The NEW ENGLAND JOURNAL of MEDICINE

ESTABLISHED IN 1812 JUNE 1, 2017 VOL. 376 NO. 22

Tracking the Evolution of Non–Small-Cell Lung Cancer

M. Jamal-Hanjani, G.A. Wilson, N. McGranahan, N.J. Birkbak, T.B.K. Watkins, S. Veeriah, S. Shafi, D.H. Johnson, R. Mitter, R. Rosenthal, M. Salm, S. Horswell, M. Escudero, N. Matthews, A. Rowan, T. Chambers, D.A. Moore, S. Turajlic, H. Xu, S.-M. Lee, M.D. Forster, T. Ahmad, C.T. Hiley, C. Abbosh, M. Falzon, E. Borg, T. Marafioti, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S.M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, R. Shah, L. Joseph, A.M. Quinn, P.A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D.A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J.F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, S. Dentro, P. Taniere, B. O'Sullivan, H.L. Lowe, J.A. Hartley, N. Iles, H. Bell, Y. Ngai, J.A. Shaw, J. Herrero, Z. Szallasi, R.F. Schwarz, A. Stewart, S.A. Quezada, J. Le Quesne, P. Van Loo, C. Dive, A. Hackshaw, and C. Swanton, for the TRACERx Consortium*

ABSTRACT

BACKGROUND

Among patients with non–small-cell lung cancer (NSCLC), data on intratumor heterogeneity and cancer genome evolution have been limited to small retrospective cohorts. We wanted to prospectively investigate intratumor heterogeneity in relation to clinical outcome and to determine the clonal nature of driver events and evolutionary processes in early-stage NSCLC.

METHODS

In this prospective cohort study, we performed multiregion whole-exome sequencing on 100 early-stage NSCLC tumors that had been resected before systemic therapy. We sequenced and analyzed 327 tumor regions to define evolutionary histories, obtain a census of clonal and subclonal events, and assess the relationship between intratumor heterogeneity and recurrence-free survival.

RESULTS

We observed widespread intratumor heterogeneity for both somatic copy-number alterations and mutations. Driver mutations in *EGFR*, *MET*, *BRAF*, and *TP53* were almost always clonal. However, heterogeneous driver alterations that occurred later in evolution were found in more than 75% of the tumors and were common in *PIK3CA* and *NF1* and in genes that are involved in chromatin modification and DNA damage response and repair. Genome doubling and ongoing dynamic chromosomal instability were associated with intratumor heterogeneity and resulted in parallel evolution of driver somatic copy-number alterations, including amplifications in *CDK4*, *FOXA1*, and *BCL11A*. Elevated copy-number heterogeneity was associated with an increased risk of recurrence or death (hazard ratio, 4.9; $P=4.4\times10^{-4}$), which remained significant in multivariate analysis.

CONCLUSIONS

Intratumor heterogeneity mediated through chromosome instability was associated with an increased risk of recurrence or death, a finding that supports the potential value of chromosome instability as a prognostic predictor. (Funded by Cancer Research UK and others; TRACERx ClinicalTrials.gov number, NCT01888601.)

The authors' full names, academic degrees, and affiliations are listed in the Appendix. Address reprint requests to Dr. Swanton at the Translational Cancer Therapeutics Laboratory, Francis Crick Institute, 3rd Fl. SW, 1 Midland Rd., London NW1 1AT, United Kingdom, or at charles.swanton@crick.ac.uk.

*A complete list of investigators in the Tracking Non–Small-Cell Lung Cancer Evolution through Therapy (TRACERx) Consortium is provided in Supplementary Appendix 1, available at NEJM.org.

Drs. Jamal-Hanjani, Wilson, McGranahan, Birkbak, and Veeriah and Mr. Watkins contributed equally to this article.

This article was published on April 26, 2017, at NEJM.org.

N Engl J Med 2017;376:2109-21.
DOI: 10.1056/NEJMoa1616288

Lung cancer is the leading cause of cancer-related death worldwide,[1,2] with non–small-cell lung cancer (NSCLC) being the most common type. Large-scale sequencing studies have revealed the complex genomic landscape of NSCLC[3-6] and genomic differences between lung adenocarcinomas and lung squamous-cell carcinomas.[7] However, in-depth exploration of NSCLC intratumor heterogeneity (which provides the fuel for tumor evolution and drug resistance) and cancer genome evolution has been limited to small retrospective cohorts.[8,9] Therefore, the clinical significance of intratumor heterogeneity and the potential for clonality of driver events to guide therapeutic strategies have not yet been defined.

Tracking Non–Small-Cell Lung Cancer Evolution through Therapy (TRACERx)[10] is a multicenter, prospective cohort study, which began recruitment in April 2014 with funding from Cancer Research UK. The target enrollment is 842 patients from whom samples will be obtained for high-depth, multiregion whole-exome sequencing of surgically resected NSCLC tumors in stages IA through IIIA. One primary objective of TRACERx is to investigate the hypothesis that intratumor heterogeneity — in terms of mutations (single or dinucleotide base substitutions or small insertions and deletions) or somatic copy-number alterations (reflecting gains or losses of chromosome segments) — is associated with clinical outcome. Here, we report on the first 100 patients who were prospectively recruited in the study.

Figure 1 (facing page). Overview of the Demographic and Clinical Characteristics of the Patients in the TRACERx Study.

Panel A shows the demographic and clinical characteristics of the 100 patients in the study, including diagnosis, tumor stage, and smoking status. Panel B shows how multiregion sequencing was performed on surgically resected tumors to analyze somatic mutations and copy-number alterations, which facilitated the assessment of intratumor heterogeneity and phylogenetic reconstruction. Stars on the schematic chromosomes indicate mutations, where yellow represents clonal pregenome doubling mutations, pink represents clonal postgenome doubling mutations, and red represents subclonal mutations. Panel C shows the key clinical questions that were addressed in the study.

METHODS

PATIENTS AND TUMOR SAMPLES

We collected tumor samples from 100 patients with NSCLC who had not received previous systemic therapy (Fig. 1A; and Fig. S1 in Supplementary Appendix 1, available with the full text of this article at NEJM.org). Identifiers of patients were reassigned to protect anonymity and were ordered according to intratumor heterogeneity and histologic subtype. Eligible patients were at least 18 years of age and had received a diagnosis of NSCLC in stages IA through IIIA (except Patient CRUK0035, whose tumor was classified as stage IIIB on the basis of postoperative histologic analysis). The cohort was representative of a population of patients with NSCLC who were eligible for curative resection. Histologic data were confirmed on central review by a lung pathologist. (Details regarding the study design are provided in the protocol, available at NEJM.org.)

To assess intratumor heterogeneity, samples of at least two tumor regions that were separated by a margin of 0.3 cm to 1 cm (depending on the size of the tumor) had to be available for study. None of the tumors carried a translocation in *ALK*, *ROS1*, or *RET* on the basis of sequencing. This finding was confirmed for *ALK* and *ROS1* with the use of immunohistochemical testing. All the patients provided written informed consent. The clinical characteristics of the patients and the study criteria are provided in Tables S1 and S2 and in the Experimental Procedures section in Supplementary Appendix 1.

