing ERBB2-L869R, either sensitive or resistant to lapatinib, were injected into the tail veins of nude mice. Six weeks later lung metastatic foci were identified. There were significantly more lung metastatic tumor foci formed by lapatinib-resistant mutant ERBB2-expressing cells compared with the sensitive cells (Fig. 5R).

Somatic Alterations in Cell Cycle Regulatory Genes

There was copy number loss of the cell cycle inhibitor *CDKN2A* in the lung tumor, whereas both the lung tumor and lymph node displayed copy-number gains of two different cyclins,
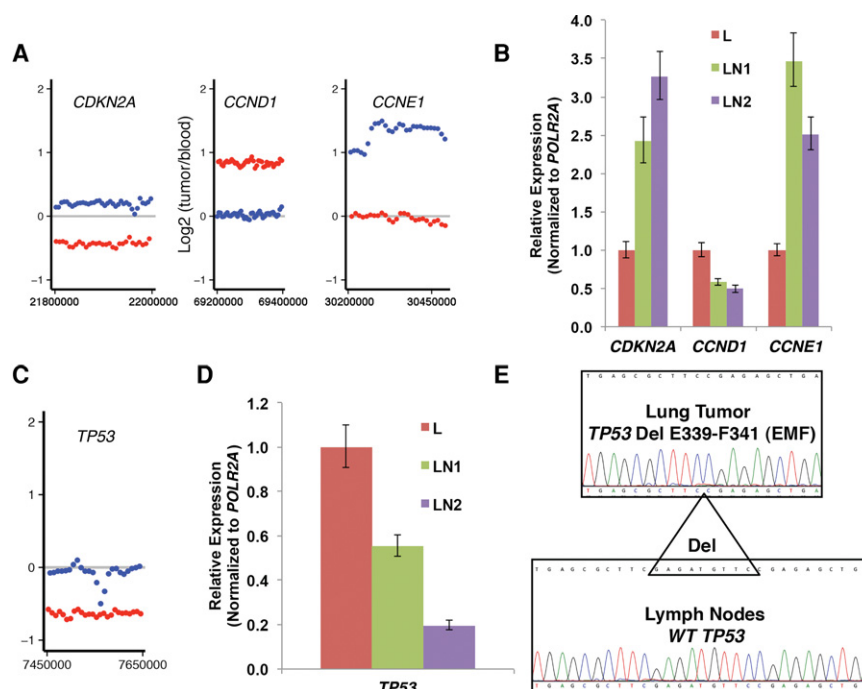
Figure 6. Cell cycle control and tumor suppressor function are disrupted by different mechanisms in both metastatic sites. (*A*) Genomic regions encompassing *CDKN2A*, *CCND1*, and *CCNE1* with copy-number gains or losses in lung tumor (red) and metastatic lymph node (blue) as assessed by CNV-seq analysis of WGS data. (*B*) Relative levels of *CDKN2A*, *CCND1*, and *CCNE1* expression in the lung tumor and lymph nodes were determined by quantitative RT-PCR using *POLR2A* mRNA as an endogenous control. Samples were analyzed in quadruplicate, and values expressed as the mean ± SD. (*C*) Region encompassing *TP53* with copy-number loss in the lung tumor (L-red) and focal loss of *TP53* in the metastatic lymph node (LN1-blue) as assessed by CNV-seq analysis of WGS data. (*D*) Expression of *TP53* in the lung tumor and metastatic lymph nodes was determined by quantitative reverse transcription–polymerase chain reaction (RT-PCR) with *POLR2A* mRNA as an endogenous control. Samples were analyzed in quadruplicate, and values expressed as the mean ± SD. (*E*) Chromatogram obtained from Sanger sequencing of cDNA shows that *TP53* with an in-frame 9-nt somatic deletion (Del EMF) is expressed in the lung tumor, whereas wild-type *TP53* is expressed in the lymph node.

both of which are involved in cell cycle progression. *CCND1* was amplified in the lung tumor and *CCNE1* was amplified in the metastatic lymph node (Fig. 6A). We used qPCR analyses to assess whether these copy number changes were reflected in the expression pattern of these cell cycle regulators. In fact, the lung tumor expressed lower levels of *CDKN2A* and higher levels of *CCND1*, whereas the metastatic lymph nodes expressed higher levels of *CCNE1* (Fig. 6B). In addition to these cell cycle regulators, the tumor suppressor *TP53* was affected in both the lung tumor and lymph node metastases. The lung tumor displayed a region of copy number loss encompassing *TP53*, whereas the lymph node showed a more focal copy number loss of *TP53* (Fig. 6C). Results of qPCR analyses show that expression of *TP53* was in fact lower in both metastatic lymph nodes in comparison to the lung tumor (Fig. 6D). Although the lung tumor expressed higher levels of *TP53*, it possessed an in-frame 9 nucleotide (nt) deletion (Del E339-F341) within the homo-oligomerization domain (Fig. 6E). Therefore, the lung tumor expressed a potentially nonfunctional TP53, whereas the lymph node expressed lower levels of a functional TP53. Taken together, results of these experiments suggest that although these tumor tissues displayed a high degree of genomic heterogeneity, deregulation of cell cycle regulatory genes and the TP53 tumor suppressor
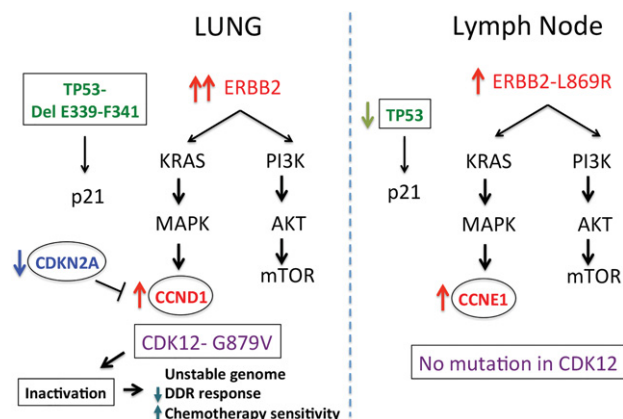
Figure 7. Schematic of key distinct somatic alterations in cell cycle–regulatory genes in the lung and lymph node metastatic sites resulting in activation of a major cancer hallmark, cell proliferation.

may have disrupted cell cycle control and tumor suppressor function thus resulting in a similar outcome (Fig. 7).

