Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers

Collin M Blakely1,2,11, Thomas B K Watkins3,11, Wei Wu1,2,11, Beatrice Gini1,2,11, Jacob J Chabon4, Caroline E McCoacht5, Nicholas McGranahan3, Gareth A Wilson3, Nicolai J Birkbak3, Victor R Olivas1,2, Julia Rotow1,2, Ashley Maynard1,2, Victoria Wang1,2, Matthew A Gubens1,2, Kimberly C Banks6,7, Richard B Lanman6,7, Aleah F Caulin7, John St John7, Anibal R Cordero7, Petros Giannikopoulos7, Andrew D Simmons8, Philip C Mack9, David R Gandara9, Hatim Husain10, Robert C Doebele5, Jonathan W Riess9, Maximilian Diehn4,8, Charles Swanton9 & Trever G Bivona1,2

A widespread approach to modern cancer therapy is to identify a single oncogenic driver gene and target its mutant-protein product (for example, EGFR-inhibitor treatment in EGFR-mutant lung cancers). However, genetically driven resistance to targeted therapy limits patient survival. Through genomic analysis of 1,122 EGFR-mutant lung cancer cell-free DNA samples and whole-exome analysis of seven longitudinally collected tumor samples from a patient with EGFR-mutant lung cancer, we identified critical co-occurring oncogenic events present in most advanced-stage EGFR-mutant lung cancers. We defined new pathways limiting EGFR-inhibitor response, including WNT/β-catenin alterations and cell-cycle-gene (CDK4 and CDK6) mutations. Tumor genomic complexity increases with EGFR-inhibitor treatment, and co-occurring alterations in CTNNB1 and PIK3CA exhibit nonredundant functions that cooperatively promote tumor metastasis or limit EGFR-inhibitor response. This study calls for revisiting the prevailing single-gene driver-oncogene view and links clinical outcomes to co-occurring genetic alterations in patients with advanced-stage EGFR-mutant lung cancer.

The current paradigm in cancer genetics and therapy is to view and treat oncogene-positive disease (e.g., EGFR-mutant non-small-cell lung cancer; NSCLC) primarily through the lens of one oncogenic alteration (for example, oncogenic-mutant EGFR; Supplementary Note). This approach does not address the potential risk of co-occurring genetic alterations present in the cancer and instead treats one ‘driver’ as mutually exclusive from any other. Despite limited and sporadic reports1–5, the prevalence of co-occurring genetic alterations that affect clinical outcomes in advanced-stage lung cancers with a primary oncogenic driver is largely unknown, although recent work has suggested a potential role of TP53 mutations6–7. This lack of data on the prevalence and effects of multiple co-occurring genetic events exists not only for treatment-naïve cancers but also for cancers that have acquired resistance to the initial targeted therapy (for example, cancers with EGFR p.Thr790Met). An open question in the field is to what extent co-occurring genetic alterations cooperate with a primary driver gene (for example, mutant EGFR) in promoting tumor progression and therapy resistance in both the targeted-therapy-naïve and acquired-resistance settings (Supplementary Note).

This knowledge gap exists because, to date, large-scale genome-sequencing efforts regarding NSCLC have included mostly early-stage tumors8, thus leading to the current prevailing model of one driver oncogene in each individual cancer. Yet, it is the patients with advanced-stage rather than early-stage NSCLC who are treated with targeted therapy. In this study, we tested the hypothesis that co-occurring genetic alterations commonly exist and cooperate with the primary driver as co-drivers, thereby promoting tumor progression and limiting targeted therapy response. Here, we link clinical outcomes to genetic co-alterations in the largest cohort of advanced-stage EGFR-mutant lung cancers profiled by multiplex sequencing to date.

RESULTS
cfDNA analysis of advanced EGFR-mutant lung cancers

To determine the prevalence of co-occurring genetic alterations in patients with advanced-stage EGFR-mutant lung cancers, we undertook a large-scale analysis, using a clinically validated cell-free (cf) DNA assay (Supplementary Tables 1 and 2, and Online Methods). This cfDNA exome platform has been approved for clinical use and...
measures single-nucleotide variants, small insertions/deletions (indels), gene rearrangements/fusions, and copy number gain (CNG) across 68 clinically relevant cancer genes\textsuperscript{9,10} (Supplementary Table 2 and Online Methods). We defined the landscape of somatic genetic alterations present in 1,122 EGFR-mutation-positive and 1,008 EGFR-mutation-negative patients with advanced-stage (stage III or IV) NSCLC (Supplementary Tables 3 and 4 and Supplementary Data Sets 1 and 2).

We filtered for mutations that were nonsynonymous and were validated or predicted to affect gene function (Online Methods), a process yielding 1,122 EGFR-mutation-positive and 944 EGFR-mutation-negative cases. This data set of advanced-stage EGFR-mutant samples differs from The Cancer Genome Atlas (TCGA) and other genomic compendia of lung cancer that largely contain early-stage tumors.

Analysis of the 1,122 EGFR-mutant patient cohort showed the widespread presence of co-occurring genetic alterations, in addition to the EGFR driver mutation (Fig. 1a). The EGFR-mutant cases contained 2.58 ± 1.7 (mean ± s.e.m) genetic alterations beyond EGFR (out of the 68 genes profiled). When the EGFR mutation was included, the range of detectable alterations was 1–13. Most patients (92.9%; 1,043/1,122) had at least one additional variant of known or likely functional importance beyond the EGFR driver mutation (Supplementary Data Set 1). Most (89.8%; 3,033/3,375) of the genetic comutations present in the EGFR-mutation-positive cohort had verified or likely functional effects (on the basis of in silico modeling; Online Methods and Supplementary Data Set 1), and only 10.2% (345/3,375) of these comutations were classified as likely passenger events (with neutral or unknown functional effects). Of the mutations present in the EGFR-mutation-negative cohort, 16.1% (415/2,578) were classified as passenger events (P = 1.3 × 10\textsuperscript{−11}, two-tailed Fisher's exact test, odds ratio (OR) = 0.64, proportions test, Supplementary Data Set 2; comparison of the prevalence of mutations classified as passenger events in the EGFR-mutation-positive cohort versus the EGFR-mutation-negative cohort) (Fig. 1a,b and Supplementary Table 3). The data showed that canonical EGFR driver mutations co-occurred with oncogenic driver alterations in several other genes, including PIK3CA, BRAF, MET, MYC, CDK6, and CTNNB1. Comparison of the frequency of genetic co-alterations present in the EGFR-mutation-positive samples (n = 1,122) with those present in the stage-matched EGFR-mutation-negative samples (n = 944) showed significant enrichment for certain genetic events (q values determined by two-tailed Fisher's exact test with Benjamini–Hochberg correction for multiple hypotheses), we observed more frequent alterations in cell-cycle genes (CDK6 CNG, 43/440 versus 36/682, q = 0.08; CCNE1 CNG, 39/440 versus 39/682, q = 0.28), WNT-pathway genes (CTNNB1 oncogenic mutations, 33/440 versus 27/682, q = 0.12), genes encoding hormone-signaling proteins (AR somatic mutations, 30/440 versus 27/682, q = 0.22), epigenetic–pathway genes (MYC CNG, 47/440 versus 41/682, q = 0.08), and KRAS (CNG and oncogenic mutations, 21/440 versus 17/682, q = 0.24), PDGFRα (CNG and oncogenic mutations, 21/440 versus 11/682, q = 0.06), and BRCAnon-synonymous alterations (31/440 versus 24/682, q = 0.10) (Fig. 2a–c, Supplementary Fig. 1a–c and Supplementary Table 4). Corresponding pathway-level differences in cell-cycle genes (106/440 versus 117/682, q = 0.07), genes encoding DNA-repair proteins (48/440 versus 51/682, q = 0.16), epigenetic–pathway genes (62/440 versus 68/682, q = 0.16), WNT-pathway genes (68/440 versus 76/682, q = 0.15), and genes encoding hormone-signaling proteins (30/440 versus 29/682, q = 0.18) were also observed (q values determined by two-tailed Fisher's exact test with Benjamini–Hochberg correction for multiple hypotheses; Fig. 2f). The finding of co-occurring oncogenic mutations in KRAS and EGFR was consistent with preclinical data\textsuperscript{13,14}. In a subgroup analysis of cases positive for the EGFR p.Thr790Met mutation (n = 15), which can arise upon acquired resistance to osimertinib (the approved third-generation EGFR TKI with activity against EGFR p.Thr790Met\textsuperscript{15}), there were recurrent activating alterations in MAPK-pathway genes (including KRAS CNG and oncogenic mutations), cell-cycle genes (CDK4 and CDK6), and AR CNGs (Supplementary Fig. 1c). These data identified extensive co-occurring alterations in advanced-stage EGFR-mutant NSCLCs, even those with EGFR TKI–resistant forms of mutant EGFR (p.Thr790Met and p.Cys797Ser).
**Figure 1** Co-occurring genomic alterations detectable in cfDNA of patients with advanced-stage EGFR-mutant-positive compared with EGFR-mutant-negative NSCLC. (a,b) Frequency of genomic alterations: nonsynonymous somatic variants of predicted functional significance (SNV; Online Methods), CNG, indels, or gene rearrangements (fusion) in the cancer-related genes listed (Supplementary Table 2), detected by next-generation sequencing of circulating tumor DNA from 1,122 patients with advanced-stage EGFR-mutant positive NSCLC (a) compared with a cohort of 944 patients with EGFR-mutant-negative NSCLC (b) (Supplementary Data Sets 1 and 2). Co-occurring alterations that occurred in at least 5% of EGFR-mutant-positive cases are shown. Asterisks indicate statistically significant differences between the cohorts (q <0.02). (c) Gene alterations with higher frequency in EGFR-mutant-positive compared with EGFR-mutant-negative samples (two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypothesis testing (q values)) performed to identify statistically significant differences in TP53, CDK6, CTNNB1, and AR. (d) Lolliplots of gene-level alterations in EGFR-mutant-positive compared with EGFR-mutant-negative samples. The functional importance of somatic variants is indicated on the basis of analysis described in Online Methods. (e) Differences in pathway-level alterations between EGFR-mutant-positive and EGFR-mutant-negative cases (two-tailed Fisher’s exact test comparing EGFR-mutant-positive to EGFR-mutant-negative samples with Benjamini–Hochberg correction for multiple hypothesis testing (q values)). Additional data in Supplementary Tables 1–3 and Supplementary Data Sets 1 and 2.
Methods. (in at least 5% of p.Thr790Met-positive cases) of circulating tumor DNA are indicated in EGFR p.T790M-positives (n = 440) (a) and EGFR p.Thr790Met-negative (n = 682) (b) cohorts (q values determined by two-tailed by Fisher’s exact test with Benjamini–Hochberg correction for multiple hypothesis testing). Asterisks indicate statistically significant differences between the cohorts; *p < 0.20.