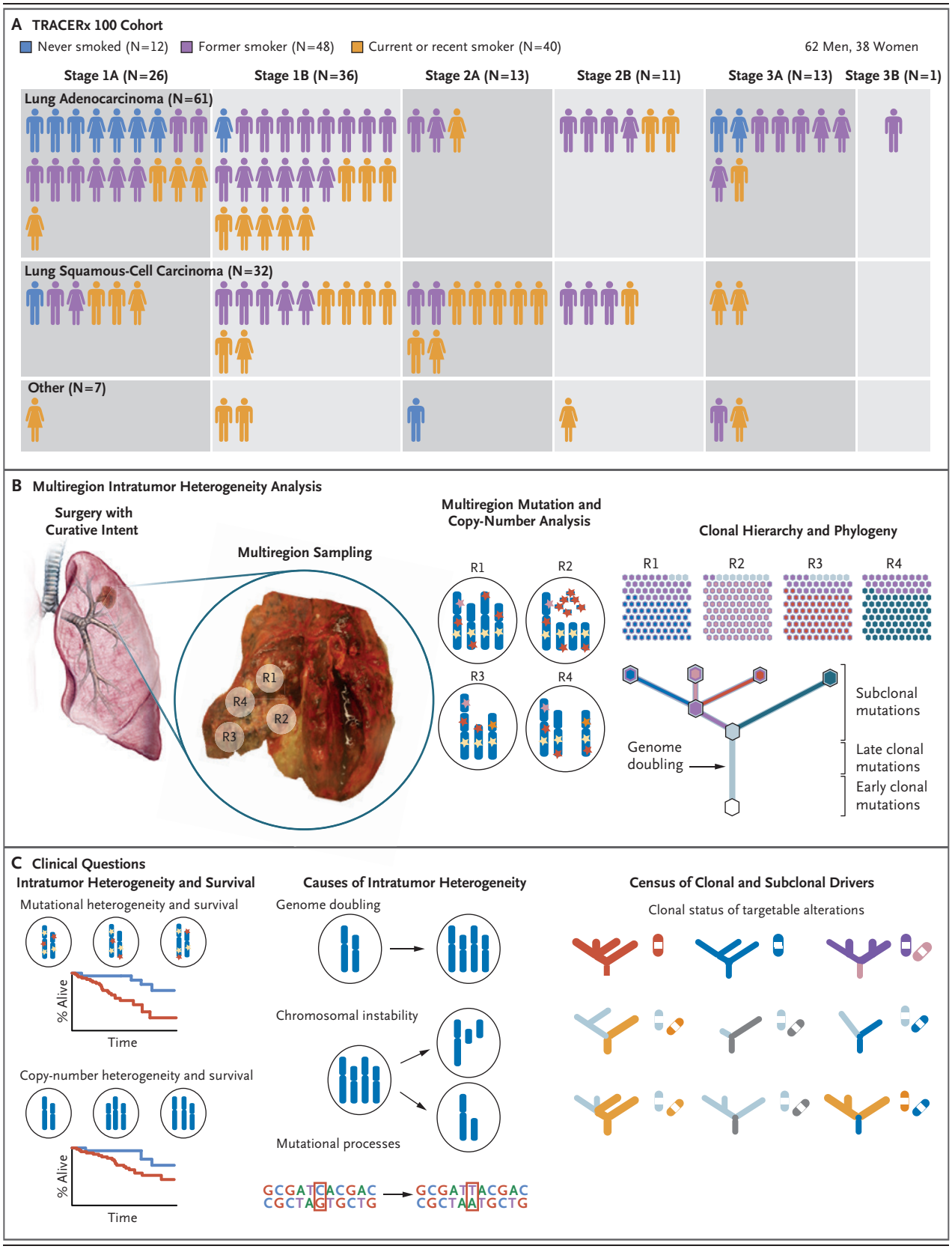
MULTIREGION WHOLE-EXOME SEQUENCING

We used the Illumina HiSeq to perform whole-exome sequencing on multiple regions collected from each tumor. We sequenced 327 tumor regions (323 primary tumor regions and 4 lymph-node metastases) and 100 matched germline samples derived from whole blood (median number, 3 regions per tumor; range, 2 to 8), to a median depth of 426× (Table S3 in Supplementary Appendix 1). Orthogonal validation was performed (Table S4 and Fig. S2 in Supplementary Appendix 1). All sequencing data have been deposited in the European Genome–Phenome Archive under accession number EGAS00001002247.

RESULTS

INTRATUMOR HETEROGENEITY IN NSCLC

Genetic diversity within tumors can act as a substrate for natural selection and tumor evolu-



tion. We performed multiregion whole-exome sequencing on 100 TRACERx tumors and classified somatic mutations, which were defined as coding and noncoding single-nucleotide variants, and copy-number alterations, which were measured as a percentage of the genome affected by such alterations, as clonal (present in all cancer cells) or subclonal (present in a subset of cancer cells) (Fig. 1).

We observed extensive intratumor heterogeneity, with a median of 30% (range, 0.5 to 93) of somatic mutations identified as subclonal and a median of 48% (range, 0.3 to 88) of copy-number alterations as subclonal (Fig. 2A, and Fig. S3 in Supplementary Appendix 1). This finding suggests that genomic-instability processes at the mutational and chromosomal level are ongoing during tumor development. Considerable variation in intratumor heterogeneity among tumors was also observed, with the number of subclonal mutations ranging from 2 to 2310 and the percentage of the genome affected by subclonal copy-number alterations ranging from 0.06 to 81% (Fig. 2A). Without multiregion whole-exome sequencing, 76% of subclonal mutations could have appeared to be clonal, which suggests the selection of subclones within individual tumor regions (Fig. S4 in Supplementary Appendix 1). Significantly more mutations were identified with the use of multiregion whole-exome sequencing than with single-sample analysis (median number, 517 vs. 398; P=0.009) or with the use of single NSCLC samples obtained from the Cancer Genome Atlas (median number, 207; P<0.001) (Fig. S5 in Supplementary Appendix 1). The Cancer Genome Atlas research network (http://cancergenome.nih.gov) was retrieved through dbGaP authorization accession number phs000178.v9.p8.