DISCUSSION

The prognosis of metastatic lung adenocarcinoma is extremely poor. During tumorigenesis, cells acquire additional genetic or epigenetic changes, which allow for invasion and dissemination of specific metastatic tumor cell clones. Thus, significant heterogeneity between a primary tumor and metastatic lesions, as well as between various metastatic sites, is quite likely and, as such, may influence treatment strategies (Campbell et al. 2010; Gerlinger et al. 2012; Vignot et al. 2012; Fisher et al. 2013; Vogelstein et al. 2013). Here, we describe a patient diagnosed with metastatic lung adenocarcinoma who survived in good performance status >7 yr while undergoing multiple rounds of therapy. We used WGS, WES, and targeted next-generation sequencing (NGS) on 10 separate metastatic lymph nodes and two separate metastatic lung tumors obtained during his course of treatment, including several metastatic lymph nodes obtained at autopsy. We discovered an unprecedented level of heterogeneity between the lung and lymph node metastatic sites with <1% similarity in SNVs and indels. It is conceivable that such extensive heterogeneity impacted our patient's treatment response.

We identified a novel translocation, *PLAG1-ACTA2*, that was common among all of the lung and lymph node metastatic sites tested, suggesting that the resulting ACTA2 overexpression may have been an early event causing metastatic spread. FISH analyses revealed copy-number gain in the lymph nodes obtained at autopsy. The copy number of the untranslocated genes also increased over time. Break-apart FISH showed that the 3′ *PLAG1* and 5′ *ACTA2* regions not involved in the translocation were retained, indicating that they were also rearranged (Supplemental Fig. S4A). However, we were unable to detect the reciprocal translocation by genomic PCR, suggesting that the rearrangement was more complex. We designed a TaqMan assay specific for the fusion transcript to use in quantitative PCR assays. Expression of this fusion transcript, in both tumors, at levels considerably higher than either untranslocated gene suggests that *PLAG1-ACTA2* expression was deregulated in both tumors. Low-level amplification, as demonstrated by FISH, may not entirely account for the high-level expression of the fusion transcript, especially in

the lung tumor and lymph node biopsies from 2011 where expression of the fusion transcript was at least 10-fold higher than either *PLAG1* or *ACTA2* alone. However, in the lymph node biopsy from 2013, there was up-regulation of *ACTA2* and the difference in expression between *ACTA2* and the fusion transcript was less than twofold. The elevated expression of *ACTA2* has been implicated in increased metastatic potential and poor outcome in human lung adenocarcinoma patients and the role of ACTA2 in these processes has been validated in vitro (Lee et al. 2012, 2013). We show that overexpression of PLAG1-ACTA2 in immortalized lung epithelial cells resulted in increased chemotaxis and migration (Fig. 2H,I; Supplemental Fig. S4C,D).

We also found a second potentially interesting structural variant, *PPM1L-SASH1*. In contrast to *PLAG1-ACTA2* which was common to both tumors, *PPM1L-SASH1* was exclusive to the lung tumor. *PPM1L*, *SASH1*, and the *PPM1L-SASH1* fusion transcripts were expressed at fairly high yet comparable levels in the lung tumor (Supplemental Fig. S11A–D). *PPM1L* and *SASH1* are both putative tumor suppressor genes: *PPM1L* is located on Chromosome 3 in a copy-number-variable region (3q26) implicated in colorectal cancer and *SASH1* is located on Chromosome 6 in a region (6q23) whose allelic loss is associated with many solid tumors including lung cancer (Zeller et al. 2003; Thean et al. 2010). PPM1L is a single-pass type I membrane protein, whereas SASH1 is normally located within the nucleus or cytoplasm (Saito et al. 2008; Martini et al. 2011). The PPM1L-SASH1 fusion juxtaposes the 133 amino-terminal hydrophobic transmembrane domain of PPM1L onto amino acids 97–1247 of SASH1 (Supplemental Fig. S11E). This fusion protein is expected to retain all the characterized functional domains of SASH1, including two nuclear localization signals, but also has the potential of homing to the estrogen receptor (ER). Determining the cellular location of this fusion protein will provide us with insight into the consequences of this translocation. If this fusion protein is a membrane protein, this may be yet another mechanism by which a tumor suppressor may become inactivated, by sequestration to an inappropriate cellular compartment. Determining whether these novel translocations play a role in tumor development and metastasis warrants further study.