(c) Frequency (percentage) of gene-level alterations detectable in the cfDNA of EGFR p.Thr790Met-positive compared with EGFR p.Thr790Met-negative samples. Somatic alterations altered in incidence of this alteration (55–65%) at acquired resistance to TKI treatment (31/53 versus 0/21, *p = 0.01, F = 4.2, degrees of freedom (df) = 97, one-way ANOVA; Supplementary Fig. 1–a). (d) Lollipops of gene-level alterations in EGFR p.Thr790Met-positive compared with EGFR p.Thr790Met-negative samples. Somatic alterations in CTNNB1 (d) and KRAS (e) are indicated. The functional importance of somatic variants is indicated on the basis of analysis described in Online Methods. (f) Differences in pathway-level alterations between EGFR p.Thr790Met-positive and EGFR p.Thr790Met-negative cases, as determined by two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypothesis testing (q value). Additional data in Supplementary Fig. 1, Supplementary Table 4, and Supplementary Data Set 1.

who were TKI naive (n = 21), who showed disease progression on first-line TKI treatment (n = 53), and who showed disease progression on second-line therapy (n = 26). The number of detectable somatic alterations increased with each line of therapy, regardless of age, sex, or tobacco exposure (pre-TKI, mean [95% confidence interval (CI)] 3.4 [2.2–4.5]; progression on first-line therapy, 3.8 [3.2–4.4]; progression on second-line therapy, 5.2 [4.1–6.2], R² = 0.064, slope 0.92, P = 0.01, F = 4.2, degrees of freedom (df) = 97, one-way ANOVA; Fig. 3b and Supplementary Fig. 2a–c). Enrichment for the EGFR p.Thr790Met alteration occurred in samples with progression on first-line EGFR TKI (31/53 versus 0/21, q = 3.6 × 10⁻⁵), as expected on the basis of the rare detection (~0.5%) of the p.Thr790Met alteration before first-generation EGFR TKI treatment² and established the incidence of this alteration (55–65%) at acquired resistance to
Figure 3  Therapy-induced evolution of genomic co-alterations detected in cfDNA of patients with advanced-stage EGFR-mutant NSCLC. cfDNA analysis of 137 samples collected from 97 patients with known clinical history (additional data in Supplementary Table 5 and Supplementary Data set 3). (a) Samples are segregated by EGFR TKI treatment; pre-TKI (n = 21), at the time of progression to first-line EGFR TKI therapy; progressive disease (PD) on first-line therapy (n = 53), or at the time of progression to second-line anticancer therapy (second- or third-generation EGFR TKI, or chemotherapy); progressive disease on second-line therapy (n = 26) Mut, mutation (indel or SNV). (b) Number of functional alterations detectable on the basis of line of therapy are indicated (mean [95% CI]). Pre-TKI (3.4, [2.2–4.5]), progressive disease on first line (3.8 [3.2–4.4]), progressive disease on second line (2.6). (c) Changes in cancer-related pathway alterations (percentage) with line of therapy. In c and d, two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypothesis testing (q values) was performed to identify statistically significant differences between pre-TKI and progressive disease on first line, between progressive disease on first line and progressive disease on second line, and between pre-TKI and progressive disease on second line. Additional data in Supplementary Figures 2–5, Supplementary Table 5, and Supplementary Data Set 3.
Effects of cfDNA detectable co-occurring genetic alterations on osimertinib clinical response in patients with advanced-stage EGFR-mutant lung cancer. (a, b) Genomic alterations detectable in cfDNA from patients with advanced EGFR-mutant NSCLC who were subsequently treated with osimertinib and exhibited a radiographic/clinical response (a) (patient response by clinician assessment, Online Methods) versus patients who did not respond (b) (on the basis of clinician assessment; Online Methods). (c) Forest plot showing effects of cfDNA detectable gene-level alterations on PFS. P values determined by Cox proportional HR with 95% CI are shown. (d) Kaplan–Meier curves showing difference in median PFS (log-rank test) in patients with cfDNA detectable alterations in CDK4 or CDK6. Alt, alteration; pts, patients; mo, months. (e) Pathway-level alterations in osimertinib responders versus nonresponders. q values were determined by two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypothesis testing. (f, g) Forest plot (f) and Kaplan–Meier curves (g) assessing the effects of the indicated cfDNA detectable pathway alterations on PFS. P values determined by Cox proportional HR with 95% CI are shown. Additional data in Supplementary Figures 4 and 5, Supplementary Table 6, and Supplementary Data Set 4.
(progression on first-line versus second-line therapy; 10/53 versus 7/26, \( q = 0.20 \)), epigenetic pathways (progression on first-line versus second-line therapy; 4/53 versus 6/26, \( q = 0.20 \)), and phosphati-
dylinositol-4,5-bisphosphate 3-kinase (PI3K) pathways (pre-TKI versus disease progression on second-line therapy; 1/21 versus 7/26, \( q = 0.20 \)) were more frequently detected in patients with progression