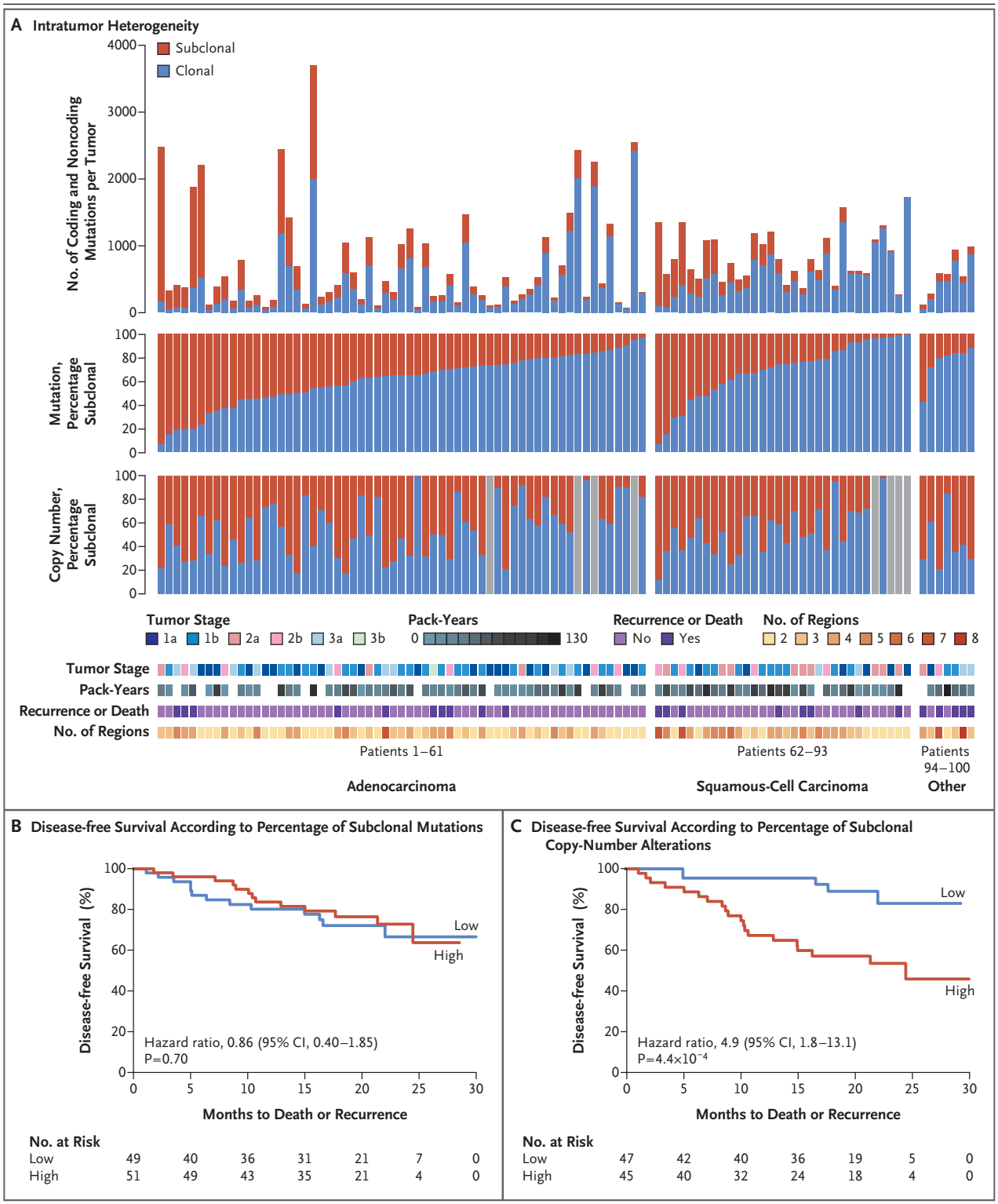
Squamous-cell carcinomas carried significantly more clonal mutations than did adenocarcinomas (P=0.003) (Fig. S6 in Supplementary Appendix 1). This finding potentially reflects differences in smoking history, with a median of 32 pack-years for adenocarcinomas and 41 pack-years for squamous-cell carcinomas (P=0.047) (Fig. S7 in Supplementary Appendix 1). There were no significant differences between squamous-cell carcinomas and adenocarcinomas in the number or proportion of subclonal mutations (P=0.72) (Fig. S6 in Supplementary Appendix 1) or within specific adenocarcinoma histopathological subtypes (Fig. S8 in Supplementary Appendix 1). In squamous-cell carcinomas, no significant relationship was observed between intratumor heterogeneity and clinical variables (Table S5 in Supplementary Appendix 1).

Figure 2 (facing page). Genomic Heterogeneity of Tumors Obtained from Patients with Non–Small-Cell Lung Cancer (NSCLC).

Panel A shows the number of coding and noncoding mutations that were detected in each tumor region in the study, according to tumor stage, smoking history, outcome of recurrence or death, and number of regions affected. The percentages of somatic mutations and copy-number alterations that were found to be clonal or subclonal in each tumor are shown below the number of mutations. The percentages of study patients who were disease-free over a 30-month period are shown according to whether the patients had a high proportion (above the median) or a low proportion (below the median) of subclonal mutations (Panel B) or of subclonal copy-number alterations (Panel C). There was no significant association between the proportion of subclonal mutations and relapse-free survival (P=0.70), but patients who had tumors with a high proportion of subclonal copy-number alterations were at significantly higher risk for recurrence or death than those with a low proportion ($P=4.4\times10^{-4}$).

In adenocarcinomas, tumor stage positively correlated with the proportion of subclonal copy-number alterations, and Ki67 staining positively correlated with the burden of both clonal and subclonal mutations, as well as with the proportion of subclonal copy-number alterations (Table S5 in Supplementary Appendix 1). Furthermore, in adenocarcinomas, a significantly higher clonal and subclonal mutational burden was observed in smokers than in patients who had never smoked (Fig. S9 in Supplementary Appendix 1).

There was no significant association between the proportion of subclonal mutations (median in the cohort, 30%) and relapse-free survival (Fig. 2B). However, in this preliminary analysis, patients who had tumors with a high proportion of subclonal copy-number alterations (≥48%, the cohort median) were at higher risk for recurrence or death than those with a low proportion (hazard ratio, 4.9; 95% confidence interval [CI], 1.8 to 13.1; $P=4.4\times10^{-4}$) (Fig. 2C). The median time until recurrence or death was 24.4 months in the higher risk group of patients compared with a median that was not reached in the lower risk group. This finding remained significant in a multivariate analysis after adjustment for age, pack-years of smoking, histologic subtype, adju-



vant therapy, and tumor stage (hazard ratio, 3.70; 95% CI, 1.29 to 10.65; P=0.01) (Table S6 in Supplementary Appendix 1). A static measure of chromosome disruption (describing the mean proportion of the genome that was aberrant across tumor regions) was not associated with

survival, which suggests that the rate of ongoing dynamic chromosomal instability, rather than the state of the genome, is prognostic (Fig. S10 in Supplementary Appendix 1).

EVOLUTIONARY HISTORIES AND TUMOR CLONAL ARCHITECTURE IN NSCLC

The number or proportion of subclonal mutations does not fully capture the extent of intratumor heterogeneity, since these measures do not reflect the number or prevalence of genetically distinct subclones that evolve in space and time. To elucidate subclones within regions and map the evolutionary history of each tumor, we clustered mutations according to their cellular prevalence. Each cluster represents a node on the phylogenetic tree of the tumor and a subclone that is present in the tumor population or has existed during its evolutionary history (Table S7, Figs. S11 and S12, and the Experimental Procedures section in Supplementary Appendix 1).

We identified 525 mutation clusters, with a median of 5 per tumor (range, 2 to 15). Most tumor regions (86%) were found to carry subclones from only a single branch of the phylogenetic tree, which emphasizes the limitations of a single diagnostic biopsy sample in accurately capturing the true extent of intratumor heterogeneity. Without the use of multiregion whole-exome sequencing, 65% of branched subclone clusters could have erroneously appeared to be clonal.

CAUSES OF INTRATUMOR HETEROGENEITY IN NSCLC

Mutational Processes

Understanding how mutational processes shape tumor evolution may inform strategies to limit tumor adaptation in the clinical setting.[11] Using published mutational signatures,[12] we analyzed clonal and subclonal mutations to determine which mutational processes contributed to intratumor heterogeneity.

