

Mutational landscape of *EGFR*-, *MYC*-, and *Kras*-driven genetically engineered mouse models of lung adenocarcinoma

David G. McFadden^{a,1,2}, Katerina Politi^{b,c,d,1}, Arjun Bhutkar^a, Frances K. Chen^a, Xiaoling Song^{d,3}, Mono Pirun^e, Philip M. Santiago^a, Caroline Kim-Kiselak^a, James T. Platt^d, Emily Lee^f, Emily Hodges^f, Adam P. Rosebrock^{f,4}, Roderick T. Bronson^g, Nicholas D. Socci^{e,5}, Gregory J. Hannon^{f,h,5,6}, Tyler Jacks^{a,i,j,1,5}, and Harold Varmus^{k,5,7}

^aDavid H. Koch Institute for Integrative Cancer Research at MIT, Massachusetts Institute of Technology, Cambridge, MA 02139; ^bDepartment of Pathology, Yale University School of Medicine, New Haven, CT 06510; ^cSection of Medical Oncology, Department of Medicine, Yale University School of Medicine, New Haven, CT 06510; ^dYale Cancer Center, Yale University School of Medicine, New Haven, CT 06510; ^eBioinformatics Core, Memorial Sloan Kettering Cancer Center, New York, NY 10065; ^fWatson School of Biological Sciences, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; ^gDepartment of Pathology, Tufts University School of Medicine and Veterinary Medicine, North Grafton, MA 01536; ^hHoward Hughes Medical Institute, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724; ⁱHoward Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139; ^jDepartment of Biology, Massachusetts Institute of Technology, Cambridge, MA 02142; and ^kCancer Biology and Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY 10065

Contributed by Tyler Jacks, August 19, 2016 (sent for review April 14, 2016; reviewed by Allan Balmain and Anton Berns)

Genetically engineered mouse models (GEMMs) of cancer are increasingly being used to assess putative driver mutations identified by large-scale sequencing of human cancer genomes. To accurately interpret experiments that introduce additional mutations, an understanding of the somatic genetic profile and evolution of GEMM tumors is necessary. Here, we performed whole-exome sequencing of tumors from three GEMMs of lung adenocarcinoma driven by mutant epidermal growth factor receptor (*EGFR*), mutant Kirsten rat sarcoma viral oncogene homolog (*Kras*), or overexpression of *MYC* proto-oncogene. Tumors from *EGFR*- and *Kras*-driven models exhibited, respectively, 0.02 and 0.07 nonsynonymous mutations per megabase, a dramatically lower average mutational frequency than observed in human lung adenocarcinomas. Tumors from models driven by strong cancer drivers (mutant *EGFR* and *Kras*) harbored few mutations in known cancer genes, whereas tumors driven by *MYC*, a weaker initiating oncogene in the murine lung, acquired recurrent clonal oncogenic *Kras* mutations. In addition, although *EGFR*- and *Kras*-driven models both exhibited recurrent whole-chromosome DNA copy number alterations, the specific chromosomes altered by gain or loss were different in each model. These data demonstrate that GEMM tumors exhibit relatively simple somatic genotypes compared with human cancers of a similar type, making these autochthonous model systems useful for additive engineering approaches to assess the potential of novel mutations on tumorigenesis, cancer progression, and drug sensitivity.

targeted therapies (11, 12). Although individual mice develop multifocal disease, only a subset of the primary *Kras* tumors progress to metastatic disease. A distinct gene expression signature has been shown to distinguish metastatic from nonmetastatic primary tumors

Significance

Knowledge of oncogenic alterations that drive lung adenocarcinoma formation has enabled the development of genetically engineered mouse models that are increasingly being used to study the biology and therapeutic vulnerabilities of this disease. Given the importance of genomic alterations in these processes in human lung cancer, information on the mutational landscape of the mouse tumors is valuable for the design and interpretation of these experiments. In this study, we compared whole-exome sequencing data from lung adenocarcinomas induced by different lung adenocarcinoma-associated drivers. In contrast to their human counterparts, oncogene-driven lung adenocarcinomas in genetically engineered mouse models harbor few somatic mutations. These results have important implications for the use of these models to study tumor progression and response and resistance to therapy.

KRAS | EGFR | MYC | GEMM | exome

Lung cancer remains the leading cause of cancer death worldwide, estimated to have caused 158,000 deaths in the United States in 2015 (seer.cancer.gov). Lung adenocarcinoma is the most common form of lung cancer, in both smokers and nonsmokers. Tobacco mutagens cause a high mutation frequency in the somatic genomes of smoking-associated tumors, complicating identification of the genetic drivers of tumor progression (1, 2). An increasing number of somatic alterations that can be targeted by existing drugs or drug candidates have been identified in lung adenocarcinoma, and several of these agents have demonstrated efficacy in patients (3).

Kirsten rat sarcoma viral oncogene homolog (*KRAS*) and epidermal growth factor receptor (*EGFR*) are the most frequently mutated oncogenes in human lung adenocarcinoma (1, 2, 4, 5). *KRAS* mutations are associated with a strong history of cigarette smoking, whereas *EGFR* mutations are the most frequent oncogene alterations in lung cancers from never-smokers (6). Our groups and others have generated genetically engineered mouse models (GEMMs) of *EGFR*- and *Kras*-mutant lung adenocarcinoma (7–10). These models recapitulate key features of the human disease, including histologic architecture and response and resistance to conventional and

Author contributions: D.G.M., K.P., G.J.H., T.J., and H.V. designed research; D.G.M., K.P., F.K.C., X.S., P.M.S., C.K.-K., E.L., E.H., and A.P.R. performed research; E.H., A.P.R., N.D.S., and G.J.H. contributed new reagents/analytic tools; D.G.M., K.P., A.B., M.P., J.T.P., R.T.B., and N.D.S. analyzed data; and D.G.M., K.P., N.D.S., G.J.H., T.J., and H.V. wrote the paper.

Reviewers: A. Balmain, UCSF Helen Diller Family Comprehensive Cancer Center; and A. Berns, The Netherlands Cancer Institute.

The authors declare no conflict of interest.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the NCBI BioProject database, www.ncbi.nlm.nih.gov/bioproject (accession no. PRJNA339310).

¹To whom correspondence may be addressed. Email: david.mcfadden@utsouthwestern.edu, tjacks@mit.edu, or katerina.politi@yale.edu.

²Present address: Division of Endocrinology, Department of Internal Medicine and Harold C. Simmons Comprehensive Cancer Center, UT Southwestern Medical Center, Dallas, TX 75235.

³Present address: Shanghai Institute for Advanced Immunochemical Studies, Shanghai Tech University, Shanghai 201210, China.

⁴Present address: Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, ON, Canada M5S 1A1.

⁵N.D.S., G.J.H., T.J., and H.V. contributed equally to this work.

⁶Present address: Cancer Research UK Cambridge Institute, University of Cambridge, Cambridge CB20RE, United Kingdom.

⁷Present address: Meyer Cancer Center, Weill Cornell Medicine, Cornell University, New York, NY 10065.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1613601113/-DCSupplemental.