Figure 5 Longitudinal genomic analysis of tumor and cfDNA in a patient with EGFR-mutant lung cancer from diagnosis to death. (a) Heat map depicting the clonal status of nonsynonymous somatic mutations including SNVs, dinucleotides, and indels from each sequenced region of the patient's disease, as determined by the subclonal copy number–corrected cancer cell fraction and PyClone cross-sample clustering. Clonal mutations were detected by WES of the tumor DNA of the patient at initial presentation and surgical resection of the primary tumor and metastases analyzed by WES. Met, metastasis; LN, lymph node. (b) Phylogenetic tree illustrating the evolutionary history of the patient's disease at the level of subclonal clusters of mutations. These subclonal clusters were inferred, using PyClone, from the samples taken from the primary and different metastases at multiple time points. The mutations were clustered on the basis of their prevalence (subclonal copy number–corrected cancer cell fraction) in the sequenced cancer cell populations across all samples, and this clustering was then used to infer the founding clone (at the bottom of the tree) and subclonal clusters. (c) Pictorial representation of primary tumor and metastatic sites analyzed by WES. Met, metastasis; LN, lymph node. (d) cfDNA detectable in plasma from the patient at the indicated time points, as determined by CAPP-seq analysis\(^5\). Additional data in Supplementary Figures 6–8 and Supplementary Data Sets 5 and 6.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
Patients with gene-level MET (0/37 responders versus 5/36 nonresponders, q = 0.23, 95% CI [0–1.15] and OR undefined), NF1 (0/37 responders versus 6/36 nonresponders, q = 0.23, 95% CI [0–0.90] and OR undefined), CDK4 (0/37 responders versus 5/36 nonresponders, q = 0.23, 95% CI [0–1.15] and OR undefined), PIK3CA (1/37 responders versus 8/36 nonresponders, q = 0.23, 95% CI [0.0026–1.005] and OR = 0.124), or APC (0/37 responders versus 5/36 nonresponders, q = 0.23, 95% CI [0–1.15] OR undefined) alterations were least likely to respond to a subsequent EGFR TKI (q values determined by two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypotheses; Supplementary Fig. 3a,b). Patients whose cfDNA contained CDK4 or CDK6 gene alterations (n = 7) also
exhibited shorter progression-free survival (PFS) (hazard ratio (HR) 13.8, 95% CI [5.1–36.8], \( P = 1.4 \times 10^{-11} \), Cox proportional-hazard regression test; Supplementary Fig. 3c,d) in response to EGFR TKI treatment, as compared with patients without detectable CDK4 or CDK6 alterations (\( n = 66 \)). Pathway-level alterations in cell-cycle genes (1/37 responders versus 12/36 nonresponders, \( q = 0.006 \) (95% CI [0.0018–0.613] and OR = 0.083)), MAPK-pathway genes (3/37 responders versus 12/36 nonresponders, \( q = 0.03 \) (95% CI [0.04–1.02] and OR = 0.247), PI3K-pathway genes (1/37 responders versus 9/36 nonresponders, \( q = 0.03 \) (95% CI [0.0024–0.867]; and OR = 0.11), and WNT-pathway genes (3/37 responders versus 8/36 nonresponders, \( q = 0.19 \) (95% CI [0.058–0.695] and OR = 0.369) also correlated with a lack of response to EGFR TKI treatment (\( q \) values determined by two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypotheses; Supplementary Fig. 3e). Cell-cycle alterations (\( n = 12 \) positive, \( n = 61 \) negative, HR 2.8, 95% CI [1.4–5.9], \( P = 0.004 \), Cox proportional-hazard regression test) and MAPK-pathway alterations (\( n = 15 \) positive, \( n = 58 \) negative, HR 1.9, 95% CI [1.0–3.7], \( P = 0.04 \), Cox proportional-hazard regression test) (Supplementary Fig. 3f–h) were biomarkers of shorter PFS during subsequent EGFR TKI treatment. Patients with CDK4 or CDK6 alterations (\( n = 7 \)) exhibited lower overall survival than did patients without those alterations (\( n = 66 \)) (HR 5.4, 95% CI [1.7–18.0], \( P = 0.002 \), Cox proportional-hazard regression test; Supplementary Fig. 4). These data suggested selection for increased genetic diversity during iterative tumor progression on therapy and identified biomarkers of poor response to EGFR TKI treatment (i.e., alterations in cell-cycle and MAPK-pathway genes). We observed further evidence for this selection in a cohort of several individual clinical cases of EGFR-mutant NSCLC in which intrapatient longitudinal cfDNA profiling was performed (Supplementary Fig. 5).

cfDNA analysis linked to differential osimertinib response

Although mechanisms of acquired resistance to the third-generation EGFR TKI osimertinib have been well described\textsuperscript{14–17}, mechanisms of primary resistance have not been well characterized. We identified 41 patients who underwent cfDNA analysis before treatment with osimertinib for whom clinical response, PFS, and overall survival after subsequent osimertinib treatment were known (Supplementary Table 6 and Supplementary Data Set 4). Alterations in MET (3/21), NFI (5/21), CDK4 (3/21), CCNE1 (3/21), CDK6 (2/21), PIK3CA (6/21), and APC (5/21) were found only in patients with primary resistance to osimertinib treatment (Fig. 4a and Supplementary Data Set 4). Patients with cfDNA alterations in CDK4 or CDK6 (\( n = 5 \)) exhibited shorter PFS after osimertinib treatment than did patients without (\( n = 36 \)) detectable CDK4 or CDK6 alterations (median PFS, 0.7 months (95% CI [0.7 to not reached (NR)]) versus 11.2 months (95% CI [6.2–NR]), HR 10.3, 95% CI [3.0–34.7], \( P = 3.7 \times 10^{-6} \), Cox proportional-hazard regression test; Fig. 4c,d), although no statistically significant differences in overall survival were observed (Supplementary Fig. 4c). Pathway-level alterations in cell-cycle genes (0/20 responders versus 8/21 nonresponders, \( q = 0.03 \) (95% CI [0–0.47] and OR undetermined), MAPK-pathway genes (1/20 responders versus 7/21 nonresponders, \( q = 0.15 \) (95% CI [0.0025–1.146] and OR = 0.126), PI3K-pathway genes (0/20 responders versus 7/21 nonresponders, \( q = 0.04 \) (95% CI [0–0.667] and OR undetermined), and WNT-pathway genes (2/21 responders versus 7/21 nonresponders, \( q = 0.33 \) (95% CI [0.023–1.65] and OR = 0.26) were associated with a lack of response to osimertinib treatment (\( q \) values determined by two-tailed Fisher’s exact test with Benjamini–Hochberg correction for multiple hypotheses; Fig. 4e). Shorter PFS in response to subsequent osimertinib treatment was also associated with cell-cycle-gene alterations (\( n = 33 \) alteration negative, median PFS 11.2, 95% CI [8.8–NR] versus \( n = 8 \) alteration positive, median 1.5 months, 95% CI [0.7–NR], HR 5.4, 95% CI [2.0–14.5], \( P = 0.0002 \), Cox proportional-hazard regression test; Fig. 4f,g), with a trend toward a difference in overall survival (overall survival 17.1 versus 4.3 months, HR 1.7, 95% CI [0.6–5.2], \( P = 0.4 \), Cox proportional-hazard regression test; Supplementary Fig. 4d). These data highlight potential roles for MAPK-, PI3K-, and WNT-pathway alterations in driving primary resistance to osimertinib and uncover cell-cycle-gene aberrations (specifically in CDK4 or CDK6) as clinical biomarkers of osimertinib nonresponse (i.e., primary resistance) in advanced-stage EGFR p.Thr790Met–positive NSCLC. Thus, co-occurring genetic alterations may function as codrivers of tumor progression and drug resistance and may create genetic diversity that is advantageous for cancer evolution.

Longitudinal spatiotemporal genomic profiling

We next leveraged the uncommon opportunity to analyze a NSCLC clinical case through both tumor-based whole-exome sequencing (WES) and cfDNA profiling over 6 years of disease progression, from the initial diagnosis of surgically resectable disease through metastatic progression, first in mediastinal lymph nodes, then in the lungs, bone, and brain, during which the patient was treated with erlotinib followed by the third-generation EGFR TKI rociletinib\textsuperscript{18} (Supplementary Fig. 6). Seven tumor specimens (four lung, two
bone, and one lymph node), including four obtained at autopsy after lethal tumor progression under rociletinib treatment, and six plasma specimens were analyzed longitudinally.