The number of early mutations (accumulated before genome doubling or copy-number change) significantly correlated with the burden of mutations associated with smoking (mutational signature 4), with Spearman's rank correlations of 0.90 ($P<1.1\times10^{-16}$) for adenocarcinomas and of 0.84 ($P=3.9\times10^{-9}$) for squamous-cell carcinomas. This finding was consistent with the identification of mutations induced by tobacco carcinogens as being a key influence on trunk length (i.e., the number of mutations found in the most recent common ancestor of all cancer cells) and was reflected in the significant correlation between pack-years and truncal signature 4 mutations in adenocarcinomas (Spearman's rank correlation, 0.63; $P=5.3\times10^{-8}$). In samples obtained from 7 of 12 patients with adenocarcinomas who were long-term former smokers (with >20 years since last tobacco exposure), a smoking signature could be detected in late clonal mutations (>30% with signature 4). This finding was suggestive of a long period of tumor latency in the evolution of lung adenocarcinomas before clinical presentation.

In squamous-cell carcinomas, no significant correlation was observed between pack-years and smoking-related signature 4 (Spearman's rank correlation, 0.10; $P=0.57$), and the timing of genome doubling (ratio of the number of early mutations to the number of late mutations) was significantly later than in adenocarcinomas (Fig. S13 in Supplementary Appendix 1). Intriguingly, Patient CRUK0093, who had squamous-cell carcinoma, had a large burden of clonal signature 4 mutations (>1000) despite having been identified as a lifelong nonsmoker. This patient's occupational history indicated exposure to chemicals that included arsenic, benzene, bisphenol, and polybrominated diphenyl ethers and coal tar, which may mimic the mutagenic effects of tobacco exposure.

There were significant correlations between the subclonal mutation burden and the number of subclonal mutations that were classified as clocklike signatures 1A (spontaneous deamination of methylated cytosines) and 5 (of unknown cause).[13] The number of subclonal mutations was also significantly correlated with signatures 2 and 13 (induced by APOBEC, a family of cytidine deaminase enzymes involved in messenger RNA editing) but not with signature 4 (smoking)[12] (Fig. S14 in Supplementary Appendix 1). (APOBEC cytidine deaminases, which are usually involved in innate immunity and RNA editing, have been found to be enriched in several tumor types and act as an important source of mutagenesis.[14]) Tumors with the largest subclonal mutation burden had extensive APOBEC-mediated mutagenesis (e.g., those obtained from Patients CRUK0001, CRUK0006, CRUK0020, and CRUK0063), and spatial heterogeneity in APOBEC mutations was observed in 15 tumors (Figs. S11 and S14 in

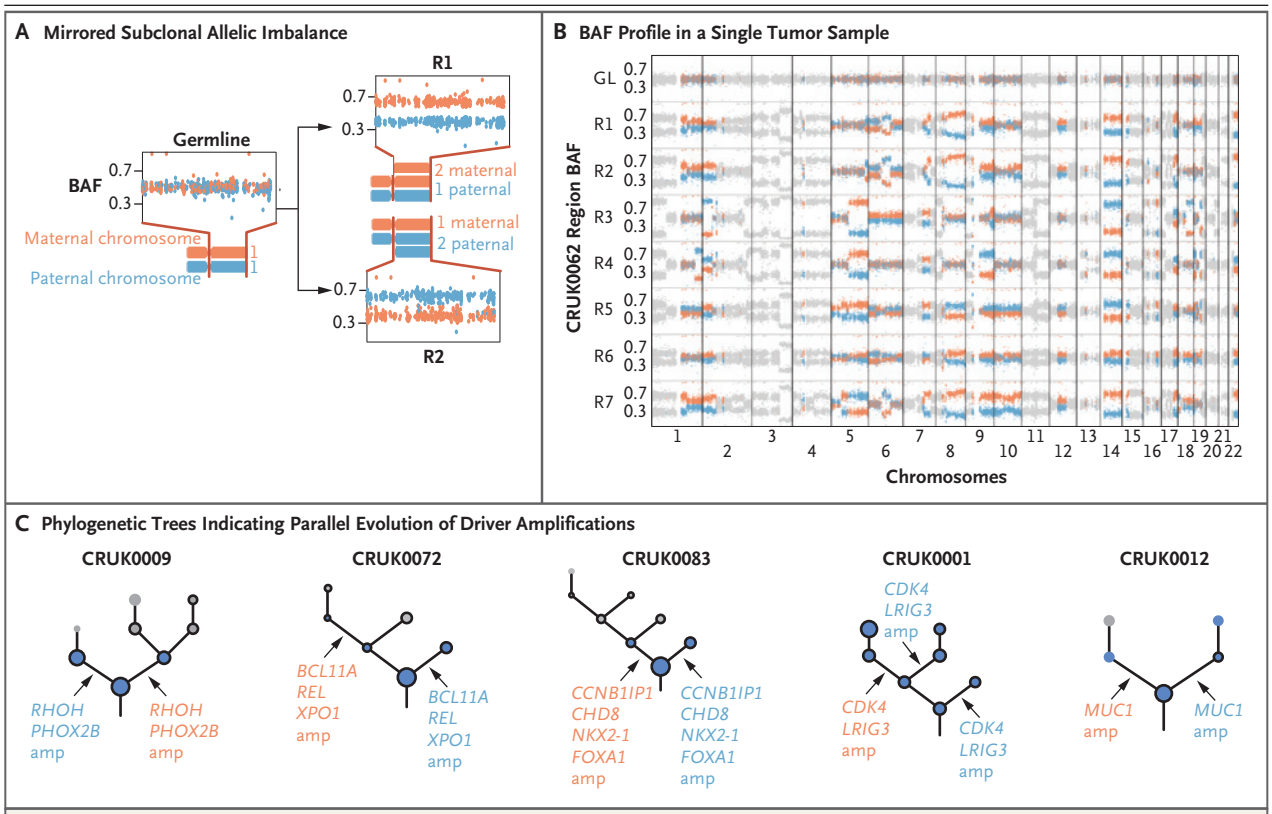


Figure 3. Drivers of Intratumor Heterogeneity.

Panel A shows an example of mirrored subclonal allelic imbalance. This occurs when the maternal allele is gained or lost in a subclone in one region and the paternal allele is gained or lost in a different subclone in another region. Such imbalance indicates additional ongoing chromosomal instability and can be inferred through multiregion whole-exome sequencing by using the frequencies at which heterozygous germline single-nucleotide polymorphisms (SNPs) (termed B-allele frequency [BAF]) are detected. The BAF of heterozygous SNPs is plotted in the same color as their parental chromosome of origin. Panel B shows the BAF profile across the genome of a tumor sample obtained from Patient CRUK0062. Areas of BAF in regions (including tumor regions R1 through R7 and a germline [GL] reference region) that have mirrored subclonal allelic imbalance are highlighted in blue or orange. Events that showed mirrored subclonal allelic imbalance were identified in more than 40% of the genome. Panel C shows phylogenetic trees that indicate parallel evolution of driver amplifications detected through the observation of mirrored subclonal allelic imbalance (arrows). Subclones that are colored blue carry a cancer driver event, and those that are colored gray carry no driver event; black outlining of the circles indicates that the subclone appears to be clonal in at least one tumor region.

Supplementary Appendix 1). Tumors obtained from 19 patients had subclonal driver mutations that could be attributed to APOBEC activity, which illustrates how APOBEC mutagenesis may frequently induce a subclonal driver event that may contribute to subclonal expansions.