We identified three cancer genes with expressed germline SNVs, *ABL1*, *ROS1*, and *RB1*. *ROS1*-N790S is a previously characterized SNP (rs34582164), and *ABL1*-P118S and *RB1*-Del A13-15 are both novel mutations (Supplemental Fig. S12). Surprisingly, we found only three nonsynonymous somatic SNVs that were common to the lung and the lymph node metastatic sites; two were later confirmed by Ion Torrent deep sequencing and neither are in the COSMIC cancer gene census list. Although both of these genes were expressed in these tissues, the mutated allele was expressed for only one of these genes. This gene encodes phosphorylase kinase gamma 1 subunit (PHKG1), a protein involved in glycogen metabolism. Although *PHKG1* is not considered a cancer gene, this *PHKG1*-A309T point mutation has been seen in endometrial cancer (COSM1090959). In addition, a SNP has been identified in *PHKG1* (rs148231327) that results in an A309S mutation. The significance of this mutation or the SNP described before at the same site remains to be determined. We found novel somatic nonsynonymous mutations in several key genes implicated in lung adenocarcinoma (Govindan et al. 2012; Imielinski et al. 2012), which were not shared between the lung and lymph node metastatic sites. We confirmed an *ERBB2* amplification previously identified by FISH in this patient (Kelly et al. 2012). The region encompassing *ERBB2* was more highly amplified in the lung tumor than in the lymph node (reads coverage ratio of $\log_2$ values of 4.7 vs. 1.0). However, the metastatic lymph node carried an L869R missense mutation in *ERBB2*, reported to be homologous to the *EGFR* L861R sensitizing mutation (Lee et al. 2006; Wu et al. 2011). *ERBB2* may be the predominant driver oncogene in this patient because of high-level amplification in the lung and amplification together with an activating mutation in the lymph node. Functional studies also showed that ERBB2 L869R mutant can drive EMT, more so upon developing lapatinib resistance. In agreement with these studies,

we also show increased expression of several mesenchymal markers in the metastatic lymph node when compared with the lung tumor (Supplemental Fig. S13).

Another gene within this shared amplified region encodes CDK12, a transcriptional cyclin-dependent kinase that in association with cyclin K regulates transcriptional elongation of long complex genes, many of which are involved in the DDR (Blazek et al. 2011; Joshi et al. 2014). In contrast to *ERBB2*, *CDK12* was mutated (G879V) and highly amplified in the lung tumor and not mutated, but moderately amplified in the lymph node. The G879V point mutation falls within the conserved DFG motif in the CDK12 protein kinase domain (Malumbres 2014). *CDK12* harboring a point mutation within its DFG motif (D877N) has been identified in a cholangiocarcinoma patient sample from TCGA, and a *CDK12* D877N expression construct has been used as a kinase dead control in various studies (Bosken et al. 2014). In one case, the D877N mutation reduced CDK12-mediated transcriptional elongation and concurrently inhibited the expression of a number of DDR genes (Joshi et al. 2014). Two independent analyses of TCGA high-grade serous ovarian carcinoma (HGS-OvCa) patient samples have provided additional insight into CDK12 function (The Cancer Genome Atlas Research Network 2011). In one study, the authors propose that CDK12 is a putative tumor suppressor gene in ovarian carcinoma because *CDK12* mutations were homozygous in seven out of 12 cases (Carter et al. 2012). In fact, we found that 941/944 WGS reads for the lung tumor sample were for the mutated *CDK12*. Taking into account the presence of normal tissue contamination, *CDK12* appears to be biallelically mutated in our patient's lung tumor. In another study, the authors found that samples containing *CDK12* mutations showed a coordinated down-regulation of DDR genes (Ekumi et al. 2015). We performed quantitative PCR and found that the lung tumor that carries the mutated *CDK12* expresses significantly lower levels in nine out of 11 of these DDR genes compared with the lymph node. In fact, although *BRCA1* and *BRCA2* were not affected in the *CDK12* mutant HGS-OvCa samples, their expression was reduced in our patient's lung tumor sample. Hence, the G879V mutation, specifically in this patient's lung tumor, may explain the genomic instability resulting in increased regions of amplification and deletion in the lung metastatic site compared with the lymph node. More importantly, this mutation and the associated reduced expression of key DDR genes may have accounted for the increased sensitivity of the lung metastatic sites to the chemotherapy combination regimens used along the course of this patient's treatment. It is noteworthy that there were no metastatic sites present in the lung at autopsy.

Our results reveal a striking difference in somatic copy-number alterations between the lung tumor and the metastatic lymph node. The lung tumor displays significantly more regions with copy-number amplifications and deletions (amplifications 258 vs. 72, and deletions 116 vs. 9) (Supplemental Tables S2A,B). In the lung tumor genome, we used Strelka to identify a large number of somatic variants with loss of heterozygosity (LOH) (SNVs and indels heterozygous in normal blood, but homozygous in the tumor). There were 90,975 LOH SNVs and 757 LOH indels identified in the lung tumor genome, whereas only 826 and three were identified in the metastatic lymph node genome. Because a majority (>90%) of these LOH sites are SNPs according to dbSNP version 137, these LOH sites were most likely caused by CNVs in the tumor genomes. Intersecting the chromosomal positions of these LOH sites with the CNV regions identified by Bayesian information criterion sequencing (BIC-seq), a large portion (>33%) of the LOH SNVs and indels were found in these CNV regions (Supplemental Tables S7A,B). Because this heterogeneity could be due to disparity in tumor content between the two sites, we used CNAnorm (Gusnanto et al. 2012) to determine tumor content. However, we found that the tumor content in the lung tumor was ~56%, whereas the tumor content in the metastatic lymph node was nearly 32% (Supplemental Fig. S14), indicating that differences in tumor content cannot completely account for the heterogeneity we see and may be due to a higher degree

of genomic instability in the lung tumor. Here we identified several regions with significant copy-number alterations containing genes previously shown to harbor frequent copy-number alterations in lung adenocarcinoma (Imielinski et al. 2012). There were copy-number gains for *ERBB2* and *CCND1* and copy-number loss for *CDKN2A* in the lung tumor. In the lymph node, there were copy-number gains for *ERBB2*, *MYC*, and *CCNE1* and copy-number loss for *TP53*.