The WES analysis showed that >75% of the coding mutational burden was truncal (ubiquitous and clonal) at diagnosis, but this proportion decreased to 50–58% at the time of full cancer evolution (patient death) via the emergence of subclonal mutations through tumor progression and first- and second-line EGFR TKI treatment and resistance (Fig. 5a). The genetic co-alterations present in this patient by longitudinal tumor-based exome sequencing were consistent with the results of our cfDNA analysis of the broader cohort of advanced-stage EGFR-mutant NSCLC (Figs. 1–4), with evidence of cell-cycle, WNT-pathway, and PI3K-pathway alterations. Multiple functionally relevant somatic co-alterations were present in early-stage disease (R1), including the clonal and truncal EGFR variant c.2233_2252delinsAATT;chr7:g.55242463_55242482delinsAATT (hg19) (p.Glu746_Thr751delinsLeu; NM_006218), CTNNB1 variant c.110C>T; chr3:g.41266113C>T (hg19) (p.Ser37Phe; NM_001904)19, SMAD4 variant c.437T>G; chr18:g.48575677T>G (hg19) (p.Leu146*; NM_005359) and RBM10 variant c.269C>A; chrX:g.47035294C>A (hg19) (p.Ser10*; NM_005676) as well as CDKN2A copy number loss (Fig. 5a, Supplementary Fig. 7 and Supplementary Data Sets 5 and 6). Acquisition of PRKCA variant c.1403_1404AC>TA, chr17:g.64738757-64738758AC>TA (hg19) (p.Asn468Ile; NM_002737)20 and PIK3CA variant c.317G>T; chr3:g.178916930G>T (hg19) (p.Gly106Val; NM_006218)21, and CNG in the genomic region containing EGFR, CDK6, MET, and BRAF all occurred during mediastinal lymph node metastasis (R2) (Fig. 5a, Supplementary Fig. 7 and Supplementary Data Sets 5 and 6). Progression under initial EGFR TKI (erlotinib) treatment occurred with acquisition of the EGFR variant c.2369C>T; chr7:g.55249071C>T (hg19) (p.Thr790Met; NM_005528), found in ~60% of patients with EGFR-mutant NSCLC progressing under first-generation EGFR TKI treatment11, and the persistence of additional co-alterations including CTNNB1 variant p.Ser37Phe and PIK3CA variant p.Gly106Val. Our data suggested that the PIK3CA variant p.Gly106Val arose before both EGFR TKI treatment and EGFR variant p.Thr790Met (Fig. 5a and Supplementary Fig. 8). The data suggested that the EGFR variant p.Thr790Met arose twice in this case in a previously unreported instance of independent dual clones, because it was found in metastatic sites containing PIK3CA variant p.Gly106Val (R3, left-lung progression under erlotinib; R4, left-lung progression under rociletinib, and R6, right-lung progression under rociletinib) and those that did not (R5, right-rib metastasis; R7, spine metastasis), although despite relatively deep sequencing coverage (250- to 600-fold across the tumor samples), we cannot completely rule out that a rare subclonal common progenitor cell bearing EGFR p.Thr790Met might have existed in the primary tumor (Supplementary Note, Supplementary Fig. 8). Additional subclonal genetic co-alterations, including PIK3CA c.3140A>G; chr3:g.178952085A>G (hg19) (p.His1047Arg; NM_005628) (R5, right rib), RB1 c.2570G>A; chr13:g.49050868G>A (hg19) (p.Arg857His; NM_000321) (R4, right lung), CHD4 c.3452A>C; chr12:g.6697477T>G (hg19) (p.His1151Pro; NM_001273) (R6, right lung), and TLR4 c.866G>A; chr9:g.124075272G>A (hg19) (p.Arg289Gln; NM_138554) (R5, right rib) arose with tumor progression under rociletinib treatment (Fig. 5a–c). The activating PIK3CA variant p.Gly106Val21 was not found in all of the post-rociletinib metastatic sites (present in R4 and R6; absent in R5 and R7), thus demonstrating lesion-specific heterogeneity (Fig. 5a, b). The subclonal PIK3CA oncogenic variant p.His1047Arg22 was found in R5 (right rib, post-rociletinib), thus suggesting another instance of parallel evolution in this cancer: two different PIK3CA oncogenes (Fig. 5a–c and Supplementary Fig. 8). Although the RB1 variant p.Arg857His was detected in R4 (left lung, progression under rociletinib), and RB1 inactivation is associated with transition from lung adenocarcinoma to small-cell carcinoma after EGFR TKI resistance is acquired23,24, there was no evidence of transition to small-cell histology in this case, perhaps because of the absence of a somatic TP53 alteration (Supplementary Fig. 6).

Although plasma samples for cfDNA analysis were unavailable for the initial clinical events (before erlotinib treatment), skipping serially acquired cfDNA data (Online Methods) with tumor-biopsy-based WES showed examples of ubiquitous (EGFR p.exon19del and CTNNB1 p.Ser37Phe) and lesion-restricted (PIK3CA p.Gly106Val, RB1 p.Arg857His, and TLR4 p.Arg289Gln) mutations in the plasma (Fig. 5d). Thus, cfDNA analysis integrates multiple metastatic tumor lesions.

Figure 8 Evolution of the understanding of the genetic pathogenesis of oncogene-positive (EGFR-mutant) lung cancer. (a) Traditional view of lung cancer, based on histopathological analysis. Lung adenocarcinoma, scale bar, 50 μm. (b) Current NSCLC molecular classification based on single-gene driver-oncogene status, depicting the current view of mutually exclusive driver oncogenes, as shown in the pie chart, with frequency of each driver alteration in lung adenocarcinoma. (c) The proposed new model of EGFR-mutant NSCLC pathogenesis arising from our findings: a reclassification of advanced-stage EGFR-mutant NSCLC, based on the co-occurring genetic alterations indicated by our data set (shown here at the pathway level). We propose that advanced-stage EGFR-mutant NSCLCs contain co-occurring genetic alterations that function collaboratively as codrivers of tumor progression and drug resistance. We now need to identify and categorize these co-occurring functional genetic alterations beyond mutant EGFR itself in patients, early and dynamically during treatment, in order to improve patient survival. The finding of extensive co-occurring alterations within advanced-stage EGFR-mutant NSCLC at scale now paves the way for studying the biological and clinical effects of genetic interactions that are created by the co-alterations present in these EGFR-mutant NSCLCs.

Functional importance of co-occurring genomic alterations

Similarly to the findings for our larger cohort of patients with EGFR-mutant NSCLC (Fig. 1–4), our results highlight the co-occurrence of genetic alterations within the WNT (CTNNB1 variant p.Ser37Phe), PI3K (PIK3CA variant p.Gly106Val), and cell-cycle (CDK6 CNG, and CDKN2A loss) pathways. We hypothesized that such co-occurring alterations might function nonredundantly in driving tumor metastasis or might limit targeted therapy response, a hypothesis for which we provide experimental evidence (Supplementary Note, Supplementary Fig. 9 and Fig. 6).
Clonal analysis of genetic alterations detected in cfDNA

Our data (Fig. 5) suggested that clonal co-occurring oncogenic driver events can influence tumor progression and response to EGFR TKI treatment. To assess whether subclonal events are common in advanced-stage EGFR-mutant lung cancers more broadly, we assessed whether co-occurring genetic alterations detected in the cfDNA of 1,122 patients with advanced-stage EGFR-mutant NSCLC were clonal or subclonal (Supplementary Note and Supplementary Fig. 10). Our analysis indicated that the founder canonical EGFR mutations (encoding p.Leu858Arg and p.Glu746_Ala750del) were mostly clonal in the cohort of 1,122 patients with advanced-stage EGFR-mutant NSCLC, as expected (~86.5%, 332/384 and ~89.7%, 350/390, respectively; \( P = 0.19 \), Fisher’s exact test; 95% CI of difference in two population proportion [−1.3%–7.7%], OR = 0.96). We further found that advanced-stage EGFR-mutation-positive NSCLCs were more likely to contain subclonal genetic alterations than advanced-stage EGFR-mutation-negative NSCLCs (Fig. 7a and Supplementary Data Sets 1 and 2; 36.6% (1,156/3,157) subclonal alterations in EGFR-mutation-positive cases versus 24.9% (572/2,291) subclonal events in EGFR-mutation-negative cases; \( P = 2.2 \times 10^{-16} \), Fisher’s exact test (95% CI of difference in two population proportion [9.2%–14.1%], OR = 1.47)). Subclonal alterations were also more commonly found in the EGFR p.Thr790Met-positive samples (Fig. 7b and Supplementary Data Set 1; 39.7% (604/1,519) in EGFR p.Thr790Met-positive samples versus 33.3% (586/1,760) in EGFR p.Thr790Met-negative samples, \( P = 0.02 \), Fisher’s exact test (95% CI of difference in two population proportion [3.1%–9.7%], OR = 1.19)). Relative to the clonal founder EGFR mutation, EGFR p.Thr790Met was more frequently subclonal (~71.1% clonal; 313/440, compared with founder EGFR alterations p.Leu858Arg and p.Glu746_Ala750del ~95% clonal (as above), \( P = 2.2 \times 10^{-16} \), Fisher’s exact test, OR = 1.83), a finding of clinical relevance given that subclonal EGFR p.Thr790Met may be linked to inferior clinical response to third-generation EGFR TKI treatment. The subclonal frequency of other common variants is also described (Supplementary Note).

DISCUSSION

This study sheds new light on the genetic basis of oncogenesis and cancer progression by showing that multiple co-occurring oncogenic events are present in most advanced-stage EGFR-mutant lung cancers. These new data call into question the current view of the genetic basis of EGFR-mutant lung cancer as a single-oncogene disease wherein oncogenic-mutant EGFR is mutually exclusive from any other oncogene (Fig. 8). Our findings highlight the importance of deploying more informed and genomically empowered molecular diagnosis, monitoring, and dynamically applied rational polytherapy strategies to address the clonal and subclonal co-alterations that drive disease progression and drug resistance, in order to better control this deadly cancer. Our results are reminiscent of recent findings in myeloproliferative neoplasms26 and prompt reexamination of the presence and clinical effects of co-occurring genetic alterations in other cancer types by using large data sets, such as the one used herein, to enable a powered analysis.