Chromosomal Instability and Genome Doubling

Given the association between intratumor heterogeneity characterized by copy-number alterations and shorter relapse-free survival, we further explored the dynamics of chromosomal alterations in different tumor regions and the extent to which chromosomal instability may drive intratumor heterogeneity. By leveraging germline heterozygous single-nucleotide polymorphisms in tumors by means of multiregion whole-exome sequencing, it is possible to determine whether the same or distinct parental alleles are gained or lost in distinct subclones on different branches of the phylogenetic tree of a tumor. Specifically, if the maternal allele is gained or lost in a subclone in one region, yet the paternal allele is gained or lost in a different subclone in another region, it will result in a mirrored subclonal allelic imbalance profile (Fig. 3A and 3B). Such an imbalance, which indicates additional ongoing chromosomal instability, may also reflect parallel

evolution involving multiple distinct events converging on the same genes in different subclones (Fig. S15 in Supplementary Appendix 1). This phenomenon was observed in 62% of 92 tumors with copy-number data on multiregion whole-exome sequencing (found in 30 adenocarcinomas, 23 squamous-cell carcinomas, and 4 other samples). In total, we detected 375 mirrored subclonal allelic imbalance events that varied in size from focal to whole chromosome and involved 1 to 43% of affected tumor genomes (Fig. S16 in Supplementary Appendix 1).

Chromosomal instability may also directly contribute to mutational heterogeneity through loss of genomic segments carrying clonal mutations. Overall, a median of 13% of subclonal mutations (range, 0 to 56) per sample are probably subclonal as a result of loss events associated with copy-number alterations, which suggests that chromosomal instability may be an initiator of both copy-number and mutational heterogeneity (Fig. S17 in Supplementary Appendix 1).

Accumulating evidence suggests that genome-doubling events are associated with the propagation of chromosomal instability by cancer cells and may predict a poor prognosis.[15-17] Genome-doubling events were identified in 76% of tumors and appeared to be clonal in all but three of these tumors (from Patients CRUK0011, CRUK0062, and CRUK0063), which suggests that whole-genome duplication is an early event in NSCLC evolution. In adenocarcinomas, we observed a significant association between genome doubling and the frequency of both subclonal mutations ($P=0.02$) and subclonal copy-number alterations ($P=0.003$) (Fig. S18 in Supplementary Appendix 1). Moreover, mirrored subclonal allelic imbalance was significantly enriched in genome-doubled tumors ($P=0.004$ by Fisher's exact test) (Fig. S16 in Supplementary Appendix 1).

SELECTION AND PARALLEL EVOLUTION

Deciphering evidence of ongoing selection in tumors may shed light on evolutionary constraints, which may identify therapeutic targets. Constraints and selection are exemplified by the occurrence of parallel evolution, in which somatic events in distinct branches within a single tumor converge on the same gene, protein complex, or pathway.

No evidence of parallel evolution was found at the mutational level. However, focal amplifications of different parental alleles in distinct subclones occurred in 5 tumors and affected known cancer genes, including *MUC1*, *CDK4*, *CHD8*, and *NKX2-1* (Fig. 3C, and Fig. S19 in Supplementary Appendix 1). At the chromosome-arm level, potential parallel evolution was observed in 13 tumors (5 adenocarcinomas, 6 squamous-cell carcinomas, and 2 other tumors). Most parallel evolution of chromosome-arm gains (in 10 of 11 samples) and losses (in 6 of 8 samples) have been previously classified as significantly gained or lost in NSCLC,[3,7] a finding that is consistent with positive selection operating later in tumor evolution (Fig. S20, S21, and S22 in Supplementary Appendix 1).

To empirically estimate positive selection at the mutational level, we used a ratio of substitution rates at nonsynonymous sites to those at synonymous sites (dN/dS) that accounts for the trinucleotide context of each mutation and determines whether there is an enrichment of protein-altering mutations as compared with the background mutation rate.[18] Evidence for positive selection (dN/dS, >1) was observed when all exonic missense mutations were considered (Table S8 in Supplementary Appendix 1). This finding suggests that mutations may be shaped by selection in NSCLC. However, when mutations were temporally dissected, significant positive selection was observed for late, but not early, mutations. Consistent with this finding, nonsense mutations were found to be depleted (dN/dS, <1) early but not late in tumor development. These data further suggest that selection is persistent in NSCLC evolution and that constraints shape evolutionary trajectories. Depletion of early nonsense mutations (dN/dS, <1) was greater in squamous-cell carcinomas than in adenocarcinomas, and the rate of acquisition of clonal driver mutations (as determined by the ratio of driver mutations to passenger mutations) was significantly greater in adenocarcinomas than in squamous-cell carcinomas ($P=0.001$ by the Wilcoxon test).

CLONAL AND SUBCLONAL DRIVER ALTERATIONS AND TIMING OF GENOMIC EVENTS

Determining whether a cancer driver event occurs early or late can indicate whether it is involved in tumor initiation or maintenance, and its clonality may inform potential therapeutic

strategies, since subclonal alterations will be present in only a proportion of cells and when targeted may result in reduced treatment efficacy.[19] We identified 795 driver events (range in adenocarcinomas, 1 to 19; range in squamous-cell carcinomas, 2 to 21). Of these events, 219 in 77 tumors were found to be subclonal (range in adenocarcinomas, 0 to 10; range in squamous-cell carcinomas, 0 to 12) and 576 to be clonal (range in adenocarcinomas, 1 to 18; range in squamous-cell carcinomas, 1 to 14) (Fig. S23 in Supplementary Appendix 1 and Table S9 in Supplementary Appendix 2). Significantly more driver alterations were identified with the use of multiregion whole-exome sequencing than with single-sample analysis ($P=0.004$ by the Wilcoxon test) (Fig. S24 in Supplementary Appendix 1).

Alterations in certain cancer genes were not only primarily clonal but almost always occurred before genome duplication, which suggests involvement in tumor initiation (Fig. 4). In adenocarcinomas, these alterations included targetable mutations or amplifications in *EGFR*, *MET*, and *BRAF*, as well as amplifications in *TERT*, 8p loss, and 5p gain. In squamous-cell carcinomas, mutations in *NOTCH1*, amplifications in *FGFR1* and in the 3q region (which includes *SOX2* and *PIK3CA*), and loss of 3p, 4p, 5q, and 17p were early clonal events. Mutations in *TP53* were predominantly clonal and early for both subtypes. Conversely, other driver events, including mutations in *KMT2C* and *COL5A2* in adenocarcinomas and in *PIK3CA* in squamous-cell carcinomas, while predominantly clonal, often occurred after genome duplication, which suggests their involvement in tumor maintenance or progression. Except for alterations in *TP53*, *ATM*, *CHEK2*, and *MDM2*, 51% of 72 driver alterations affecting chromatin remodeling, histone methylation, or DNA damage response and repair were subclonal or late in both histologic subtypes (23 of 41 events in adenocarcinomas and 14 of 31 events in squamous-cell carcinomas) (Fig. S25 in Supplementary Appendix 1). *UBR5*, with a known role in differentiation and DNA damage response, was one of the most frequently altered genes later in evolution in both adenocarcinomas and squamous-cell carcinomas. Other genes that were subject to frequent subclonal or late alterations in adenocarcinomas included *NF1* and *NOTCH1*, along with 3p, 13q, and 21p loss and 7q and 8q gain, whereas in squamous-cell carcinomas, alterations in *MLH1* and *KRAS*, along with 10q loss and 7p, 8q, and 20q gain, were late events.