TP53 with a three-amino-acid deletion (E339–F341) within the tetramerization domain α-helix was virtually the only form expressed in the lung tumor (Fig. 6E). In fact, 27 out of 35 WGS reads were of the mutated form, suggesting that this tumor suppressor may have been biallelically mutated in this tumor. Somatic mutations within the TP53 tetramirization domain are rare in cancer (Chene 2001; Kamada et al. 2015). However, TP53 tetramerization is essential for DNA binding and transcriptional activity, as well as for a variety of its posttranslational modifications and proteasome-mediated degradation (Pietenpol et al. 1994; Kawaguchi et al. 2005; Itahana et al. 2009; Lang et al. 2014). Two out of three of the amino acids deleted (M340 and F341) are hydrophobic residues previously shown to mediate TP53 tetramerization (Jeffrey et al. 1995; Kamada et al. 2011), suggesting that the *TP53*-Del E339-F341 mutation may have resulted in diminished tumor suppressor function in this patient's lung tumor. Whereas *TP53* was mutated in the lung tumor, a different tumor suppressor, *MLL3*, was mutated in the lymph node. *MLL3* with a C>A (S2663*) nonsense mutation is expressed together with wild type in the lymph node. This mutated *MLL3* encodes a truncated protein lacking the FY-rich and SET methyltransferase domains. *MLL3* carboxy-terminal truncation mutants exist in a variety of cancer types (COSMIC) including lung adenocarcinoma (Q2363*; COSM364771 [Imielinski et al. 2012] and Q2635*; COSM745879; TCGA-39-5036-01). Parsons et al. found *MLL2* and *MLL3* inactivating mutations in 16% of medulloblastoma patients and also found that 3/88 have *MLL3* truncating mutations predicted to abolish key methyltransferase activity; they conclude that *MLL3* is a tumor suppressor gene inactivated by mutation (Parsons et al. 2011). Moreover, *MLL3* haploinsufficiency, in cooperation with NF1 suppression and TP53 deficiency, promoted acute myeloid leukemia in a mouse model (Chen et al. 2014). Although our patient's lymph node expressed roughly equal proportions of both wild-type and truncated *MLL3* (Supplemental Fig. S7C), this may be sufficient, in the context of reduced *TP53* levels (Fig. 6C,D) together with aberrant signaling due to expression of a constitutively active *ERBB2* (Fig. 4C) to impact his tumorigenic phenotype.

A variety of resistance mechanisms have been described for the ERBB2-directed therapies used in this study; most applicable to the treatment response of our patient are those involving loss of cell cycle control. We saw deregulated expression of CDK4/6 regulatory genes in the lung tumor via *CDKN2A* copy-number loss and decreased expression together with copy-number gain and increased expression of *CCND1*. On the other hand, in the metastatic lymph node, the CDK2 pathway may have been affected by amplification and overexpression of *CCNE1*. In models of acquired resistance to ERBB2-targeted therapies, cyclin D1 was shown to be deregulated and CDK4/6 inhibition was effective at overcoming such resistance (Witkiewicz et al. 2014). However, tumors with deregulated expression of *CCNE1* would be resistant to such targeted agents, suggesting that treatment with CDK2 inhibitors may be a valid strategy in such patients (Scaltriti et al. 2011).

Taken together, our findings provide additional evidence for heterogeneity among different tumor sites within one patient and elucidate how this may potentially impact the selection of treatment strategies (Jamal-Hanjani et al. 2014). Although we saw such a high degree of heterogeneity in our patient, there seems to have been selection for specific pathways. The novel *PLAG1-ACTA2* translocation common to both disease sites may explain the early metastatic spread in this patient. Thus, the lung and lymph node metastatic sites likely evolved from two discrete clones that diverged quite early in the disease and progressed

independently. The consistent responses to ERBB2-directed therapies, at least in combination treatments, suggest a driver role for ERBB2 in tumor development and/or metastasis. CDK12 with the G879V mutation and the associated lower expression of DDR genes may explain the sensitivity to chemotherapy and cure of the lung metastases in this patient. On the other hand, resistance to HER2-directed therapies, perhaps due to *CCNE1* amplification, together with a functional DDR may explain why the lymph nodes were refractory to combination treatments. Although the specific cell cycle regulatory genes affected by copy number gains and losses differ between the lung tumor and lymph node, there is a commonality in the pathways involved. The ultimate effect of these changes would be disruption of normal cell cycle regulation, a hallmark event in tumorigenesis (Fig. 7). This underscores the importance of studying the impact of such extensive genomic heterogeneity in the context of pathway alterations and specific hallmarks of tumorigenesis. It is clear, however, that understanding the genomic differences between primary and metastatic sites or between different metastatic sites will require the sequencing of a substantial number of matched tumor genomes.

METHODS

Sample Preparation

Tumor samples were obtained, with written informed consent, from a patient who participated in a molecular profiling study at the National Cancer Institute (NCI). The metastatic tumor tissues were collected from the left lung lower lobe wedge (L) and a PET avid right neck lymph node resection (LN1), which were used for WGS. The surgically collected tumor tissues were snap-frozen in liquid nitrogen, and a pathologist from a medical pathology laboratory at NCI examined the tumor cellularity. The All Prep DNA/RNA/protein kit (QIAGEN) was used to simultaneously isolate genomic DNA and total RNA from tumor samples and the DNAeasy Blood and Tissue kit (QIAGEN) was used to isolate genomic DNA from blood. The QIAamp DNA FFPE Tissue kit (QIAGEN) was used to purify DNA from the diagnostic FFPE LBX. Time points of all tissue procurement are shown in Figure 1A.