We identified new pathways that promote EGFR-mutant lung cancer progression and limit EGFR TKI response. Examples include WNT/β-catenin and cell-cycle gene alterations and cell-cycle-gene mutations (Figs. 4–6 and Supplementary Note). Overall, the widespread presence, evolution, and clinical effects of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers uncovered here reshape the current view of oncogene-positive lung cancer and offer future directions for both basic and clinical research that hold promise for improving current treatments for this aggressive cancer.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

The authors acknowledge funding support from the NIH (R01CA169338, NIH Director’s New Innovator Award NCI-DP2CA174497), the Pew Charitable Trust, Stewart Foundation, and Searle Foundation (to T.G.B.), and the AACR and Lung Cancer Research Foundation (C.M.B.). The authors thank J. Blakey for artwork and A. Sabnis, R. Okimoto, A. Tulpule, and M. Hutchinson for critical reading of the manuscript. The authors acknowledge the following researchers for providing plasmids through Addgene: E. Campeau (University of Massachusetts Medical School); H. Land and J. Morgenstern (Imperial Cancer Research Fund); B. Weinberg (Whitehead Institute for Biomedical Research); J. Zhao (Dana-Farber Cancer Institute, Harvard Medical School); and E. Fearon (University of Michigan School of Medicine).

AUTHOR CONTRIBUTIONS

C.M.B., T.B.K.W., C.S., and T.G.B. designed the study. C.M.B. performed medical-record review, analyzed data and prepared tables and figures. T.B.K.W. performed WES and clonality analysis and prepared tables and figures with assistance from N.M., G.A.W., and N.I.B. W.W. performed analysis of cfDNA-sequencing data on patient cohorts and prepared tables and figures. B.G. performed cell-line experiments and prepared figures with assistance from A.M. J.C. and M.D. performed cancer personalized profiling by deep sequencing (CAPP-seq) analysis. V.R.O. and J.R. performed immunohistochemistry analysis. C.E.M., M.A.G., V.W., A.D.S., P.C.M., D.R.G., H.H., R.C.D., and J.W.R. performed medical-record review and provided clinical data. K.C.B. and R.B.L. compiled and annotated cfDNA data from 1,150 patients with EGFR-mutant-positive NSCLC and 1,008 patients with EGFR-mutant-negative NSCLC. A.R.C. extracted DNA and prepared exome libraries from patient tumor samples. A.E.C. and J.S.J. performed exome sequencing alignment and quality analysis. P.G. harvested autopsy tissue and performed pathological assessments. C.M.B. and T.G.B. wrote the manuscript, to which all authors contributed.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

1. de Bruin, E.C. et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 346, 251–256 (2014).
2. Zhang, J. et al. Intrapathogenic heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 346, 256–259 (2014).
3. Turke, A.B. et al. Preexistence and clonal selection of MET amplification in EGFR-mutant NSCLC. Cancer Cell 17, 77–88 (2010).
4. Zhang, Z. et al. Activation of the AKT kinase causes resistance to EGFR-targeted therapy in lung cancer. Nat. Genet. 44, 852–860 (2012).
5. Chabon, J.J. et al. Circulating tumour DNA profiling reveals heterogeneity of EGFR inhibitor resistance mechanisms in lung cancer patients. Nat. Commun. 7, 11815 (2016).
6. VanderLaan, P.A. et al. Mutations in TP53, PIK3CA, PTEN and other genes in EGFR-mutated lung cancers: correlation with clinical outcomes. Lung Cancer 106, 17–21 (2017).
7. Canale, M. et al. Impact of TP53 mutations on outcome in EGFR-mutated patients treated with first-line tyrosine kinase inhibitors. Clin. Cancer Res. 23, 2195–2202 (2017).
8. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 511, 543–550 (2014).
9. Lanman, R.B. et al. Analytical and clinical validation of a digital sequencing panel for quantitative, highly accurate evaluation of cell-free circulating tumor DNA. PLoS One 10, e0140712 (2015).
10. Thompson, J.C. et al. Detection of therapeutically targetable driver and resistance mutations in lung cancer patients by next-generation sequencing of cell-free circulating tumor DNA. *Clin. Cancer Res.* **22**, 5772–5782 (2016).

11. Yu, H.A. et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. *Clin. Cancer Res.* **19**, 2240–2247 (2013).

12. Yu, H.A. et al. Poor response to erlotinib in patients with tumors containing baseline EGFR T790M mutations found by routine clinical molecular testing. *Ann. Oncol.* **25**, 422–428 (2014).

13. Politi, K., Fan, P.D., Shen, R., Zakoenski, M. & Varmus, H. Erlotinib resistance in mouse models of epidermal growth factor receptor-induced lung adenocarcinoma. *Dis. Model. Mech.* **3**, 111–119 (2010).

14. Eberlein, C.A. et al. Acquired resistance to the mutant-selective EGFR inhibitor AZD9291 is associated with increased dependence on RAS signaling in preclinical models. *Cancer Res.* **75**, 2489–2500 (2015).

15. Thess, K.S. et al. Acquired EGFR C797S mutation mediates resistance to AZD9291 in non-small cell lung cancer harboring EGFR T790M. *Nat. Med.* **21**, 560–562 (2015).

16. Ou, S.I., Agarwal, N. & Ali, S.M. High MET amplification level as a resistance mechanism to osimertinib (AZD9291) in non-small cell lung cancer harboring EGFR T790M. *Lung Cancer* **98**, 59–61 (2016).

17. Ho, C.C. et al. Acquired BRAF V600E mutation as resistant mechanism after treatment with osimertinib. *J. Thorac. Oncol.* **12**, 567–572 (2017).

18. Sequist, L.V. et al. Rociletinib in EGFR-mutated non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 1700–1709 (2015).

19. Gamallo, C. et al. Beta-catenin expression pattern in stage I and II ovarian carcinomas relationship with beta-catenin gene mutations, clinicopathological features, and clinical outcome. *Am. J. Pathol.* **155**, 527–536 (1999).

20. Wang, X.Y., Repasky, E. & Liu, H.T. Antisense inhibition of protein kinase Co reverses the transformed phenotype in human lung carcinoma cells. *Exp. Cell Res.* **250**, 253–263 (1999).

21. Burke, J.E., Ponsic, O., Masson, G.R., Vadas, O. & Williams, R.L. Oncogenic mutations mimic and enhance dynamic events in the natural activation of phosphoinositide 3-kinase p110α (PIK3CA). *Proc. Natl. Acad. Sci. USA* **109**, 15259–15264 (2012).

22. Trejo, C.L. et al. Mutationally activated PIK3CA(H1047R) cooperates with BRAF(V600E) to promote lung cancer progression. *Cancer Res.* **73**, 6448–6461 (2013).

23. Dorantes-Heredia, R., Ruiz-Morales, J.M. & Cano-Garcia, F. Histopathological transformation to small-cell lung carcinoma in non-small cell lung carcinoma tumors. *Transl. Lung Cancer Res.* **5**, 401–412 (2016).

24. Sequist, L.V. et al. Genotypic and histological evolution of lung cancers acquiring resistance to EGFR inhibitors. *Sci. Transl. Med.* **3**, 75ra26 (2011).

25. Piotrowska, Z. et al. Heterogeneity underlies the emergence of EGFR T790 wild-type clones following treatment of T790M-positive cancers with a third-generation EGFR inhibitor. *Cancer Discov.* **5**, 713–722 (2015).

26. Otvos, H.M. et al. Effect of mutation order on myeloproliferative neoplasms. *N. Engl. J. Med.* **372**, 601–612 (2015).
**ONLINE METHODS**

**Patients.** IRB approval for the study (no. 16-19636) was granted by the UCSF IRB on May 13, 2016. Per the UCSF IRB, the study did not involve human subjects, as defined by the federal regulations summarized in 45 CFR 46.102(f). Hence, further IRB oversight was not required, and the requirement for informed consent was waived. For EGFR-mutation-positive and EGFR-mutation-negative cohorts, selection criteria for inclusion were met if patients had a known diagnosis of stage III or stage IV NSCLC. For EGFR-mutation-positive NSCLC, the analysis included 1,122 consecutive samples collected in March 2015 to April 2016 from 1,006 patients in whom a nonsynonymous mutation in EGFR of known or predicted functional significance (described below) was identified in the Guardant 360 clinical assay. EGFR CNG by itself was included in the EGFR-mutation-negative cohort. The EGFR-mutant-negative cohort consisted of all other patients with advanced-stage NSCLC from the time period of January 2016 to April 2016 (1,008 samples from 999 patients). Chart review of EGFR-mutant Guardant cases from patients at UCSF, UCSD, UC Davis, and the University of Colorado was carried out by the study investigators to identify patient demographic information and to determine when the Guardant 360 assay was sent in relation to the patients’ treatment course. Objective response, PFS, and overall survival to EGFR TKI therapy were determined by retrospective chart review for clinical assessment and direct radiographic review by study investigators when possible. Composite clinical evaluation that integrated clinical and radiographic information was used to discriminate responders from nonresponders. (Responders were defined as showing radiographic stable disease or progressive disease (by RECIST 1.1 criteria, or clinical decline, or death before imaging). The time-to-event outcomes, including PFS and overall survival were estimated with the Kaplan–Meier method.