Driver mutations that occurred early showed a significantly greater tendency to occur in established histologic-subtype–specific cancer genes than did late or subclonal driver mutations, which affected a broader selection of pan-cancer genes[20] (Fig. S26 in Supplementary Appendix 1). These data are consistent with the dN/dS mutation-selection analysis and suggest that constraints inherent in cancer evolution vary as tumors develop, which potentially renders more evolutionary paths permissive for progression.

Overall, 86 of the 100 tumors in our study had alterations that are being investigated in NSCLC in genomically profiled drug studies, including the National Lung Matrix Trial (NLMT)[21] and the Molecular Analysis for Therapy Choice (MATCH) trial.[22] Of these 86 tumors, 17 (20%) had subclonal targetable mutations and copy-number alterations. In 12 of these 17 tumors (71%), both a clonal and a subclonal targetable alteration were present, which indicates how targets might be prioritized for therapeutic intervention (Fig. S27 in Supplementary Appendix 1).

DISCUSSION

Intratumor heterogeneity provides the fuel for tumor evolution and drug resistance.[23] Here, we have provided an analysis of NSCLC evolution, which has shown that intratumor heterogeneity and branched evolution are almost universal across the cohort. We also observed a common pattern of early clonal genome doubling, followed by extensive subclonal diversification.

These data may have important implications for our understanding of tumor biology and therapeutic control in NSCLC. Certain targetable driver mutations, including those in *EGFR*, *MET*, and *BRAF*, were almost exclusively clonal and early, which explains the robust and uniform responses that are often seen across multiple sites of disease when these alterations are targeted.[24-26] However, more than 75% of the tumors in our study carried a subclonal driver alteration, including in genes such as *PIK3CA*, *NF1*, *KRAS*, *TP53*, and *NOTCH* family members. Moreover, a large fraction of subclonal driver mutations appeared to be clonal in a single region but were

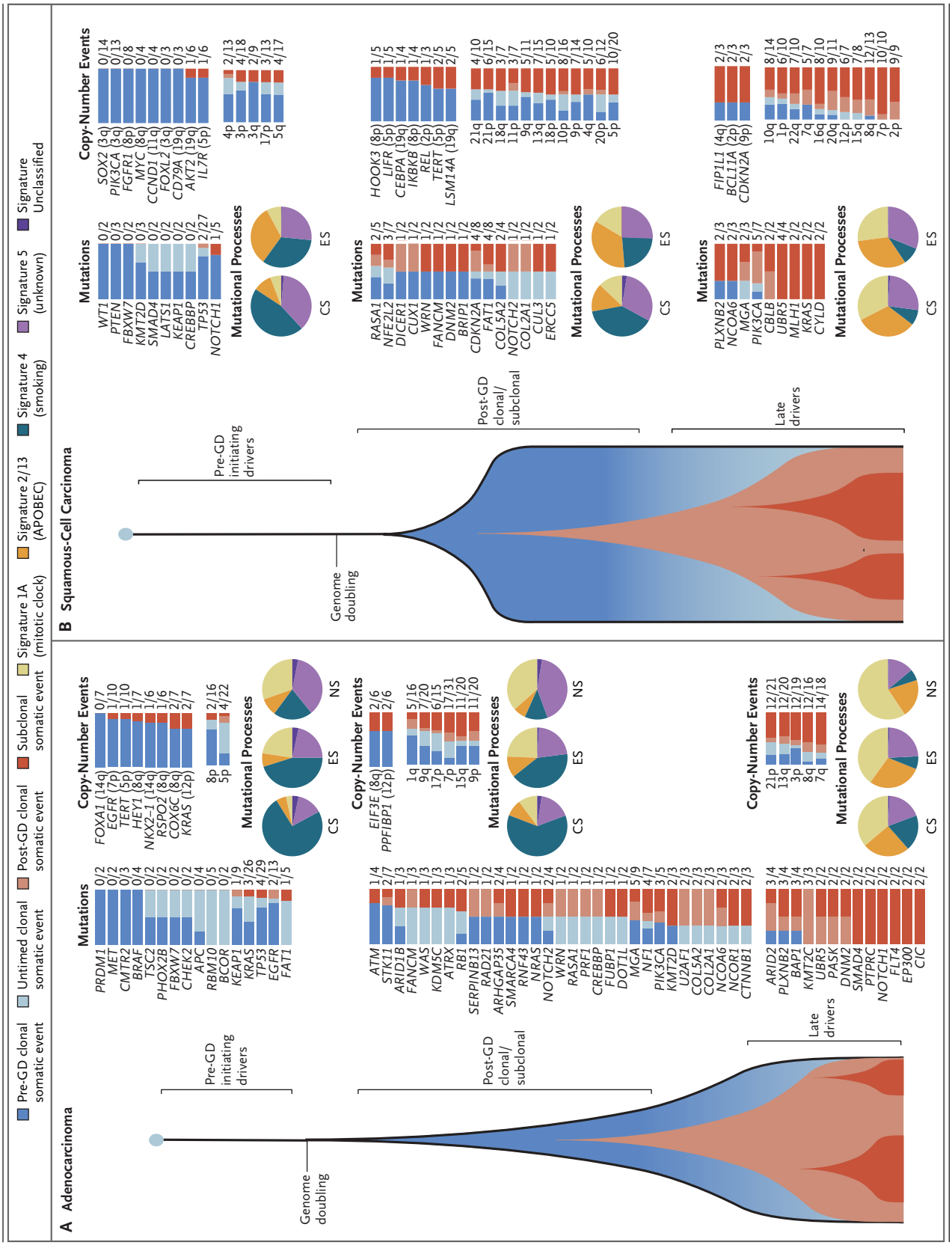


Figure 4 (facing page). Timing of Somatic Events in NSCLC Evolution.

A diagram of tumor evolution in adenocarcinoma (Panel A) and squamous-cell carcinoma (Panel B) shows the approximate timing of genomic aberrations with respect to the cancer life history. The timing of mutations and copy-number events is shown as bars indicating whether the events are clonal or subclonal. Clonal mutations and chromosome-arm events are further timed as early or late with respect to genome doubling (GD). The frequency of mutations and copy-number alterations (subclonal and total) is indicated on the right side of the bars. Pie charts show the fraction of estimated mutations for each signature, averaged across current smokers or recent ex-smokers (CS), long-term (>20 years) former smokers (ES), and lifelong never smokers (NS) at three different time points. Only genes that were mutated in at least two patients or that had copy-number alterations in at least 20% of the patients in the cohort are shown.

absent or subclonal in other regions, which confirmed the limitations of sampling single tumor regions and emphasized the ability of multiregion whole-exome sequencing to define the clonality of driver events for prioritization of drug targets.

Late mutations in tumor-suppressor genes that occur after genome doubling often affected only one allele, which potentially left the wild-type alleles intact. Although this finding could indicate that late tumor-suppressor mutations are often passenger events that do not contribute to tumor progression, it is also plausible that germline defects, subclonal copy-number loss, haploinsufficiency, or transcriptional regulation may act to limit wild-type expression. In contrast to early mutations, late driver mutations were not specific to the NSCLC subtype and often occurred in cancer genes that have been identified in other tumor types; a high proportion occurred in genes that are involved in the maintenance of genome integrity through DNA damage response and repair, chromatin remodeling, and histone methylation. Such mutations may remove tissue-specific constraints on the cancer genome and provide advantages to emerging subclones later in evolution. However, the observation of parallel evolution of driver copy-number alterations that were identified through mirrored subclonal allelic imbalance, including in *CDK4*, *FOXA1*, and *BCL11A*, suggests that despite extensive diversity, specific constraints, which could be therapeutically exploited, may operate later in tumor evolution.