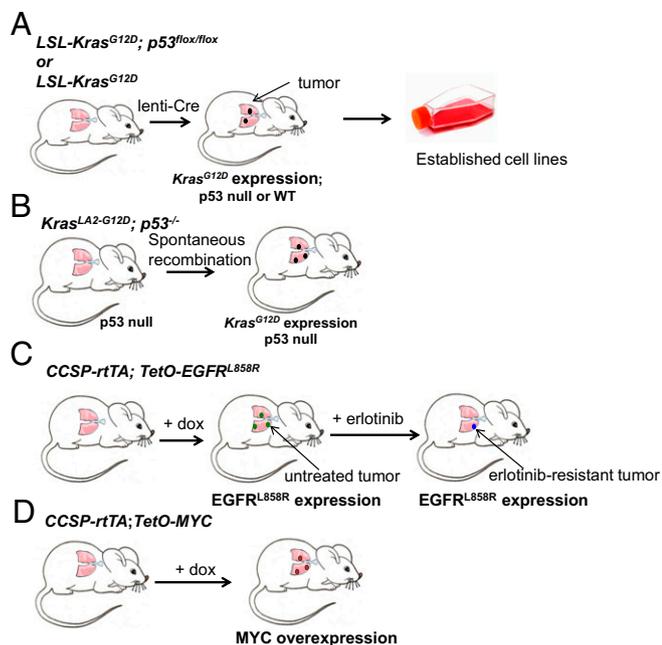


Fig. 1. Diagrams illustrating the mouse models of mutant *Kras*-, mutant *EGFR*-, and *MYC*-induced lung adenocarcinoma used in whole-exome sequencing. (A and B) *Kras* models. (A) Mice carrying conditional *Kras*^{LSL-G12D} and *p53*^{flox/flox} alleles develop lung adenocarcinomas upon administration of lenti-cre. Cell lines were generated from primary and metastatic lung tumors. Tumors and cell lines were collected for exome sequencing. (B) Mice carrying *Kras*^{LA2-G12D}; *p53*^{-/-} form lung adenocarcinomas spontaneously. Primary tumors were collected for exome sequencing. (C) *EGFR* model: Bitransgenic *CCSP-rtTA*; *TetO-EGFR*^{L858R} mice were treated with doxycycline at weaning to induce transgene expression (10). Tumors were collected from untreated tumor-bearing mice, or mice were treated with erlotinib as described until the appearance of resistant tumors (12). Untreated and erlotinib-resistant lung tumors were collected and used for exome sequencing. (D) *MYC* model: Bitransgenic *CCSP-rtTA*; *TetO-MYC* mice were treated with doxycycline at weaning to induce transgene expression. Overexpression of *MYC* in type II pneumocyte leads to the development of lung adenocarcinomas that were collected for whole-exome sequencing (26).

in a *Kras*^{G12D}-driven GEMM, suggesting that acquired genetic or epigenetic alterations underlie metastatic progression (13).

Somatic genome evolution in tumors produced in GEMMs remains incompletely characterized. Several studies have described the spectrum of acquired DNA copy number alterations in murine models (14–20). Although these studies reached varying conclusions, it appears that somatic alterations in DNA copy number, especially changes in copy number of certain whole chromosomes, are a common somatic event during tumor evolution in GEMMs. We also recently studied a GEMM of small cell lung cancer using

exome and whole-genome sequencing. In this model, which is initiated by mutation of the *p53* and retinoblastoma tumor suppressors, we identified recurrent alterations in the *PTEN/PI3K* pathway, in addition to previously identified focal DNA amplifications targeting the *Myc11* oncogene (21–23). Together, these studies suggest that, similar to human cancers, GEMM tumors can undergo extensive genome remodeling during tumor progression and that a subset of these acquired events contributes to cancer progression.

The mutational landscape of carcinogen-induced murine lung adenocarcinomas was also recently described and compared with that in tumors initiated by expression of an oncogenic *Kras* allele (24). Not surprisingly, single nucleotide mutations, including *Kras* mutations, were more frequently observed in carcinogen-induced tumors. In contrast, secondary DNA copy number alterations were more prevalent in tumors arising in genetically engineered mice. This finding further suggests that the path of somatic alteration and selection during tumor progression depends on the specific events that initiate tumorigenesis. As previously described, the carcinogen-treated tumors acquired clonal oncogenic *Kras* mutations. However, it remains unknown whether murine lung adenocarcinomas initiated by other oncogenic drivers, or those harboring combined loss of the tumor suppressor *p53*, acquire similar or distinct patterns of somatic alteration during tumor evolution and progression. Here, we describe the somatic evolution of a panel of tumors and tumor-derived cell lines derived from GEMMs of lung adenocarcinoma initiated by *Kras*, *EGFR*, and *MYC* (10, 25–27).

Results

Mouse Models of Lung Adenocarcinoma Acquire Few Somatic Point Mutations. We generated a panel of tumor specimens and tumor cell lines from GEMMs of *EGFR*-, *KRAS*-, and *MYC*-mutant lung adenocarcinomas (Fig. 1). We performed whole-exome sequencing on tumors and cell lines from these models to profile the spectrum of genomic alterations acquired during tumorigenesis and progression (Table 1 and Tables S1 and S2).

We initially focused on somatic point mutations. Several methods of mutation calling have been developed, but agreement among the various methods is poor (28); at the beginning of our studies, there was no independent evaluation of which method had optimal sensitivity and specificity. An added challenge was that available methods were developed to call mutations in human tumors, and many had parameters (e.g., background mutation rate) that were optimized for human samples. Therefore, we created a controlled dataset to assess the performance of our variant-calling pipeline in a murine background by simulating mixtures of tumor and normal DNA using exon capture of mixtures of germ-line DNA from different inbred mouse strains (Fig. S14).

As a first step, we generated exon-capture sequencing libraries with tail DNA from inbred C57BL/6 and 129S1/SvImJ mice. To simulate tumor subclonal heterogeneity and contamination with infiltrating nontumor stromal cells, we serially diluted the 129S1/SvImJ library (mimicking “tumor DNA”) into the C57BL/6 library

Table 1. Lung adenocarcinoma GEMMs used for the sequencing study

Initiating driver	Cell line or tumor	Genotype	Tumor Induction	Treatment	No. studied
<i>Kras</i> -G12D	Cell line	<i>Kras</i> -LSL-G12D; <i>Trp53</i> FL/FL	Lenti-Cre	n/a	15
<i>Kras</i> -G12D	Tumor	<i>Kras</i> -LSL-G12D; <i>Trp53</i> FL/FL	Lenti-Cre	n/a	9
<i>Kras</i> -G12D	Tumor	<i>Kras</i> -LSL-G12D	Lenti-Cre	n/a	8
<i>Kras</i> -G12D	Tumor	<i>Kras</i> -LA2; <i>Trp53</i> ^{-/-}	Spontaneous recombination	n/a	4
<i>EGFR</i> -L858R	Tumor	<i>CCSP-rtTA</i> ; <i>tetO::EGFR</i> -L858R	Doxycycline	None	10
<i>EGFR</i> -L858R	Tumor	<i>CCSP-rtTA</i> ; <i>tetO::EGFR</i> -L858R	Doxycycline	Erlotinib-resistant	6
<i>MYC</i>	Tumor	<i>CCSP-rtTA</i> ; <i>tetO::MYC</i>	Doxycycline	n/a	5

GEMMs are grouped by initiating driver event, and columns show the tumor type (studied as a cell line or primary tumor tissue), genotype of animals, method of cancer gene induction, and number of samples from each model included in the study. Additional details on the models are provided in [Appendix](#) and [Fig. S1](#). n/a, not applicable.