Whole-Genome and Whole-Exome Sequencing

We performed WGS on metastatic tumor tissues collected from the left lower lobe lung wedge (L) resection, a PET avid right neck lymph node resection (LN1), and whole blood as normal control. WES was performed on the initial diagnostic FFPE LBX and whole blood with the Agilent SureSelect Human All Exon V5 kit (Agilent Technologies). Paired-end sequencing was done on Illumina's next-generation sequencing platforms (HiSeq2000 and NextSeq500, Illumina). The sequence reads were aligned against the reference human genome (hg19) with BWA (Li et al. 2009) and further processed following GATK's best practice pipeline (McKenna et al. 2010) (see Supplemental Methods for additional information).

Identification of Somatic Alterations in Tumors

Somatic SNVs and indels in the tumor exome and genomes were identified with Strelka (Saunders et al. 2012). snpEff and snpSift (Cingolani et al. 2012) were used for variant functional annotation and filtration. Somatic CNVs were analyzed using BIC-seq (Xi et al. 2011) and CNV-seq (Xie and Tammi 2009). Somatic SVs were identified with BreakDancer (Chen et al. 2009) and CREST (Wang et al. 2011). SVs identified by both BreakDancer and CREST were chosen for further analyses. Array comparative genomic hybridization (aCGH) was performed on both lung tumor (L) and metastatic lymph node (LN1) using Agilent's SurePrint G3 Human CGH Microarray kit, 180K (Agilent Technologies) (see Supplemental Methods for additional information).

Validation of Key Somatic Variants

We used Ion AmpliSeq Custom Panel and Ion Torrent PGM (Life Technologies) deep sequencing to validate all potentially function-altering somatic variants (missense, nonsense, and splicing site SNVs and in-frame and frameshift short indels). Barcoded libraries were pooled and subjected to sequencing on Ion 318 Chip using Ion Torrent PGM following the manufacturer's instructions. Chromosomal translocations were validated using PCR amplification of genomic DNA and reverse-transcribed cDNA from total RNA using specifically designed PCR primers (Supplemental Table S8A,B). Sanger sequencing was performed to confirm the breakpoints and exon fusions to the nucleotide level of resolution. FISH analysis was performed to detect the *PLAG1-ACTA2* translocation.

FISH Analysis of the *PLAG1* (8q21.1) and *ACTA2* (10q23.31) Gene Fusion on FFPE Tissue Sections

The *PLAG1* and *ACTA2* gene fusion was confirmed by FISH on tumor specimens and metastatic lymph nodes collected in 2011, 2013, and 2015. For interphase FISH analysis, the following probes were purchased from Empire Genomics: *ACTA2* (5′ Spectrum Green/3′ Spectrum Red) break-apart probe, *PLAG1* (5′ Spectrum Green/3′ Spectrum Red) break-apart probe, BAC DNA probe RP11-248B17, labeled with Spectrum Gold. Gene rearrangements for *ACTA2* were tested using *ACTA2* break-apart probes. Gene rearrangements for *PLAG1* were tested using *PLAG1* break-apart probes. *PLAG1-ACTA2* fusion was detected using the *PLAG1*-5′ probe (Spectrum Gold) combined with the *ACTA2* BA probes (5′ Spectrum Green/3′ Spectrum Red) (see Supplemental Methods for additional information).

Generation of Stable Transfectants Overexpressing ACTA2 or PLAG1-ACTA2

The open reading frame (ORF) of *ACTA2* and *PLAG1-ACTA2* was amplified by PCR using in vitro mutagenesis to introduce a strong Kozak consensus sequence before the initiating ATG of each gene. These fragments were then cloned into the pCR8GWTOPO entry vector, using TOPO TA Cloning (Invitrogen) and the ORFs confirmed by Sanger sequencing. Multisite gateway recombinational cloning (LR Clonase II Plus Enzyme Mix, Invitrogen) was used to generate *ACTA2_CMV51_*pDest-665 and *PLAG1-ACTA2_CMV51_*pDest-665 lentiviral expression constructs by combining each entry vector with the *C413-E36 CMV51p>* entry vector and the pDest-665 destination vector (gift from PEL). The HPL1D (immortalized human small airway epithelial cell line) cells stably expressing either *ACTA2* or *PLAG1-ACTA2* were generated by transduction of the cells with lentivirus particles harboring the expression constructs followed by selection using 5 µg/mL of blasticidin.

Western Immunoblot and Simple Western Analyses

For western blot analysis, 80 µg of lysates were electrophoresed in SDS-PAGE (4%–15%) followed by wet transfer to nitrocellulose membrane. Anti-ACTA monoclonal antibody (1D11-2B9, Sigma-Aldrich) and anti-mouse secondary antibodies (1:5000) coupled to horseradish peroxidase (HRP) were used to develop the blots.

Simple western (ProteinSimple) assay was performed on 40 ng protein lysates (or 10 ng for ERBB2 target). The primary antibodies used in the study were anti-EGFRpY1068, anti-ERK1/2, anti-phospho-ERK1/2, anti-AKT, anti-AKTpS473, anti-PLCγ, anti-PLCγ pY783, anti-SRC, anti-SRC pY416 (Cell Signaling Technology), anti-EGFR (BD Biosciences Pharmingen), HER2 (Thermo Fisher), HER2 pY1248 (R&D systems), and Rho GDI (Sigma-Aldrich) (see Supplemental Methods for additional information).