Tumors with the highest subclonal mutational burden had extensive APOBEC-mediated mutagenesis, and 19 tumors carried subclonal driver mutations within an APOBEC context. This finding suggests that targeting the enzymatic activity of APOBEC may provide a means of limiting subclone diversification. The clonal mutation burden was significantly enriched in patients with a smoking history. Conceivably, this finding could be exploited for therapeutic benefit through the use of peptide vaccines or adoptive cell therapy against clonal neoantigens that are present in every tumor cell. However, the observation that clonal mutations can be lost owing to later copy-number events could limit the efficacy of such strategies, especially in tumors with high chromosome instability.

Finally, although a single sample can provide a static measure of chromosomal complexity,[27] the use of multiregion whole-exome sequencing enables the assessment of dynamic chromosome instability, which may lead to differences in chromosomal karyotypes between NSCLC subclones. The onset of chromosome instability appears to have a considerable effect on the evolution of NSCLC; such instability appears to be the predominant driver of parallel evolution and can lead to both mutational and copy-number diversity among subclones. Elevated copy-number heterogeneity was associated with shorter relapse-free survival, which suggests that patients who have early-stage tumors with high levels of copy-number heterogeneity may represent a high-risk group who may benefit from close monitoring and early therapeutic intervention during follow-up. We are continuing to assess this association in the next 742 patients enrolled in TRACERx. Whether noninvasive prognostic approaches, such as liquid biopsy, can be used to prospectively assess the levels of chromosomal instability in the clinical setting warrants further attention.[28] In addition to ongoing efforts to target single genetic alterations, there is a need to develop a greater understanding of chromosomal instability, which can alter the copy number of a multitude of genes simultaneously. Indeed, therapeutic efforts that can attenuate this process may limit the ensuing heterogeneity and tumor evolution that drive poor rates of relapse-free survival. In the analysis presented here, we provide a

census of driver events in early-stage NSCLC in relation to clonality and show that chromosomal instability is not only a significant driver of parallel evolution but also a predictor of poor outcome.

Supported by Cancer Research UK (CRUK), the CRUK Lung Cancer Centre of Excellence, the University College London Hospitals Biomedical Research Centre, the CRUK University College London Experimental Cancer Medicine Centre, the Rosetrees Trust, the Francis Crick Institute (which receives its core funding from CRUK [FC001169 and FC001202]), the U.K. Medical Research Council (FC001169 and FC001202), and the Wellcome Trust (FC001169 and FC001202). Dr. Swanton is a Royal Society Napier Chair in Oncology and is funded by CRUK (TRACERx and CRUK Cancer Immunotherapy Catalyst Network), the National Institute for Health Research, Novo Nordisk Foundation (ID16584), the European Research Council, and PloidyNet (a Marie Curie Initial Training Network). Dr. Van Loo is a Winton Group Leader in recognition of the support of the Winton Charitable Foundation in the establishment of the Francis Crick Institute.

Disclosure forms provided by the authors are available with the full text of this article at NEJM.org.

We thank all the patients who participated in this study and representatives of Illumina and Agilent who provided sequencing infrastructure support.

APPENDIX

The authors' full names and academic degrees are as follows: Mariam Jamal-Hanjani, M.D., Ph.D., Gareth A. Wilson, Ph.D., Nicholas McGranahan, Ph.D., Nicolai J. Birkbak, Ph.D., Thomas B.K. Watkins, M.C.I.T., Selvaraju Veeriah, Ph.D., Seema Shafi, Ph.D., Diana H. Johnson, B.Sc., Richard Mitter, M.Sc., Rachel Rosenthal, M.Sc., Max Salm, Ph.D., Stuart Horswell, M.Math., Mickael Escudero, M.Sc., Nik Matthews, B.Sc., Andrew Rowan, B.Sc., Tim Chambers, M.Sc., David A. Moore, M.D., Samra Turajlic, M.D., Ph.D., Hang Xu, Ph.D., Siow-Ming Lee, M.D., Ph.D., Martin D. Forster, M.D., Ph.D., Tanya Ahmad, M.D., Crispin T. Hiley, M.D., Ph.D., Christopher Abbosh, M.D., Mary Falzon, M.D., Elaine Borg, M.D., Teresa Marafioti, M.D., David Lawrence, M.D., Martin Hayward, M.D., Shyam Kolvekar, M.D., Nikolaos Panagiotopoulos, M.D., Sam M. Janes, M.D., Ph.D., Ricky Thakrar, M.D., Asia Ahmed, M.D., Fiona Blackhall, M.D., Ph.D., Yvonne Summers, M.D., Ph.D., Rajesh Shah, M.D., Leena Joseph, M.D., Anne M. Quinn, M.D., Ph.D., Phil A. Crosbie, M.D., Ph.D., Babu Naidu, M.D., Gary Middleton, M.D., Gerald Langman, M.D., Simon Trotter, M.D., Marianne Nicolson, M.D., Hardy Remmen, M.D., Keith Kerr, M.D., Mahendran Chetty, M.D., Lesley Gomersall, M.D., Dean A. Fennell, M.D., Ph.D., Apostolos Nakas, M.D., Sridhar Rathinam, M.D., Girija Anand, M.D., Sajid Khan, M.D., Peter Russell, M.D., Ph.D., Veni Ezhil, M.D., Babikir Ismail, M.D., Melanie Irvin-Sellers, M.D., Vineet Prakash, M.D., Jason F. Lester, M.D., Malgorzata Kornaszewska, M.D., Ph.D., Richard Attanoos, M.D., Haydn Adams, M.D., Helen Davies, M.D., Stefan Dentro, M.Sc., Philippe Taniere, M.D., Ph.D., Brendan O'Sullivan, B.Sc., Helen L. Lowe, Ph.D., John A. Hartley, Ph.D., Natasha Iles, Ph.D., Harriet Bell, M.Sc., Yenting Ngai, B.Sc., Jacqui A. Shaw, Ph.D., Javier Herrero, Ph.D., Zoltan Szallasi, M.D., Roland F. Schwarz, Ph.D., Aengus Stewart, M.Sc., Sergio A. Quezada, Ph.D., John Le Quesne, M.D., Ph.D., Peter Van Loo, Ph.D., Caroline Dive, Ph.D., Allan Hackshaw, M.Sc., and Charles Swanton, M.D., Ph.D.