**Cell-free DNA analysis.** Samples were shipped to a Clinical Laboratory Improvement Act (CLIA)-certified, College of American Pathologists–accredited laboratory (Guardant Health). cfDNA was extracted from whole blood collected in 10-mL Streck tubes. After double centrifugation, 5–30 ng of cfDNA was isolated for digital sequencing as previously described. For EGFR-mutant-positive NSCLCs, samples were run on a 68-gene panel or 70-gene panel, but only the 68 genes in common were included in this analysis (Supplementary Table 2). Sequencing data were analyzed with the Guardant Health clinical analysis bioinformatics pipeline to identify single-nucleotide variants (SNVs) in 68 genes (150-kb panel footprint), CNGs in 16 genes, indels in EGFR, and fusions in ALK, RET, ROS1, NTRK3, FGFR2, and FGFR3 (refs. 9, 27). All cell-free leukocyte- and tumor-derived DNA fragments were simultaneously sequenced. The variant allele fraction (VAF) was calculated as the proportion of cfDNA bearing the variant in a background of wild-type cfDNA. Reporting thresholds for SNVs, indels, and fusions were one or two molecules and 0.01–0.04% allelic fraction with 0.2–0.3% 95% limits of detection and >99.9999% per-position analytical specificity. To identify CNGs in a large training set, probe-level unique-molecule coverage was normalized for signal saturation, individual-probe efficiency, GC content, and overall unique-molecule throughput, and was robustly summarized to generate a quantitative gene-level unique-molecule representation. Relative quantification was determined by comparison of this representation to the inferred diploid baseline on a per-gene level. Reporting thresholds were based on training-set-established decision thresholds for both absolute copy number deviation from the per-sample diploid baseline and deviation from the baseline variation of probe-level-normalized signal in the context of background variation within each sample’s own diploid baseline. The reporting threshold for CNGs was 2.12 copies with a 2.24–2.76 gene-specific 95% limits of detection and 100% analytical specificity. For clonality analysis, first, the mutational allele frequency (MAF) was normalized by the copy numbers of the same genes with CNGs, and the largest MAF within each was selected as the normalized maximum MAF; second, the ratio of the MAF of each mutational allele over the maximum percentage detection within a case was computed, and the probability distribution was plotted by using kernel density estimation. To determine the cutoff of the percentage of normalized MAF as clonal or subclonal of each mutation, we implemented the aforementioned algorithms to the case with both tumor tissue exome sequencing and cfDNA sequencing encoding EGFR exon19del, EGFR p.Thr790Met, PIK3CA p.Gly106Val, CTNNB1 p.Ser37Phe, RB1 p.Arg857His, and TLR4 p.Arg289Gln reported in this study (Fig. 5 and Supplementary Fig. 10). The value of 0.2 was defined as a robust cutoff for subclonal or clonal mutations, thus resulting in 100% sensitivity and 100% specificity, because all somatic variants identified through this method as subclonal or clonal in cfDNA were also correctly identified as subclonal or clonal in patient tumor samples (Supplementary Fig. 10). For the longitudinal case (Fig. 5d), cfDNA was isolated from 1 ml of frozen plasma and analyzed as previously described. Clinical data were collected by review of medical records under an IRB-approved protocol (UCSF). Nonsynonymous mutations from EGFR-mutant-positive and EGFR-mutant-negative data sets were further processed with the R statistical computing program (version 3.3). Unknown significant variants were filtered out by using COSMIC (V79), GENIE, and prediction algorithms (URLs).

**Whole-exome sequencing and analysis.** Informed consent was obtained from the patient and patient’s family for study of biological materials and clinical records obtained from the patient. DNA was extracted from formalin-fixed paraffin-embedded (FFPE) primary and frozen tumor tissue samples and matched nontumor tissue with a Qiagen Allprep DNA/RNA Mini Kit. The library preparation protocol was based on the Agilent SureSelect Library Prep and Capture System (Agilent Technologies). Quantification and quality were assessed with a Qubit Fluorometer (Thermo Fisher). DNA concentration was determined to be greater than 2.5 ng/µL, and the overall quantity was >500 ng. Analysis with a NanoDrop spectrophotometer indicated that the 260/280 ratio was >1.7. DNA was resuspended in a low-TE buffer and sheared (duty cycle, 5%; intensity, 175; cycles/burst, 200; time, 300 s) with a Covaris S2 Ultrasonicator Barcoded exome libraries were prepared with an Agilent Sure Select V5 library kit, per the manufacturer’s specifications. The libraries were run on the HiSeq2500 platform.

**Alignment.** Raw paired-end reads (100 bp) in FastQ format generated by the Illumina pipeline were aligned to the full hg19 genomic assembly obtained from UCSC, gencode 14, with bwa version 0.7.12. Picard tools version 1.117 was used to sort, remove duplicate reads, and generate QC statistics. Tumor DNA was sequenced to a median depth of 303× (range 114.39–383.41), and the matched germline DNA was sequenced to an average depth of 231.65.

**Exome analysis.** SNV, indel, and dinculeotide-substitution calling; identification and classification of driver mutations; somatic copy number–aberration calling; subclonal deconstruction; and phylogenetic-tree construction were performed as previously described.

**Classification of somatic copy number alterations (SCNAs).** SCNAs were defined as segments called by ASCAT that were ≥400 kb in size and met set thresholds. Segments with a combined raw nMinor and nMajor ≥1.5 times the ASCAT-derived ploidy for their specific tumor region were considered SCNAs. SCNA losses had an integer nMinor value of 0 and a combined raw nMinor and nMajor of less than 1.25 times ploidy for their specific tumor region.

**Incorporation of the EGFR p.Thr790Met mutation into phylogenetic reconstruction.** To create an accurate subclonal phylogeny, it is necessary to remove mutation clusters that violate two evolutionary principles. First is the ‘pigeonhole principle’, wherein two mutation clusters cannot be considered to be on separate branches of an evolutionary tree and thus to be independent if the cancer cell fraction values of the two clusters together exceed 100% within region of a tumor. Second is the ‘crossing rule’, wherein a descendant clone is required to have a smaller cancer cell fraction than its ancestor within each and every tumor region. With these principles, it can be determined whether particular mutation clusters conflict with each other and cannot be fitted to the same evolutionary tree.

The subclonal phylogeny illustrating the entire course of the patient’s disease was derived by following these two principles and the methods of multisample subclonal deconstruction and tree construction in ref. 29. However, the SNV encoding EGFR p.Thr790Met did not cluster with any other SNVs through these methods, owing to its unique cancer-cell-fraction profile across
R3, R4, R5, R6, and R7. No other SNV appeared to be clonal in all these regions and absent from both R1 and R2. Because cluster 7 and the EGFR p.Thr790Met mutation appeared to be clonal in R3, R4, and R6, but cluster 7 was absent from R5 and R7, and EGFR p.Thr790Met was present, on the basis of the crossing rule, they could not be present in the same population of cells. In addition, because cluster 7 was present only in R2 before erlotinib treatment, but EGFR p.Thr790Met was absent from R2, it follows that cluster 7 is likely to have arisen before the EGFR p.Thr790Met SNV.

The most parsimonious solution to this violation of the crossing rule, assuming that the cancer cell fractions were correct, is that there are two independent origins of EGFR p.Thr790Met. EGFR p.Thr790Met SNV (A) would occur in a cell already containing the SNVs from cluster 7 and would go on to become clonal after erlotinib treatment in R3, R4, and R6. EGFR p.Thr790Met SNV (B) would occur in a population of cells lacking the SNVs present in cluster 7 and would go on to become clonal in R5 and R7 after erlotinib treatment. These possible origins of the EGFR p.Thr790Met are indicated on the subclonal phylogeny in Supplementary Figure 8 by the placement of a magenta square on the relevant branches.