(mimicking “normal DNA”) (Fig. S14). The starting libraries and mixtures were sequenced to a median average depth of 132× (range from 92× to 205×). Somatic mutations were identified by using both muTect (Version 1.1.4) and a somatic mutation caller developed by our group (hereafter referred to as the HaJaVa caller) based on the GATK UnifiedGenotyper with filtering to call somatic events (as described in detail in *Methods* and *SI Appendix*) (29). By using this approach, individual germ-line polymorphisms can be traced through the serially diluted libraries, mimicking somatic variant detection in tumors. This dataset was used to estimate the false-positive and false-negative rates at decreasing allelic fraction.

At the indicated depth of coverage, muTect was highly sensitive, particularly at low allelic fractions that might be found in samples with a low fraction of tumor cells or as a result of subclonal mutational events, but it exhibited a higher false-positive rate. In contrast, the HaJaVa caller exhibited a higher true-positive rate, but also a higher false-negative rate at low allelic fraction (Fig. S1 B–D, *SI Appendix*, and *Table S3*). We found that the intersection of the two callers exhibited a lower false-positive rate than either caller alone, reducing missed calls by ~50%, with a minimal increase in the false-negative rate. Evidence that aggregating calls from multiple methods improves performance was also recently described, supporting this approach (24, 30). Therefore, we used the intersection of the HaJaVa and muTect calling algorithms to identify somatic mutations and to compare datasets among the EGFR, MYC, and Kras models.

Kras-Driven Models of Lung Adenocarcinoma. We generated DNA from a large panel of tumors and tumor-derived cell lines from *Kras*^{LSL-G12D}-based mouse models for whole-exome sequencing (Table 1). We sequenced DNA from 15 tumor cell lines derived from tumor-bearing *Kras*^{LSL-G12D};*Trp53*^{fl/fl} mice, following the lentivirus-based delivery of cre recombinase (Fig. 1) (31). In nine cases, we also sequenced DNA from the parental tumor from which the cell lines were derived. In addition, to determine whether expression of cre recombinase generated unexpected mutations, we sequenced DNA from four tumors arising in the *Kras*^{LA2-G12D};*Trp53*^{-/-} model, in which spontaneous recombination at the *Kras* locus, rather than cre-induced recombination, initiated tumorigenesis (32). Finally, to assess the potential impact of p53 loss on mutation frequency, we sequenced DNA from eight tumors initiated in *Kras*^{LSL-G12D};*Trp53*^{WT} animals.

We observed a median nonsynonymous mutation frequency of 0.07 per Mb (including missense and nonsense mutations and mutations affecting splicing signals; range 0.00–0.46; Fig. 2A) in *Kras*-driven tumors (cell lines are excluded from this analysis; see below). Interestingly, we did not observe statistically different mutation frequencies between *Trp53*-null (0.07 mutations per Mb) and wild-type (0.06 mutations per Mb; $P = 0.60$; Fig. 2B) tumors. Tumors initiated in the *Kras*^{LSL-G12D};*Trp53*^{fl/fl} model harbored numbers of mutations (0.09 mutations per Mb) similar to those in the *Kras*^{LA2-G12D};*Trp53*^{-/-} model (0.03 mutations per Mb; $P = 0.1$; Fig. 2C). With the *Kras*^{LSL-G12D};*Trp53*^{fl/fl} model, we observed a statistically significant increase in the mutation frequency in tumor-derived cell lines (0.25 mutations per Mb) compared with primary tumor specimens (0.07 mutations per Mb; $P = 0.001$; Fig. 2D). This result may reflect the presence of subclonal mutations present in the genomes of cell line-founder clones that would be expected to be enriched during the generation of tumor cell lines; however, we cannot exclude the possibility that new mutations arose de novo during generation of the cell lines.

If the higher mutation frequency observed in tumor cell lines compared with dissected tumor specimen reflects expansion of a tumor subclone during generation of the cell line rather than acquisition of new mutations in vitro, the mutational burden of single cell genomes within tumors must be similar to the tumor cell lines. Therefore, the lower frequency of mutations observed in tumors compared with cell lines suggests that many of the

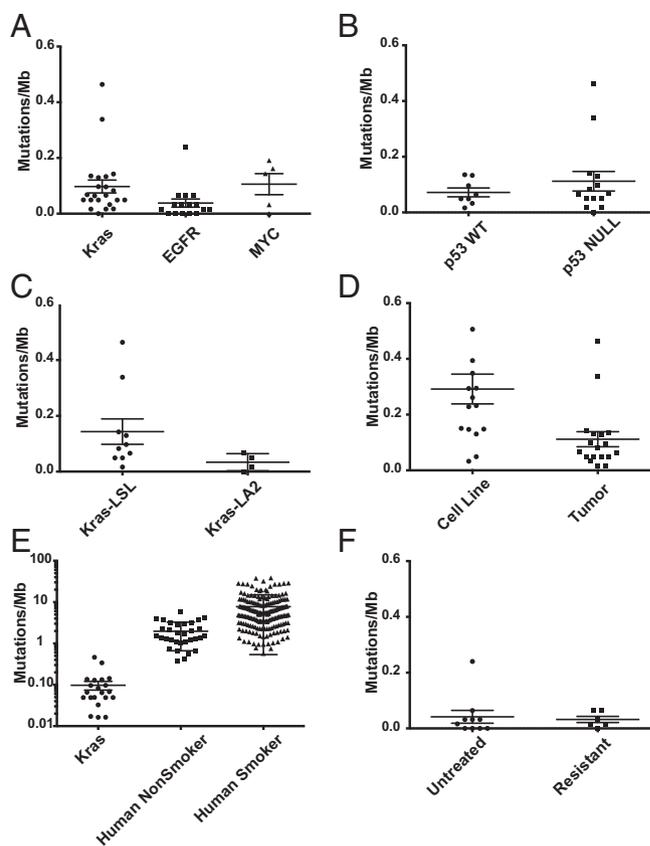


Fig. 2. Low mutational burden in GEMM models of lung cancer. (A) Dot plots showing the nonsynonymous mutation frequency observed from whole-exome sequencing datasets in murine LUADs induced by oncogenic *Kras* (tumors from either the LA2 or LSL models; tumor-derived cell lines are excluded; see below), EGFR, or overexpression of MYC. (B) *Trp53* null vs. wild-type tumors. (C) *Kras*^{LA2-G12D};*p53*^{-/-} vs. *Kras*^{LSL-G12D};*Trp53*^{fl/fl} tumors. (D) *Kras*^{LSL-G12D};*Trp53*^{fl/fl} tumors vs. tumor-derived cell lines. (E) Comparison of *Kras*^{G12D}-induced tumors to human lung adenocarcinomas (ref. 4) from smoking and nonsmoking patients, shown in log scale. (F) Untreated vs. drug-resistant EGFR^{L858R}-induced LUADs. Mean and SEM are shown.