Chemotaxis and Invasion Assays

MCF10A cells were serum starved overnight before the day of the experiment. Cells were dislodged with trypsin-EDTA solution, then neutralized with trypsin neutralization solution (Lonza). Three hundred microliters of cell suspension (assay medium without EGF) containing 10^5 cells were added in the upper chamber of 8.0-µm pore size inserts (BD with fluoroblok). The lower chamber contained 600 µL of complete medium with EGF. For invasion assays, BD inserts with basal Matrigel were used. Cells were allowed to migrate for 12 or 24 h toward the EGF gradient depending on the assay. Cells were then fixed with 4% paraformaldehyde and stained with DAPI. Nonmigrating cells were removed with a cotton swab. Migrating cells were then counted under fluorescent microscope.

Real-Time PCR Assays for Expression Analyses

Real-time PCR was conducted in 384-well plates using a ViiA7 Real-time PCR system (Applied Biosystems). Singleplex reactions (10 µL) containing an FAM-MGB expression assay for the gene of interest (see Supplemental Table S8D for TaqMan assays used) or an endogenous control was performed using 4.5 ng template and 1× Universal Master Mix (Applied Biosystems-without Amp Erase UNG) (see Supplemental Methods for additional information).

Molecular Modeling

Active conformation of CDK12 with cyclin K was obtained from PDB structure 4CXA (chains A and B) (Dixon-Clarke et al. 2015), respectively. FoldX4 program (Guerois et al. 2002) was used to repair the initial structure, generate the structure of the point mutant, and estimate changes in free energy upon mutation (ΔΔG). Contacting residues were determined based on 4 Å distance between any heavy atoms of the amino acid side chains criterion. Visualization and structure analysis was performed in Visual Molecular Dynamics (VMD) (Humphrey et al. 1996) sequences displayed using TexShade (Beitz 2000) (see Supplemental Methods for additional information).

ADDITIONAL INFORMATION

care are acknowledged. Patient and family members are acknowledged for their generous involvement in the clinical protocol.

Author Contributions

R.B., S.G., C.M.C., J.K., and U.G. designed the study. R.B., C.M.C., T.K.M., A.V., Z.A., Y.S., S.P., K.C., S.B., J.B., K.-I.H., J.C., S.P., and U.G. performed experiments and analyzed data. S.G., A.K.S., and A.R.P. analyzed data. C.A.C., A.T., A.R., J.C.Y., G.G., D.S.S., and U.G. participated in the treatment of patients and tissue collection. R.B., S.G., C.M.C., and U.G. wrote the manuscript. All coauthors assisted in preparing and reviewing the manuscript.

Funding

REFERENCES

Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, et al. 2013. Signatures of mutational processes in human cancer. *Nature* **500:** 415–421.

Beitz E. 2000. TEXshade: shading and labeling of multiple sequence alignments using LATEX2ε. *Bioinformatics* **16:** 135–139.

Blazek D, Kohoutek J, Bartholomeeusen K, Johansen E, Hulinkova P, Luo Z, Cimermancic P, Ule J, Peterlin BM. 2011. The cyclin K/Cdk12 complex maintains genomic stability via regulation of expression of DNA damage response genes. *Genes Dev* **25:** 2158–2172.

Bosken CA, Farnung L, Hintermair C, Merzel Schachter M, Vogel-Bachmayr K, Blazek D, Anand K, Fisher RP, Eick D, Geyer M. 2014. The structure and substrate specificity of human Cdk12/cyclin K. *Nat Commun* **5:** 3505.

Campbell PJ, Yachida S, Mudie LJ, Stephens PJ, Pleasance ED, Stebbings LA, Morsberger LA, Latimer C, McLaren S, Lin ML, et al. 2010. The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467:** 1109–1113.

Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, Laird PW, Onofrio RC, Winckler W, Weir BA, et al. 2012. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* **30:** 413–421.

Chen K, Wallis JW, McLellan MD, Larson DE, Kalicki JM, Pohl CS, McGrath SD, Wendl MC, Zhang Q, Locke DP, et al. 2009. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* **6:** 677–681.

Chen C, Liu Y, Rappaport AR, Kitzing T, Schultz N, Zhao Z, Shroff AS, Dickins RA, Vakoc CR, Bradner JE, et al. 2014. MLL3 is a haploinsufficient 7q tumor suppressor in acute myeloid leukemia. *Cancer Cell* **25:** 652–665.

Chene P. 2001. The role of tetramerization in p53 function. *Oncogene* **20:** 2611–2617.

Cingolani P, Patel VM, Coon M, Nguyen T, Land SJ, Ruden DM, Lu X. 2012. Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift. *Front Genet* **3:** 35.

Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD, Cibulskis K, Sougnez C, Greulich H, Muzny DM, Morgan MB, et al. 2008. Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455:** 1069–1075.

Dixon-Clarke SE, Elkins JM, Cheng SW, Morin GB, Bullock AN. 2015. Structures of the CDK12/CycK complex with AMP-PNP reveal a flexible C-terminal kinase extension important for ATP binding. *Sci Rep* **5:** 17122.

Ekumi KM, Paculova H, Lenasi T, Pospichalova V, Bosken CA, Rybarikova J, Bryja V, Geyer M, Blazek D, Barboric M. 2015. Ovarian carcinoma CDK12 mutations misregulate expression of DNA repair genes via deficient formation and function of the Cdk12/CycK complex. *Nucleic Acids Res* **43:** 2575–2589.