**Cell lines and reagents.** The HCC827 (EGFR p.Glu746_Ala750del) and HEK293-FT cell lines were obtained, authenticated, and cultured as recombinant with the American Type Culture Collection (ATCC). These cell lines were confirmed to be negative for mycoplasma. HCC827 cells were cultured in RPMI 1640 medium (HyClone, GE Healthcare) supplemented with 10% FBS (SAFC, Sigma-Aldrich) and 1× penicillin and streptomycin (UCSF, Cell Line Technologies) in RPMI 1640 medium (HyClone, GE Healthcare) supplemented with 10% FBS (SAFC, Sigma-Aldrich) and 1× penicillin and streptomycin (UCSF, Cell Culture Facility). HEK293-FT cells were cultured in DMEM (HyClone, GE Healthcare) supplemented with 10% FBS, 0.1× penicillin and streptomycin. All cell lines were grown at 37 °C in a humidified atmosphere with 5% CO2. Erlotinib and rociletinib were purchased from Selleckchem.

Mammalian expression vector pQCXIB empty (w335-1) was a gift from E. Campeau (Addgene no. 17487)30, pBABE-puro was a gift from H. Land, J. Morgenstern, and B. Weinberg (Addgene plasmid no. 1764)31; pCMV-SV5-G (Addgene plasmid no. 8454) and pUMVC (Addgene plasmid no. 8449)32 were gifts from B. Weinberg; pBabe puro HA PIK3CA was a gift from J. Zhao (Addgene plasmid no. 12523)33, human β-Catenin pcDNA3 was a gift from E. Fearon (Addgene plasmid no. 16828)34. The PIK3CA and β-Catenin constructs were engineered to express PIK3CA p.Gly106Val and β-Catenin p.Ser37Phe with the QuikChange II XL Site-Directed Mutagenesis Kit protocol (Agilent Technologies). p.Ser37Phe encoding the CTNNB1 fragment was then cloned in a pQCXIB retroviral construct for stable overexpression, through sticky-end ligation with Apal and BamHI-TF (New England BioLabs) restriction enzymes, per the manufacturer’s instructions. HEK293-FT cells were transfected with pBABE (empty vector), pBabe-PIK3CA encoding p.Gly106Val, pQCXIB (empty vector) and pQCXIB-CTNNB1 encoding p.Ser37Phe with Fugu 6 reagent (Promega), per the manufacturer’s instructions. Virus-containing medium was harvested 24 h and 48 h after transfection. HCC827 cells were infected for 24 h with virus-containing medium supplemented with polybrene (8 μg/mL; Sigma-Aldrich). The culture medium was changed to standard growth medium for an overnight incubation, after which cells were incubated in antibiotic selection medium containing puromycin (1 μg/mL; Gibco) for the p-Babe construct or blasticidin (2.5 μg/mL; Gibco) for pQCXIB constructs. Antibiotic-resistant cells were used in the subsequent tests.

**Cell viability and growth assays.** We seeded 100,000 HCC827 cells engineered with the β-Catenin p.Ser37Phe and PIK3CA p.Gly106Val expression constructs, under puromycin (1 μg/mL) and blasticidin (2.5 μg/mL) selection, in 12-well plates. After 24 h, the cells were treated with DMISO (control), erlotinib (50–100 nM), or blasticidin (100 nM) in 2% FBS for 3 d. Cells were then air-dried for 5 min, fixed for 5 min in paraformaldehyde (PFA, 4% (vol/vol); Santa Cruz Biotechnology) and stained in 0.05% (wt/vol) crystal violet (Sigma-Aldrich) solution for 30 min. Each well was washed twice with tap water and air-dried. Plates were scanned with an ImageQuant LAS4000 instrument (GE Healthcare Life Sciences). Each image is representative of a triplicate experiment. Cell viability was assessed by using the above culture conditions, with 200 cells seeded per well. Cell count was registered after 3 d of growth and assessed with a Vi-CELL XR cell counter. Each test was run in triplicate.

**Invasion and migration assays.** Transwell migration and invasion assays were performed as previously described15. Briefly, 8-μm-pore Matrigel-coated (invasion) or noncoated (migration) Transwell inserts (BD Biosciences) were added at the top of a Transwell chamber filled with RPMI medium with 10% FBS. To each insert, 2.4 × 10^5 cells in serum-free medium were added. The Transwell chambers were incubated for 20 h at 37 °C. Cells that did not migrate through the pore or invade the Matrigel were scraped off; the membranes were fixed in methanol for 15 min and then stained with crystal violet for 30 min. The surface of the membrane was imaged in five distinct fields with a Zeiss AxioPlan II immunofluorescence microscope at 10× magnification. Invasion and migration were assessed by counting the average imaged cells in the five regions. Results presented are from three independent experiments.

**Immunoblotting and qRT–PCR.** The HCC827 cells engineered with the β-Catenin p.Ser37Phe and PIK3CA p.Gly106Val expression constructs were drug treated, in serum free conditions, with DMISO (control), erlotinib (100 nM), or rociletinib (100 nM) for 24 h. Protein lysates were collected in RIPA buffer supplemented with protease inhibitors (Roche) and phosphatase inhibitors (Roche). Western blotting was performed by loading 10 μg of lysed proteins. Precast 4–15% acrylamide gels (Bio-Rad) were used for the monodimensional protein separation. Proteins were transferred on nitrocellulose membranes with a Trans-blot Turbo Transfer system (Bio-Rad). Blots were then blocked in Tris-buffered saline, 0.1% Tween20 (vol/vol) and 5% (vol/vol) BSA (Fisher Scientific) for 1 h at room temperature. The following primary antibodies were incubated overnight at 4 °C: anti-pY1068-EGFR D7A5 (no. 3777), anti–total EGFR D38B1 (no. 4267), anti-β-Catenin D10A8 (no. 8480), anti-pS473-AKT D9E (no. 4060), anti–total AKT (no. 9272), anti-pT202/Y204-ERK1/2 (no. 9101), anti–total ERK1/2 (no. 9102), and anti–cleaved PARP (no. 9541) from Cell Signaling Technology; and anti-actin AC-74 (no. A2228) from Sigma-Aldrich. The membranes were washed twice in wash buffer (Tris-buffered saline, 0.1% (vol/vol) Tween20) and then incubated with secondary HRP-conjugated antibodies (Cell Signaling Technology, anti-rabbit IgG, no. 7074, anti-mouse IgG, no. 7076) for 1 h, at room temperature. An ECL kit (GE Healthcare) was used as the chemiluminescent substrate. Blots were developed and scanned with ImageQuant LAS4000 (GE Healthcare Life Sciences). ImageJ (NIH) was used to quantify the western blots. All western blots represent the result of three independent experiments.

The RNA was purified from the HCC827 cells engineered with the β-Catenin p.Ser37Phe and PIK3CA p.Gly106Val expression constructs with an RNeasy Micro Kit (Qiagen). One microgram of total RNA was used for the reverse transcriptase reaction with a SensiFAST cDNA Synthesis Kit (BIOLINE). The qPCR was performed with six replicates per condition, with a 1:3 dilution of the template cDNA, Human MYC, CCND1, LEF1, HOXB9, and endogenous-control GAPDH genes were amplified with TaqMan gene expression assays (Applied Biosystems). Gene expression analysis was computed with QuantStudio 12K Flex Software (Applied Biosystems). Data were analyzed with the 2^(-ΔΔCt) method and are expressed as relative mRNA levels.

**Immunohistochemistry.** Immunohistochemistry was performed as previously described16. Briefly, 5-μm-thick FFPE human tissue sections were stained with antibodies to β-Catenin D10A8 (no. 8480, Cell Signaling, 1:100 dilution) or pSer473-Akt D9E (no. 4060, Cell Signaling, 1:100 dilution), per the manufacturer’s instructions. Stained slides were digitized with an Aperio ScanScope CS Slide Scanner (Aperio Technologies) with a 20× objective. The proportion of cells exhibiting nuclear β-Catenin staining was determined by using the ScanScope default nuclear algorithm. pSer473-Akt quantification was determined by using the ScanScope default membrane algorithm. Three fields of view per section were used to determine the mean and s.e.m. of positively staining cells.