mutations observed in cell lines existed in tumor subclones occupying a minority of the tumor mass, below the resolution of detection by exome sequencing. In addition, not all mutations observed in dissected tumor specimens, including events with relatively high allelic fraction (>25%), were detected in the tumor cell lines derived from the tumor. Together, these results suggest a degree of clonal heterogeneity within the tumors, which is consistent with enrichment of subsets of shared mutations between cell lines and tumor specimens and, conversely, absence of other mutations detected in the tumors, but not cell lines. The absence of clonal enrichment of these events within the tumor dissected from the animal further predicts that the majority of mutations observed only in the cell lines are passenger events that do not endow selective advantage in autochthonous tumors, at least within the number of cell divisions these tumors have undergone by the time of animal necropsy. We also did not observe a predilection for specific base transitions or transversions in the tumors or cell lines (Fig. S2).

We compared these datasets to available sequencing data from human lung adenocarcinoma (Fig. 2E). We observed significantly fewer nonsynonymous mutations in GEMM models than in either smoker- or nonsmoker-associated human lung adenocarcinomas (*Kras* GEMM, 0.07 mutations per Mb; nonsmoker-associated, 1.97 mutations per Mb; $P < 0.0001$; and smoker associated, 7.76

mutations per Mb). Even assuming that the true cancer cell mutation frequency reflects that observed in cell lines (0.25 mutations per Mb), the mutational burden in the mouse models remains significantly lower than that observed in human lung cancers. The relatively low number of mutations observed in murine tumors might reflect a lower number of cell divisions that the cells of origin and tumors have undergone relative to human cancers, a difference in mutation rate per cell division, or a combination of these. We cannot discriminate between these possibilities with available datasets because the number of cell divisions in murine tumors was neither quantified nor estimated in this study.

We identified independent recurrent mutations in a number of genes, including *C5ar1*, *Dnahc5*, *Nyap2*, *Pcdh15*, *Pclo*, *Rngtt*, *Stil*, *Tenm4*, and *Xirp2*. Seven genes (*Csmd1*, *Hmcn1*, *Kmt2c*, *Pcdh15*, *Pclo*, *Ttn*, and *Xirp2*) mutated in the mouse *Kras*^{G12D} cell lines and tumors were also recurrently mutated in >15% of samples in human lung adenocarcinomas (LUADs), as reported by The Cancer Genome Atlas (TCGA). Therefore, *Pcdh15*, *Pclo*, *Ttn*, and *Xirp2* were recurrently mutated in both the mouse model and human lung adenocarcinoma (4) (Fig. 3, denoted with an asterisk). However, mutations in *PCDH15*, *PCLO*, *TTN*, and *XIRP2* were evenly distributed across the coding sequence of these genes, rather than clustered as hot spot events in human cancers (www.cbioportal.org). Therefore, we refrain from speculating that these mutations act as driver alterations, considering ambiguous evidence for a functional role of each of these genes in human cancer.

As one example, *Xirp2* encodes an actin-binding protein implicated in the maintenance of inner ear hair cell stereocilia and cardiac myocyte remodeling. *Xirp2* was mutated in two

independent primary tumors and one related primary tumor-metastasis cell line pair, the latter of which harbored the same *Xirp2* mutation within a highly conserved region of the Xin actin binding repeats (33–35). *Xirp2* has no known role in cancer, and despite 21% of human lung adenocarcinomas in the TCGA study harboring mutations in *XIRP2*, it was not considered a “significantly mutated” gene (4). Review of RNA-seq data from the TCGA study revealed very low expression of *XIRP2* mRNA in human lung adenocarcinoma, suggesting that mutations in *XIRP2* may be passenger events, despite their high frequency. Furthermore, *XIRP2* is primarily expressed in muscle tissues, and it is not known to be expressed in lung tissue (36). In addition, *XIRP2* mutations observed in human lung adenocarcinoma are distributed across the entire coding sequence and the human and mouse *XIRP2* genes are large, encoding peptides >3,500 amino acids in length. Theoretically, it is possible that *XIRP2* mutations might have been selected at an early stage of tumorigenesis, but may not have been advantageous during outgrowth of the dominant tumor subclone before clinical detection. However, the preponderance of evidence suggests that mutations in *Xirp2* are passenger alterations, although the mechanism behind frequent mutation of this gene in human and mouse cancers remains undefined.

We also observed recurrent mutation of *Pclo*, which has been shown to be important for axonal guidance during central nervous system development. *PCLO*, which was mutated in 21% of lung adenocarcinomas in the TCGA study, was also recently identified as a recurrently mutated gene in liver cancers exhibiting a biliary phenotype (37). Knockdown of *PCLO* RNA