Ferlay J, Soerjomataram I, Ervik M, Forman D, Bray F. 2013. GLOBOCAN 2012 v1.0, Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. http://globocan.iarc.fr/Default.aspx.

Fisher R, Pusztai L, Swanton C. 2013. Cancer heterogeneity: implications for targeted therapeutics. *Br J Cancer* **108:** 479–485.

Gerlinger M, Rowan AJ, Horswell S, Larkin J, Endesfelder D, Gronroos E, Martinez P, Matthews N, Stewart A, Tarpey P, et al. 2012. Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N Engl J Med* **366:** 883–892.

Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, et al. 2012. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150:** 1121–1134.

Guerois R, Nielsen JE, Serrano L. 2002. Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. *J Mol Biol* **320:** 369–387.

Gusnanto A, Wood HM, Pawitan Y, Rabbitts P, Berri S. 2012. Correcting for cancer genome size and tumour cell content enables better estimation of copy number alterations from next-generation sequence data. *Bioinformatics* **28:** 40–47.

Hartman Z, Zhao H, Agazie YM. 2013. HER2 stabilizes EGFR and itself by altering autophosphorylation patterns in a manner that overcomes regulatory mechanisms and promotes proliferative and transformation signaling. *Oncogene* **32:** 4169–4180.

Hirsch FR, Varella-Garcia M, Cappuzzo F. 2009. Predictive value of EGFR and HER2 overexpression in advanced non-small-cell lung cancer. *Oncogene* **28:** S32–S37.

Humphrey WA, Schulten K. 1996. VMD: visual molecular dynamics. *J Mol Graphics* **14:** 33–38.

Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, et al. 2012. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150:** 1107–1120.

Itahana Y, Ke H, Zhang Y. 2009. p53 Oligomerization is essential for its C-terminal lysine acetylation. *J Biol Chem* **284:** 5158–5164.

Jamal-Hanjani M, Hackshaw A, Ngai Y, Shaw J, Dive C, Quezada S, Middleton G, de Bruin E, Le Quesne J, Shafi S, et al. 2014. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. *PLoS Biol* **12:** e1001906.

Jeffrey PD, Gorina S, Pavletich NP. 1995. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* **267:** 1498–1502.

Joshi PM, Sutor SL, Huntoon CJ, Karnitz LM. 2014. Ovarian cancer-associated mutations disable catalytic activity of CDK12, a kinase that promotes homologous recombination repair and resistance to cisplatin and poly(ADP-ribose) polymerase inhibitors. *J Biol Chem* **289:** 9247–9253.

Kamada R, Nomura T, Anderson CW, Sakaguchi K. 2011. Cancer-associated p53 tetramerization domain mutants: quantitative analysis reveals a low threshold for tumor suppressor inactivation. *J Biol Chem* **286:** 252–258.

Kamada R, Toguchi Y, Nomura T, Imagawa T, Sakaguchi K. 2015. Tetramer formation of tumor suppressor protein p53: structure, function, and applications. *Biopolymers* doi:10.1002/bip.22772.

Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, Xie M, Zhang Q, McMichael JF, Wyczalkowski MA, et al. 2013. Mutational landscape and significance across 12 major cancer types. *Nature* **502:** 333–339.

Kawaguchi T, Kato S, Otsuka K, Watanabe G, Kumabe T, Tominaga T, Yoshimoto T, Ishioka C. 2005. The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library. *Oncogene* **24:** 6976–6981.

Kelly RJ, Carter C, Giaccone G. 2010. Personalizing therapy in an epidermal growth factor receptor-tyrosine kinase inhibitor-resistant non-small-cell lung cancer using PF-00299804 and trastuzumab. *J Clin Oncol* **28:** e507–e510.

Kelly RJ, Carter CA, Giaccone G. 2012. HER2 mutations in non-small-cell lung cancer can be continually targeted. *J Clin Oncol* **30:** 3318–3319.

Kozak M. 1996. Interpreting cDNA sequences: Some insights from studies on translation. *Mamm Genome* **7:** 563–574.

Lang V, Pallara C, Zabala A, Lobato-Gil S, Lopitz-Otsoa F, Farras R, Hjerpe R, Torres-Ramos M, Zabaleta L, Blattner C, et al. 2014. Tetramerization-defects of p53 result in aberrant ubiquitylation and transcriptional activity. *Mol Oncol* **8:** 1026–1042.

Lee JW, Soung YH, Seo SH, Kim SY, Park CH, Wang YP, Park K, Nam SW, Park WS, Kim SH, et al. 2006. Somatic mutations of ERBB2 kinase domain in gastric, colorectal, and breast carcinomas. *Clin Cancer Res* **12:** 57–61.

Lee W, Jiang Z, Liu J, Haverty PM, Guan Y, Stinson J, Yue P, Zhang Y, Pant KP, Bhatt D, et al. 2010. The mutation spectrum revealed by paired genome sequences from a lung cancer patient. *Nature* **465:** 473–477.

Lee HW, Seol HJ, Choi YL, Ju HJ, Joo KM, Ko YH, Lee JI, Nam DH. 2012. Genomic copy number alterations associated with the early brain metastasis of non-small cell lung cancer. *Int J Oncol* **41:** 2013–2020.