**Statistics.** To determine differences in cDNA alterations between cohorts, we used two-tailed Fisher’s exact tests with Benjamini–Hochberg correction for multiple hypothesis testing to generate q values (Figs. 1c, e, 2f, 3c, d and 4e and Supplementary Figs. 3 and 9). We considered the false discovery rate to be controlled under 20% (q ≤ 0.2)77. For Supplementary Tables 3 and 4, two-tailed t tests were used for two-population mean difference with 95% confidence intervals. The effect size, Cohen’s d was determined with the equation

\[
\text{d} = \frac{\bar{X}_1 - \bar{X}_2}{S_p}
\]

where \(\bar{X}_1\) and \(\bar{X}_2\) are the means of the two populations and \(S_p\) is the pooled standard deviation.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
\[(\text{mean}_1 - \text{mean}_2)/\text{s.d.}_\text{pooled}, \text{ where } \text{s.d.}_\text{pooled} = (\text{s.d.}_1 + \text{s.d.}_2)/2; \text{ proportions tests were used for two-population proportion comparison with 95% confidence intervals, and no correction was used. In some conditions, the 95% confidence interval for the single-population proportion was used. OR values were calculated as a measure of the effect size between two populations with proportion comparison. For assessments of PFS (Fig. 4 and Supplementary Fig. 3) and overall survival (Supplementary Fig. 4), the 95% CI for the median duration of PFS and overall survival were computed with the robust non-parametric Brookmeyer and Crowley method. HRs with 95% CI and P values were calculated with the ‘Cox proportional-hazards regression model with survival’ package in R. For qPCR, cell growth, invasion, and migration analysis (Fig. 6) one-way ANOVA with Bonferroni correction were used to determine P values (GraphPad Prism).

**Code availability.** Most bioinformatics tools used in the analysis of this data set are publicly available; any that are not are available on request.

**Data availability.** The data supporting the findings of the study are available within the paper and its supplementary information files and have been deposited publically in the European Genome-phenome Archive (EGA) under accession number EGAS00001002604. A Life Sciences Reporting Summary for this paper is available.
Nature Research wishes to improve the reproducibility of the work we publish. This form is published with all life science papers and is intended to promote consistency and transparency in reporting. All life sciences submissions use this form; while some list items might not apply to an individual manuscript, all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

### Experimental design

1. **Sample size**
   
   Describe how sample size was determined.
   
   This is a retrospective analysis of 1122 consecutive EGFR-mutant positive patients collected from March 2015 to April 2016 and from 1008 EGFR-mutant negative patients collected from January 2016 to April 2016. This sample size provided sufficient power to detect statistically significant differences between these two cohorts (page 4).

2. **Data exclusions**
   
   Describe any data exclusions.
   
   Samples with only synonymous mutations or mutations of unknown functional significance were excluded from the analysis in order to identify alterations that were most likely to have a functional impact on tumor biology.

3. **Replication**
   
   Describe whether the experimental findings were reliably reproduced.
   
   All functional experiments were performed in replicate and results were reliably reproduced. The results presented in figure 5 serve as validation of the results presented in figures 1-4.

4. **Randomization**
   
   Describe how samples/organisms/participants were allocated into experimental groups.
   
   This is a retrospective study cohort study in which genomic alterations detectable in cfDNA of patients with 1. EGFR-mutations were compared to patients without EGFR-mutations, 2. EGFR-T790M-positive compared to EGFR-T790M negative, and EGFR TKI responders compared to non-responders were compared. Due to the retrospective aspect of this analysis, randomization was not possible. To account for possible confounding variables between the groups, we controlled for patient demographic variables (e.g. age, gender, and smoking status).

5. **Blinding**
   
   Describe whether the investigators were blinded to group allocation during data collection and/or analysis.
   
   Bioinformatics analysis was performed in a manner blinded to patient demographic or clinical outcomes data. Retrospective clinical chart review was performed in a manner blinded to patient cfDNA genomic data.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.
6. Statistical parameters
For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or the Methods section if additional space is needed).

| n/a | Confirmed |
|-----|-----------|
| ☐   | ☒         |

- The exact sample size \((n)\) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly.
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. \(p\) values) given as exact values whenever possible and with confidence intervals noted
- A summary of the descriptive statistics, including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

7. Software
Policy information about availability of computer code

Describe the software used to analyze the data in this study. Open source R program was used for data visualization and statistic analysis as described in Methods. Whole exome sequencing data was performed according to the Tracer-X pipeline published in Jamal-Hanjani, M. et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med 376, 2109-2121 (2017).

For all studies, we encourage code deposition in a community repository (e.g. GitHub). Authors must make computer code available to editors and reviewers upon request. The Nature Methods guidance for providing algorithms and software for publication may be useful for any submission.

8. Materials and reagents
Policy information about availability of materials

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions on availability of unique materials.
9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

The primary antibodies used for the western blots were: pY1068-EGFR D7A5 (#3777) (1), total EGFR D38B1 (#4267) (1), β-Catenin D10A8 (#8480) (2), pS473-AKT D9E (#4060) (1), total AKT (#9272) (1), pT202/Y204-ERK1/2 (#9101) (3), total ERK1/2 (#9102) (4) and cleaved PARP (#9541) (5) from Cell Signaling Technology; Actin AC-74 (#A2228) from Sigma-Aldrich (1).

The secondary HRP-conjugated antibodies used for the western blots were: HRP-conjugated anti-rabbit (#7074) or HRP-conjugated anti-mouse IgG (Cell Signaling Technology) (6).

All of the antibodies were validated in Antibodypedia. References are provided below. The citations, catalog numbers, clone numbers and references to the antibody validation profile validated the antibodies used for the system under study (western blot assay, human antigens).

1. Blakely CM, Pazarentzos E, Olivas V, et al. NF-kappaB-activating complex engaged in response to EGFR oncogene inhibition drives tumor cell survival and residual disease in lung cancer. Cell Rep 2015;11:98-110.
2. Ando F, Sohara E, Morimoto T, et al. Wnt5a induces renal AQP2 expression by activating calcineurin signalling pathway. Nat Commun 2016;7:13636.
3. Kovacs J, Poor P, Kaschani F, et al. Proteasome Activity Profiling Uncovers Alteration of Catalytic beta2 and beta5 Subunits of the Stress-Induced Proteasome during Salinity Stress in Tomato Roots. Front Plant Sci 2017;8:107.
4. Yang X, Lou Y, Liu G, et al. Microglia P2Y6 receptor is related to Parkinson’s disease through neuroinflammatory process. J Neuroinflammation 2017;14:38.
5. Yang HJ, Ju F, Guo XX, et al. RNA-binding protein RBM3 prevents NO-induced apoptosis in human neuroblastoma cells by modulating p38 signaling and miR-143. Sci Rep 2017;7:41738.
6. Yang Y, Jiang Z, Bolnick A, Dai J, Puscheck EE, Rappolee DA. Departure from optimal O2 level for mouse trophoblast stem cell proliferation and potency leads to most rapid AMPK activation. J Reprod Dev 2017;63:87-94.

For IHC:

β-Catenin D10A8 (#8480 Cell Signaling). Manufacturer datasheet describes cross-reactivity for human samples and incudes immunohistochemistry among applications. References provided, including Yueh et al (Colon Cancer Tumorigenesis Initiated by the H1047R Mutant PI3K. PLoS ONE. 2016. 11(2): e0148730), who demonstrate use of this antibody for immunohistochemistry on human tumor samples.

pSer473-Akt D9E (#4060, Cell Signaling). Manufacturer datasheet describes cross-reactivity for human samples and incudes immunohistochemistry among applications. References provided, including Manchado et al (A combinatorial strategy for treating KRAS mutant lung cancer. Nature. 2016. 534(7609): 647-651), who demonstrate use of this antibody for immunohistochemistry on patient-derived xenograft samples.

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

b. Describe the method of cell line authentication used.

c. Report whether the cell lines were tested for mycoplasma contamination.

d. If any of the cell lines used in the paper are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.
Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals
   Provide details on animals and/or animal-derived materials used in the study.

No animals were used.

Policy information about studies involving human research participants

12. Description of human research participants
   Describe the covariate-relevant population characteristics of the human research participants.

This information is provided in Supplementary Tables 1, 5, and 6.
To summarize:
EGFR-mutation positive: Total Samples 1122 (1006 patients)
EGFR-mutation negative: Total samples 1008 (999 patients)
Date range March 2015 - April 2016 Jan 2016 April 2016
EGFR-mutant positive male: 346 (34.4%)
EGFR-mutant negative male: 457 (45.7%)
EGFR-mutant positive female: 660 (65.6%)
EGFR-mutant negative female: 542 (54.3%)
Mean Age EGFR-mutant positive: 64.6
Mean Age EGFR-mutant negative: 67.4
EGFR-mutant positive Adenocarcinoma: 747 (74.3%)
EGFR-mutant negative Adenocarcinoma: 558 (55.9%)
EGFR-mutant positive NSCLC-NOS: 259 (25.7%)
EGFR-mutant negative NSCLC-NOS: 441 (44.1%)
EGFR-mutant positive Stage III/IV: 1006
EGFR-mutant negative Stage III/IV: 999