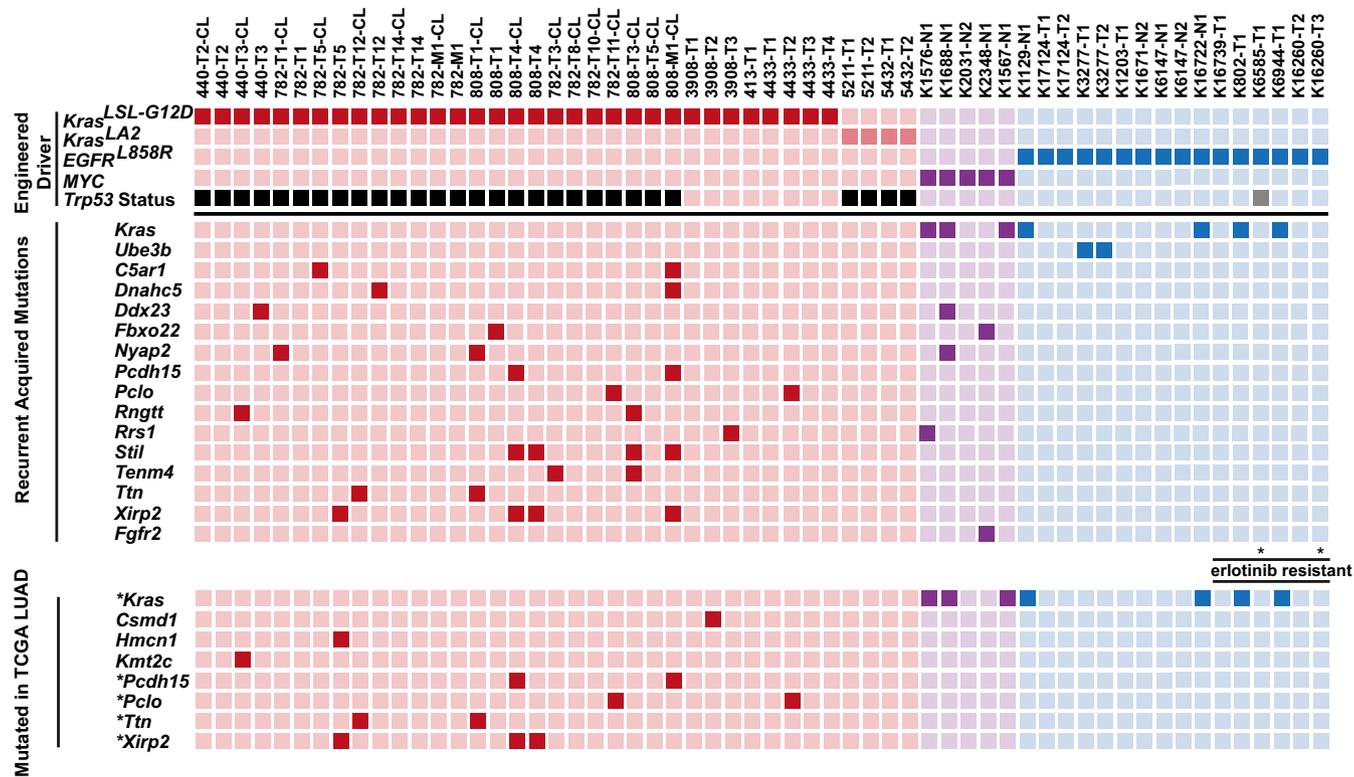


Fig. 3. Mutational landscape of oncogene-induced mouse lung adenocarcinomas. Schematic diagram of genes mutated in the mouse lung adenocarcinoma and cell lines is shown. (Top) *Kras*-, *EGFR*-, and *MYC*-induced tumors are indicated and shaded in red, blue, and purple, respectively. The *Trp53* status is indicated (black, null; gray, heterozygous). (Middle) Genes mutated in two or more samples are indicated. (Bottom) Genes mutated in the murine tumors that are also mutated in >15% of lung adenocarcinomas analyzed in the TCGA. Note that *Csmd1*, *Hmcn1*, and *Kmt2c* were not recurrently mutated in murine tumor, whereas asterisk-indicated genes were recurrently mutated in human and murine tumors. Erlotinib-resistant tumors are indicated. Stars are used to highlight tumors harboring an *EGFR* T790M mutation (by conventional Sanger sequencing).

in human liver cancer cells led to an increase in cell migration. We identified two mutations in *Pclo*, a nonsense mutation (E577X) and E1850K; the latter resides in a conserved region of the protein with unknown function. Further investigation of the role of *Pclo* alterations, especially in the context of *Kras* mutations, seems warranted.

Manual review of mutations occurring in a single tumor revealed mutations in several genes encoding regulators of transcription, including several factors involved in chromatin modification and regulation. Among these are mutations in *Brd4* (H965P, within a proline-rich domain of unknown function), *Chd7* (a nonsense mutation, R977X), *Chd8* (H2198R), *Mill3* (T1798S), *Mlxip* (V453G), *Smarca1* (M27R), *Smyd4* (H769R), and *Tet1* (C1784X) (Table S4). None of these specific mutations has been identified in human cancers. Therefore, it is difficult to determine whether any of these mutations represent driver events. However, mutations in many epigenetic regulators in human tumors are not clustered into hot spots, so it is premature to conclude that these represent passenger mutations in the mouse model.

EGFR-Driven Model of Lung Adenocarcinoma. Tetracycline-inducible expression of the EGFR^{L858R} mutant in the lung epithelium of transgenic mice leads to the formation of lung adenocarcinomas with bronchioalveolar carcinoma features that are sensitive to treatment with EGFR tyrosine kinase inhibitors (TKIs) like erlotinib (10). Long-term intermittent dosing of the mice with erlotinib leads to the emergence of drug-resistant tumors that harbor some of the molecular features of TKI-resistant human tumors, including a secondary mutation in EGFR, EGFR^{T790M} (12). To determine the mutational load of these tumors, compared with tumors initiated by alternate oncogenes, and to seek genetic differences between untreated and erlotinib-resistant tumors, we performed whole-exome sequencing of DNA from 10 TKI-naïve and 6 erlotinib-resistant EGFR^{L858R}-induced mouse lung adenocarcinomas. We observed a lower mutational burden in EGFR mutant lung adenocarcinomas compared with *Kras*-driven tumors (0.02 vs. 0.07 mutations per Mb, $P = 0.002$; Fig. 2A). Erlotinib-resistant tumors exhibited no difference in mutation frequency compared with untreated EGFR mutant tumors (0.02 vs. 0.02, $P = 0.49$; Fig. 2F). The mutational signature present in the EGFR^{L858R}-induced lung adenocarcinomas exhibits a preponderance of C > T transitions, consistent with findings in EGFR mutant human lung adenocarcinomas and all adenocarcinomas from never-smokers (4) (Fig. S2). Interestingly, this result was significantly different from the signature observed in the *Kras* mutant tumors ($P = 0.0008$), although the mechanistic basis for this observation remains to be determined.

Recurrent mutations in the tumors were found in *Kras* ($n = 4$ total; 2 G12V and 2 Q61R) and *Ube3b* ($n = 2$). We previously described an oncogenic *Kras* mutation in an erlotinib-resistant murine tumor; however, such mutations have not been described in patients (12). Interestingly, two of the *Kras* mutations observed here (with nonreference allele fractions of 0.38–0.55) were found in the 10 tumors not treated with TKIs, suggesting that these mutations can arise during tumor development independent of treatment. A recent report shows that coexpression of mutant EGFR and KRAS in the same human lung tumor cells can be toxic (38). It is possible that detection of mutations in both oncogenes in some untreated tumors indicates that expression of one of the oncogenes has been down-regulated in at least some tumor cells; in other tumors with both mutations, inhibition of the EGFR kinase activity with a TKI may have been a permissive feature.

Ube3b, an E3 ubiquitin ligase, was mutated in two individual tumors from a single untreated mouse. The same variant was found in both tumors, suggesting that the two tumors are clonally related. According to TCGA reports, *Ube3b* is altered in 4% of

human LUADs; however, there is no indication at this point of a functional relationship between EGFR mutations and alterations in *Ube3b*.

Because a major mechanism of resistance to EGFR inhibitors is a secondary mutation in EGFR (EGFR^{T790M}), we examined the whole-exome sequencing data to determine whether we could detect reads corresponding to human EGFR (because the transgene encodes human EGFR). Indeed, we unequivocally detected the EGFR^{T790M} mutation in one erlotinib-resistant tumor that we had previously shown to harbor this mutation. However, we cannot exclude inadequate depth of sequencing of the human EGFR transgene as an explanation of our failure to identify other cases of secondary T790M mutations of EGFR in these tumors, especially because the exon capture probes did not target human EGFR sequences.