Lee HW, Park YM, Lee SJ, Cho HJ, Kim DH, Lee JI, Kang MS, Seol HJ, Shim YM, Nam DH, et al. 2013. Alpha-smooth muscle actin (ACTA2) is required for metastatic potential of human lung adenocarcinoma. *Clin Cancer Res* **19:** 5879–5889.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25:** 2078–2079.

Lynch TJ, Bell DW, Sordella R, Gurubhagavatula S, Okimoto RA, Brannigan BW, Harris PL, Haserlat SM, Supko JG, 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.

Malumbres M. 2014. Cyclin-dependent kinases. *Genome Biol* **15:** 122.

Martini M, Gnann A, Scheikl D, Holzmann B, Janssen KP. 2011. The candidate tumor suppressor SASH1 interacts with the actin cytoskeleton and stimulates cell-matrix adhesion. *Int J Biochem Cell Biol* **43:** 1630–1640.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20:** 1297–1303.

Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, Boca SM, Carter H, Samayoa J, Bettegowda C, et al. 2011. The genetic landscape of the childhood cancer medulloblastoma. *Science* **331:** 435–439.

Pietenpol JA, Tokino T, Thiagalingam S, el-Deiry WS, Kinzler KW, Vogelstein B. 1994. Sequence-specific transcriptional activation is essential for growth suppression by p53. *Proc Natl Acad Sci* **91:** 1998–2002.

Saito S, Matsui H, Kawano M, Kumagai K, Tomishige N, Hanada K, Echigo S, Tamura S, Kobayashi T. 2008. Protein phosphatase 2Cε is an endoplasmic reticulum integral membrane protein that dephosphorylates the ceramide transport protein CERT to enhance its association with organelle membranes. *J Biol Chem* **283:** 6584–6593.

Saunders CT, Wong WSW, Swamy S, Becq J, Murray LJ, Cheetham RK. 2012. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28:** 1811–1817.

Scaltriti M, Eichhorn PJ, Cortes J, Prudkin L, Aura C, Jimenez J, Chandarlapaty S, Serra V, Prat A, Ibrahim YH, et al. 2011. Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2+ breast cancer patients. *Proc Natl Acad Sci* **108:** 3761–3766.

Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba KM II, Lee H, Toyooka S, Shimizu N, et al. 2005. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* **97:** 339–346.

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S-i, Watanabe H, Kurashina K, Hatanaka H, et al. 2007. Identification of the transforming *EML4-ALK* fusion gene in non-small-cell lung cancer. *Nature* **448:** 561–566.

Subramanian J, Govindan R. 2007. Lung cancer in never smokers: a review. *J Clin Oncol* **25:** 561–570.

Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, et al. 2012. RET, ROS1 and ALK fusions in lung cancer. *Nat Med* **18:** 378–381.

Thean LF, Loi C, Ho KS, Koh PK, Eu KW, Cheah PY. 2010. Genome-wide scan identifies a copy number variable region at 3q26 that regulates *PPM1L* in *APC* mutation-negative familial colorectal cancer patients. *Genes Chromosomes Cancer* **49:** 99–106.

The Cancer Genome Atlas Research Network. 2011. Integrated genomic analyses of ovarian carcinoma. *Nature* **474:** 609–615.

The Cancer Genome Atlas Research Network. 2014. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511:** 543–550.

Thomas A, Rajan A, Lopez-Chavez A, Wang Y, Giaccone G. 2013. From targets to targeted therapies and molecular profiling in non-small cell lung carcinoma. *Ann Oncol* **24:** 577–585.

Vignot S, Besse B, Andre F, Spano J-P, Soria J-C. 2012. Discrepancies between primary tumor and metastasis: a literature review on clinically established biomarkers. *Crit Rev Oncol Hematol* **84:** 301–313.

Vignot S, Frampton GM, Soria J-C, Yelensky R, Commo F, Brambilla C, Palmer G, Moro-Sibilot D, Ross JS, Cronin MT, et al. 2013. Next-generation sequencing reveals high concordance of recurrent somatic alterations between primary tumor and metastases from patients with non-small-cell lung cancer. *J Clin Oncol* **31:** 2167–2172.

Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Kinzler KW. 2013. Cancer genome landscapes. *Science* **339:** 1546–1558.

Wang J, Mullighan CG, Easton J, Roberts S, Heatley SL, Ma J, Rusch MC, Chen K, Harris CC, Ding L, et al. 2011. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat Methods* **8:** 652–654.

Witkiewicz AK, Cox D, Knudsen ES. 2014. CDK4/6 inhibition provides a potent adjunct to Her2-targeted therapies in preclinical breast cancer models. *Genes Cancer* **5:** 261–272.

Wu JY, Wu SG, Yang CH, Chang YL, Chang YC, Hsu YC, Shih JY, Yang PC. 2011. Comparison of gefitinib and erlotinib in advanced NSCLC and the effect of EGFR mutations. *Lung Cancer* **72:** 205–212.

Xi R, Hadjipanayis AG, Luquette LJ, Kim T-M, Lee E, Zhang J, Johnson MD, Muzny DM, Wheeler DA, Gibbs RA, et al. 2011. Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proc Natl Acad Sci* **108:** 1128–1136.

Xie C, Tammi MT. 2009. CNV-seq, a new method to detect copy number variation using high-throughput sequencing. *BMC Bioinformatics* **10:** 80–80.

Zeller C, Hinzmann B, Seitz S, Prokoph H, Burkhard-Goettges E, Fischer J, Jandrig B, Schwarz LE, Rosenthal A, Scherneck S. 2003. SASH1: a candidate tumor suppressor gene on chromosome 6q24.3 is down-regulated in breast cancer. *Oncogene* **22:** 2972–2983.