The authors' affiliations are as follows: the Cancer Research UK Lung Cancer Centre of Excellence (M.J.-H., G.A.W., N. McGranahan, N.J.B., S.V., S.S., D.H.J., R.R., S.-M.L., M.D.F., C.A., S.M.J., C.D., C.S.), London and Manchester, Good Clinical Laboratory Practice Facility, University College London (UCL) Experimental Cancer Medicine Centre (H.L.L., J.A.H.), Bill Lyons Informatics Centre (J.H.), and Cancer Immunology Unit (S.A.Q.), UCL Cancer Institute, the Translational Cancer Therapeutics Laboratory (G.A.W., N. McGranahan, N.J.B., T.B.K.W., A.R., T.C., S. Turajlic, H.X., C.T.H., C.S.), Department of Bioinformatics and Biostatistics (R.M., M.S., S.H., M.E., A.S.), Advanced Sequencing Facility (N. Matthews), and Cancer Genomics Laboratory (S.D., P.V.L.), Francis Crick Institute, the Renal and Skin Units, Royal Marsden Hospital (S. Turajlic), the Departments of Medical Oncology (M.J.-H., S.-M.L., M.D.F., T.A., C.A., C.S.), Pathology (M.F., E.B., T.M.), Cardiothoracic Surgery (D.L., M.H., S. Kolvekar, N.P.), Respiratory Medicine (S.M.J., R.T.), and Radiology (A.A.), UCL Hospitals, Lungs for Living, UCL Respiratory, UCL (S.M.J.), the Department of Radiotherapy, North Middlesex University Hospital (G.A.), the Department of Respiratory Medicine, Royal Free Hospital (S. Khan), and UCL Cancer Research UK and Cancer Trials Centre (N.I., H.B., Y.N., A.H.), London, Cancer Studies, University of Leicester (D.A.M., D.A.F., J.A.S., J.L.Q.), the Department of Thoracic Surgery, Glenfield Hospital (A.N., S.R.), and the Medical Research Center Toxicology Unit (J.L.Q.), Leicester, the Institute of Cancer Studies, University of Manchester (F.B.), the Christie Hospital (F.B., Y.S.), the Departments of Cardiothoracic Surgery (R.S.) and Pathology (L.J., A.M.Q.) and the North West Lung Centre (P.A.C.), University Hospital of South Manchester, and Cancer Research UK Manchester Institute (C.D.), Manchester, the Departments of Thoracic Surgery (B.N.) and Cellular Pathology (G.L., S. Trotter), Birmingham Heartlands Hospital, Molecular Pathology Diagnostic Services, Queen Elizabeth Hospital (P.T., B.O.), and Institute of Immunology and Immunotherapy, University of Birmingham (G.M.), Birmingham, the Departments of Medical Oncology (M.N.), Cardiothoracic Surgery (H.R.), Pathology (K.K.), Respiratory Medicine (M.C.), and Radiology (L.G.), Aberdeen University Medical School and Aberdeen Royal Infirmary, Aberdeen, the Department of Respiratory Medicine, Barnet and Chase Farm Hospitals, Barnet (S. Khan), the Department of Respiratory Medicine, Princess Alexandra Hospital, Harlow (P.R.), the Department of Clinical Oncology, St. Luke's Cancer Centre, Guildford (V.E.), the Departments of Pathology (B.I.), Respiratory Medicine (M.I.-S.), and Radiology (V.P.), Ashford and St. Peters' Hospitals, Surrey, the Department of Clinical Oncology, Velindre Hospital (J.F.L.), the Departments of Radiology (H.A.) and Respiratory Medicine (H.D.), University Hospital Llandough, the Departments of Pathology (R.A.) and Cardiothoracic Surgery (M.K.), University Hospital of Wales, and Cardiff University (R.A.), Cardiff, and Wellcome Trust Sanger Institute, Hinxton, and Big Data Institute, University of Oxford, Oxford (S.D.) — all in the United Kingdom; the Center for Biological Sequence Analysis, Department of Systems Biology, Technical University of Denmark, Lyngby (Z.S.); the Computational Health Informatics Program, Boston Children's Hospital and Harvard Medical School, Boston (Z.S.); MTA-SE-NAP, Brain Metastasis Research Group, 2nd Department of Pathology, Semmelweis University, Budapest, Hungary (Z.S.); Berlin Institute for Medical Systems Biology, Max Delbrueck Center for Molecular Medicine, Berlin (R.F.S.); and the Department of Human Genetics, University of Leuven, Leuven, Belgium (P.V.L.).

REFERENCES

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. CA Cancer J Clin 2013;63:11-30.
2. Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. CA Cancer J Clin 2011;61:69-90.
3. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543-50.
4. Cancer Genome Atlas Research Network. Comprehensive genomic character-

ization of squamous cell lung cancers. Nature 2012;489:519-25.
5. Imielinski M, Berger AH, Hammerman PS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 2012;150:1107-20.
6. Govindan R, Ding L, Griffith M, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell 2012;150:1121-34.
7. Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet 2016;48:607-16.
8. de Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 2014;346:251-6.
9. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 2014;346:256-9.
10. Jamal-Hanjani M, Hackshaw A, Ngai Y, et al. Tracking genomic cancer evolution for precision medicine: the lung TRACERx study. PLoS Biol 2014;12(7):e1001906.
11. Alexandrov LB, Nik-Zainal S, Siu HC, Leung SY, Stratton MR. A mutational signature in gastric cancer suggests therapeutic strategies. Nat Commun 2015;6:8683.
12. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. Nature 2013;500:415-21.
13. Alexandrov LB, Jones PH, Wedge DC, et al. Clock-like mutational processes in human somatic cells. Nat Genet 2015;47:1402-7.
14. Roberts SA, Lawrence MS, Klimczak LJ, et al. An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. Nat Genet 2013;45:970-6.
15. Carter SL, Cibulskis K, Helman E, et al. Absolute quantification of somatic DNA alterations in human cancer. Nat Biotechnol 2012;30:413-21.
16. Dewhurst SM, McGranahan N, Burrell RA, et al. Tolerance of whole-genome doubling propagates chromosomal instability and accelerates cancer genome evolution. Cancer Discov 2014;4:175-85.
17. Fujiwara T, Bandi M, Nitta M, Ivanova EV, Bronson RT, Pellman D. Cytokinesis failure generating tetraploids promotes tumorigenesis in p53-null cells. Nature 2005;437:1043-7.
18. Martincorena I, Roshan A, Gerstung M, et al. Tumor evolution: high burden and pervasive positive selection of somatic mutations in normal human skin. Science 2015;348:880-6.
19. Lohr JG, Stojanov P, Carter SL, et al. Widespread genetic heterogeneity in multiple myeloma: implications for targeted therapy. Cancer Cell 2014;25:91-101.
20. Lawrence MS, Stojanov P, Mermel CH, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495-501.
21. Middleton G, Crack LR, Popat S, et al. The National Lung Matrix Trial: translating the biology of stratification in advanced non-small-cell lung cancer. Ann Oncol 2015;26:2464-9.
22. Abrams J, Conley B, Mooney M, et al. National Cancer Institute's Precision Medicine Initiatives for the new National Clinical Trials Network. Am Soc Clin Oncol Educ Book 2014;:71-6.
23. Greaves M. Evolutionary determinants of cancer. Cancer Discov 2015;5:806-20.
24. Chapman PB, Hauschild A, Robert C, et al. Improved survival with vemurafenib in melanoma with BRAF V600E mutation. N Engl J Med 2011;364:2507-16.
25. Mok TS, Wu Y-L, Thongprasert S, et al. Gefitinib or carboplatin–paclitaxel in pulmonary adenocarcinoma. N Engl J Med 2009;361:947-57.
26. Cao Y, Xiao G, Qiu X, Ye S, Lin T. Efficacy and safety of crizotinib among Chinese EML4-ALK-positive, advanced-stage non-small cell lung cancer patients. PLoS One 2014;9(12):e114008.
27. McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C. Cancer chromosomal instability: therapeutic and diagnostic challenges. EMBO Rep 2012;13:528-38.
28. Ni X, Zhuo M, Su Z, et al. Reproducible copy number variation patterns among single circulating tumor cells of lung cancer patients. Proc Natl Acad Sci U S A 2013;110:21083-8.