MYC Model of Lung Adenocarcinoma. The low mutation burden observed in the *Kras*- and EGFR-induced lung tumors prompted us to hypothesize that tumors induced by strong oncogenic lung drivers might have a lower mutation burden than tumors induced by a less potent lung oncogene. We therefore performed whole-exome sequencing of DNA from five lung adenocarcinomas that arose in a mouse model initiated by overexpression of wild-type human MYC, which has been shown to be a less potent oncogene than mutant KRAS or EGFR in the murine lung (26). MYC-induced tumors also exhibited a low mutation frequency, comparable with that observed in *Kras*-induced mouse lung adenocarcinomas (0.14 vs. 0.07 mutations per Mb, $P = 0.57$; Fig. 2A). The mutations found in the MYC-induced tumors included oncogenic *Kras* mutations in three of five tumors and an oncogenic mutation in *Fgfr2* (Fgfr2 K659M) in one tumor. The *Kras* and *Fgfr2* mutations were independently validated by using Sanger sequencing. Mutations at this residue in *FGFR2* have been shown to activate the intrinsic protein-tyrosine kinase and to cooperate with MYC in tumorigenesis (39, 40). The identification of known cancer driver mutations in four of five MYC-driven tumors is consistent with the suggestion that MYC acts as a less potent tumor initiator than mutant *Kras* or EGFR in the murine lung.

Acquired Whole-Chromosome Copy Number Changes Are Common in Mouse Lung Adenocarcinomas. Given the low point-mutation burden observed in the mouse tumors in our GEMM models, we asked whether the tumors harbored alterations in chromosomal or subchromosomal copy numbers. To examine somatic changes in DNA copy number in the GEMM tumors, we analyzed the datasets from whole-exome sequencing using validated computational methods (41, 42). We first examined the *Trp53* locus, which was anticipated to be deleted in the tumors when Cre recombinase was expressed to initiate tumorigenesis (25). We detected deletion of exons 2–10 in all tumor cell lines derived from those animals, suggesting that the method based on sequence data accurately identified small regions of deletion (Fig. S3).

When these methods were applied to the complete set of exome-sequencing data, we primarily detected putative whole chromosome gains and losses in the murine lung adenocarcinomas (Fig. 4), consistent with prior studies of cancers arising in GEMMs (16, 17). Manual review of putative focal amplifications and deletions revealed that many likely resulted from artifacts or biases in the dataset. In particular, a set of breakpoints exhibiting the same start and stop positions in several tumors matched to the same normal sample (i.e., these tumors were isolated from the same animal). These events likely represented artifacts of variable coverage during sequencing of normal (tail) DNA. Another set of events consisted of regions with nearly balanced amplifications and deletions, which overlapped annotated duplication regions. These events likely represented copy

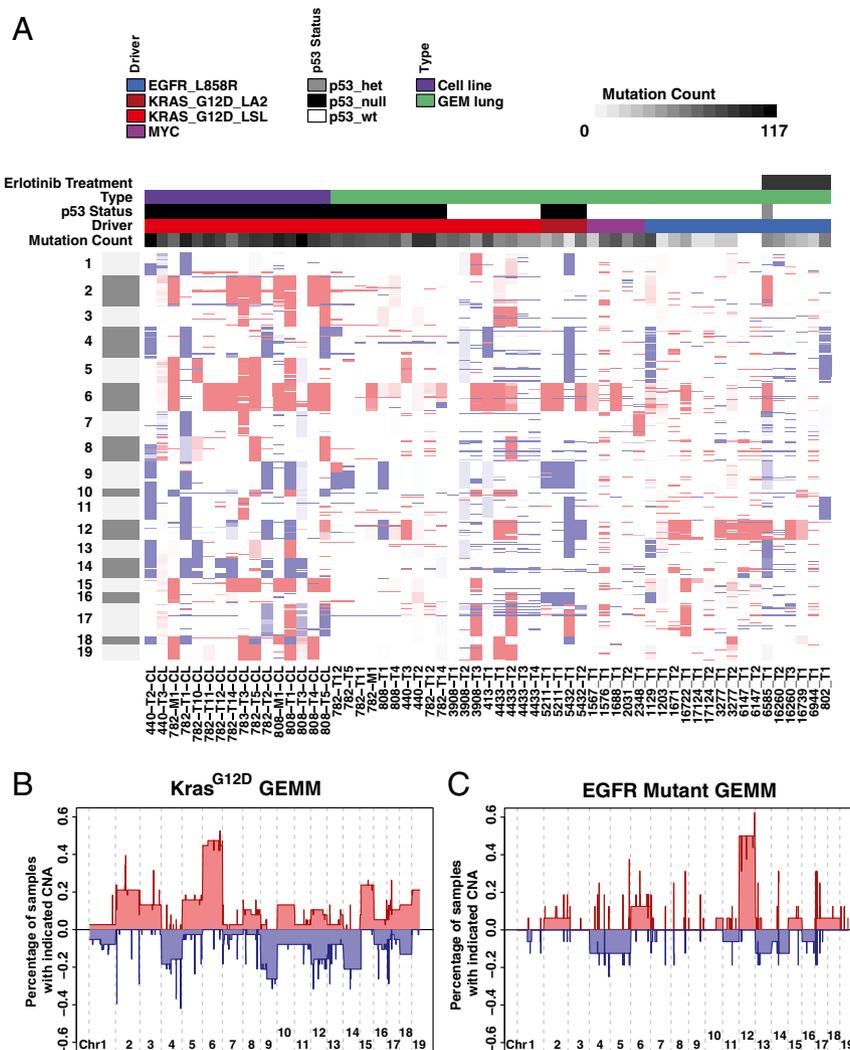


Fig. 4. Distinct patterns of DNA copy number alterations in *Kras*- and *EGFR*-driven GEMMs. (A) Heat map of DNA copy number alterations across all samples. Red, DNA copy number amplification; blue, DNA copy number loss. Models are grouped by initiating driver allele. Also shown is p53 status and whether the sample was a tumor or tumor-derived cell line. Point mutation frequency is shown in gray-black box, with the darker shade representing a higher mutation frequency. (B) Recurrent whole-chromosome DNA copy number gains in *Kras*^{G12D}-driven GEMM tumors and cell lines. (C) Recurrent whole-chromosome DNA copy number gains in *EGFR*-driven GEMM tumors.

number polymorphisms, some of which could be germ-line events (Table S5).

Although the models used in this study all produced one histological tumor type, lung adenocarcinoma, the tumors that develop in each GEMM showed a distinct pattern of recurrent DNA copy number gain or loss (Fig. 4 A–C and Table S5). *Kras*-driven tumors and cell lines harbored recurrent gain of Chr6, consistent with prior studies of *Kras*-induced lung tumors (16, 17, 24, 43). In addition to extra copies of Chr6, we observed recurrent whole-chromosome amplification of Chr2, Chr15, and Chr19 (in $\geq 20\%$ of samples) and whole-chromosome loss of Chr9 and Chr14 (Table S5). These alterations have been observed in murine tumors by other investigators using different methods for copy-number determination, lending additional confidence in our exon capture-based copy number analysis (16, 17, 24). Because Chr6 carries the *Kras* locus, we determined the allelic fraction with the G12D mutation; this analysis suggested that the chromosome with the engineered G12D mutant, not the chromosome with the wild-type allele, was responsible for the gain in chromosome number (Fig. S4). Similarly, the proto-oncogene *Myc* is on Chr15, and gain of copies of Chr15 is the

second most frequent whole-chromosome alteration observed in this model and is consistent with previous work suggesting that *Myc* function may be necessary for tumor maintenance in *Kras*-driven GEMMs (44).

Recurrent whole-chromosome DNA copy number changes appeared to be less frequent in the *EGFR*-driven GEMMs (Fig. 4 and Table S6), and a different pattern of changes was observed. For unexplained reasons, an increased number of copies of Chr12 was the most recurrent alteration. It is possible that the unmapped *TetO-EGFR* transgene is integrated in Chr12; gain in copy number might then increase signaling from the *EGFR* oncogene.

We did not observe an anticorrelation between the frequency of point mutations and the fraction of the genome affected by DNA copy number alterations, as previously described (24). This finding might in part be due to the overall low mutational burden observed in these tumors, which is approximately an order of magnitude lower than that observed in carcinogen-induced models (mean of 185 mutations in urethane-induced and 728 in methylnitrosourea-induced tumors) (24). Therefore, it is possible that, when higher mutation loads are present, there is less selection for

large-scale DNA copy number alterations across the genome in these models.

Discussion

Recent improvements in DNA-sequencing technologies have spurred the genomic characterization of many types of human cancers, including lung adenocarcinomas. These datasets have revealed much information about the mutational profiles of this cancer and identified several novel putative oncogenes and tumor suppressors. However, unraveling the complexity of these datasets and distinguishing driver and passenger mutations remain significant challenges, particularly in highly mutated genomes such as smoking-associated lung adenocarcinoma. In contrast, in this work, we have observed a very low mutation burden in EGFR-, *Kras*-, and *MYC*-driven GEMM tumors, regardless of tumor genotype. The low mutation burden in the murine tumors is consistent with prior studies in other GEMMs (24). Together, these findings suggest that the number of mutations necessary for the development and progression of invasive lung adenocarcinomas in mice is small. To conduct these analyses, we developed a pipeline for somatic mutation calling in mouse exome datasets that was adjusted to minimize the calling of false positives. This performance was achieved by testing two different SNP callers on a set of germline polymorphisms that distinguish the inbred mouse strains C57BL/6 and 129sV/J. Through this analysis, we found that the highest sensitivity and lowest false-positive rates were observed when the intersection of both callers was used in the analysis, consistent with recent findings in the literature (30).

We detected recurrent whole-chromosome gains and losses in the EGFR- and *Kras*-driven models. These observations raise the question of whether copy number alterations are contributing to tumorigenesis in these models. Considering that we observed recurrent Chr6 amplification—which encodes the engineered *Kras*^{LSL-G12D} allele as well as several components of the MAPK signaling pathway—in *Kras*-driven models, we speculate that cooperating oncogenes and/or tumor suppressors located in the regions of whole-chromosome gain or loss indeed contribute to tumor progression. However, the identity of the driver events in the amplified and deleted regions remains to be determined. These observations suggest that amplification of the signal from the initiating oncogene may be the most important somatic event in these models to drive tumor progression (43).

Previous work also described changes in gene expression during tumor progression, suggesting that epigenetic alterations contribute to progression in these models (13). The detection of several mutations in transcriptional regulators and chromatin-remodeling factors in our models (Fig. 3) is consistent with these findings, although we have not determined directly whether any of the observed mutations accelerate tumor progression or lead to specific changes in chromatin in these models.

We found evidence for clonal selection during the emergence of drug resistance and during the generation of tumor-derived cell lines. Tumors harvested from mice with EGFR-mutant lung cancer harbored oncogenic lesions known to confer primary or acquired resistance to TKIs (for example, *Kras* mutations and the EGFR T790M mutation, respectively). In addition, in the *Kras*-driven model, tumor cell lines harbored a higher mutation load compared with the parental tumors, suggesting clonal selection during outgrowth of cell lines.

Despite the low frequency of observed somatic events in the GEMM tumors, each model exhibited distinct features. In contrast to the *Kras* and EGFR models, which harbored few mutations known to act as cancer drivers, four of five *MYC*-induced tumors harbored oncogenic mutations in *Kras* or *Fgfr2*. The acquisition of potent driver mutations in these tumors suggests that *MYC* overexpression sensitizes cells to transformation by cooperating with spontaneous *Kras* or *Fgfr* somatic mutations. In contrast, even in the absence of *Trp53*, *Kras*-driven tumors did

not acquire mutations in known tumor suppressors or oncogenes. This finding highlights the potency of this oncogene and strongly suggests that the initiating genetically engineered allele is a critical determinant of acquired events in these models.

The fact that tumor-initiating overexpression of *MYC* required additional spontaneous mutations in *Kras* or *Fgfr* is a strong testament to the centrality of MAPK signaling in lung adenocarcinoma. In addition, the fact that *Kras*-mutant models acquired few other recurrent drivers further suggests that initiating MAPK activation is sufficient for tumorigenesis. The most frequently observed somatic alteration in *Kras*-mutant tumors was amplification of Chr6, which encodes the initiating oncogene and additional signaling components of the MAPK pathway. Although additional acquired driver alterations, or independent tumor-initiating events (e.g., *MYC* expression), may cooperate to drive tumor development, these data underscore the importance of maintaining signaling through the MAPK pathway in lung adenocarcinoma.

The overall nonsynonymous mutation burden in human lung adenocarcinomas is 6.86 mutations per Mb (lung TCGA). This frequency is ~50-fold higher than the median mutation burden observed in any of the mouse lung adenocarcinomas studied here. In part, this finding is likely to reflect the lack of carcinogen exposure. However, the mutation frequency in never-smokers (1.97 mutations per Mb) remains >10-fold higher than that observed in our lung cancer models (4). The rapid development of tumors in mice may also contribute to the reduced complexity of the cancer genome in these models compared with human lung adenocarcinoma. As previously described, the copy number profiles of mouse tumors were generally characterized by large-scale whole-chromosome gains or losses (14, 15, 17, 18). In contrast, human tumors exhibited both large-scale and focal amplifications and deletions, perhaps also reflecting differences in carcinogen exposures in tumors in the two species (45).

Our findings have important implications for the optimal use and further development of GEMMs, particularly considering the ease with which these modifications can be generated by using new genome-editing methods (46–48). Although we have not sequenced to the extreme depth necessary to identify mutations in very small subpopulations of tumor cells, the genomic profile of these models appears to be much less complex than most human cancers. The genomic complexity of lung cancer is at the heart of drug resistance and appears to be an important determinant of the response to immune checkpoint inhibitors (49).

We previously reported that tumors in the *Kras*^{LSL-G12D}; *Trp53*^{fl/fl} model exhibit very modest immune cell infiltrates (50). This finding is consistent with very few neoantigens generated as a result of the very low number of somatic mutations that arise during tumor development as we describe here. However, expression of a strong T-cell antigen in the model induces a potent T-cell response, which is subsequently suppressed at later stages of tumor progression (50). Therefore, it is important to consider the low mutational burden exhibited in tumors in these GEMMs when designing therapeutic studies or studies of drug resistance. At the same time, the uniformity of the programmed somatic mutations and low acquired mutation frequency observed in tumors in these models are important experimental strengths, making the models well suited to reproducible mechanistic studies and genetic screening. Efforts to model genomic complexity in GEMMs, by using mutagens, transposons, or engineered loss of DNA repair pathways, are approaches to further optimize GEMMs for studies of sensitivity and resistance to therapies and could identify new drivers of progression and metastasis that cooperate with the initiating engineered mutations.

Methods

Mouse Models. All animal studies were performed under approved Institutional Animal Care and Use Committee protocols at the Massachusetts Institute of Technology and Memorial Sloan Kettering Cancer Center. Tumor

induction was performed in *Kras^{LSL-G12D};p53^{fl/fl}* and *Kras^{LSL-G12D}* mice as described with 2.5×10^4 lentivirus particles per animal (31). Tumors were isolated and tumor-derived cell lines were generated as described (13). Histological analysis was performed on a piece of each tumor to assess tumor purity and histological subtype of lung cancer. Tumor-derived cell lines were probed by Southern blotting using a radioactive probe to cre cDNA sequences to confirm a single viral integration and independent origin of tumor cell lines as described (13). Tumors were induced in the *TetO-EGFR^{L858R}* and *TetO-MYC* models by feeding the mice doxycycline-impregnated food as described (10).

Exome Sequencing. DNA was purified from tumor tissue and tumor cell lines by using standard methods. Sonication of 2 μ g of genomic DNA was performed by using a Diagenode Bioruptor, and size selection was performed by using dual selection using AMPure beads as described (51). Exon capture was performed by using Roche SeqCap EZ all-exon mouse kits. Postcapture libraries were sequenced on an Illumina HiSeq instrument.

Variant Calling. The pipeline to call somatic mutations was predominately composed of standard programs that have been used in human tumor analysis, with the addition of a custom caller that was optimized to reduce false positives at the expense of some sensitivity to low-allele frequency events. The final mutation list was the intersection of these two calling algorithms in the hope that artifacts giving rise to false positives in one would be filtered out by the other.

Raw sequence files were first preprocessed to remove the sequencing adapters. Then, clipped reads were mapped to the standard mm9 genome for the *Kras* model or to mm9+transGene (hEGFR or hMYC) hybrid genomes for the EGFR and MYC models. BWA ALN (Version 0.5) was used to make maps. The reads were marked with read groups and sorted, and then duplicates were removed by using the PICARD toolkit. These initial bam files were postprocessed by using the standard GATK packages; indel realignment was followed by base quality recalibration. The postprocessed bam files were then called by using two separate mutation callers: MuTect (Version 1.1.4) in high-confidence mode and a custom caller built around the GATK Unified Genotyper, with a set of filters to improve specificity. An intersection of

these calls were postfiltered for artifacts by removing any events that were in a database of likely germ-line events (see *SI Appendix* for details). Approximately 50 somatic mutation calls were independently validated by PCR and Sanger sequencing, which suggested very high (>95%) accuracy of the calling method. The calls were annotated with Annotvar, and a list of “functional” mutations was created that contained missense, nonsense, and splice-site mutations.

DNA Copy Number. Copy number was determined by first computing normalized log ratios between tumors and matched normals. This analysis was performed by taking the bam files from the mutation-calling pipeline and computing the coverage for each exon target region using bedtools. The raw coverage numbers for each tumor normal pair were normalized with a robust regression method that normalized not only for total depth but also for the local GC content around each target region using the loess function from R. The log (base 2) ratio of T to N was computed from the normalized coverage, and this was then segmented by using the circular binary segmentation method of ref. 27. A postsegmentation normalization was then used to center the diploid peak at $\log R = 0$. To find segments that were either amplified or deleted, we used the RAE algorithm (28), which computes a sample-dependent soft threshold (sigmoid function) for each tumor/normal pair based on the noise of that pair. This algorithm gives a value of 0–1 for both amplifications and deletions, which approximately indicates fractional amount of each alteration. These values were then averaged over all samples to give the fraction of each region that was amplified or deleted in a given set of samples.

ACKNOWLEDGMENTS. We thank Richard Cook and Alla Leshinsky (Biopolymers Core Facility, KI Swanson Biotechnology Center) and the Hope Babette Chang Histology Facility in the Swanson Biotechnology Center. This work was supported by a grant from the Starr Cancer Consortium (to T.J., G.J.H., and H.V.). Additional support was provided by the Howard Hughes Medical Institute (T.J. and G.J.H.); NIH Grants NCI K08 160658 (to D.G.M.), R01CA120247 (to K.P. and H.V.), and R01 CA121210 (to K.P.); Cancer Center Support Grants P30 CA016359 and P30 CA008748; Specialized Program of Research Excellence Grants P50 CA140146-05 and UL1TR000457; and by Uniting Against Lung Cancer (X.S.).

- Govindan R, et al. (2012) Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 150(6):1121–1134.
- Imielinski M, et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150(6):1107–1120.
- Politi K, Herbst RS (2015) Lung cancer in the era of precision medicine. *Clin Cancer Res* 21(10):2213–2220.
- Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511(7511):543–550.
- Ding L, et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* 455(7216):1069–1075.
- Sun S, Schiller JH, Gazdar AF (2007) Lung cancer in never smokers—a different disease. *Nat Rev Cancer* 7(10):778–790.
- Fisher GH, et al. (2001) Induction and apoptotic regression of lung adenocarcinomas by regulation of a K-Ras transgene in the presence and absence of tumor suppressor genes. *Genes Dev* 15(24):3249–3262.
- Jackson EL, et al. (2001) Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev* 15(24):3243–3248.
- Ji H, et al. (2006) The impact of human EGFR kinase domain mutations on lung tumorigenesis and in vivo sensitivity to EGFR-targeted therapies. *Cancer Cell* 9(6):485–495.
- Politi K, et al. (2006) Lung adenocarcinomas induced in mice by mutant EGF receptors found in human lung cancers respond to a tyrosine kinase inhibitor or to down-regulation of the receptors. *Genes Dev* 20(11):1496–1510.
- Oliver TG, et al. (2010) Chronic cisplatin treatment promotes enhanced damage repair and tumor progression in a mouse model of lung cancer. *Genes Dev* 24(8):837–852.
- Politi K, Fan PD, Shen R, Zakowski M, Varmus H (2010) Erlotinib resistance in mouse models of epidermal growth factor receptor-induced lung adenocarcinoma. *Dis Model Mech* 3(1-2):111–119.
- Winslow MM, et al. (2011) Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* 473(7345):101–104.
- Holstege H, et al. (2010) Cross-species comparison of aCGH data from mouse and human BRCA1- and BRCA2-mutated breast cancers. *BMC Cancer* 10:455.
- Maser RS, et al. (2007) Chromosomally unstable mouse tumours have genomic alterations similar to diverse human cancers. *Nature* 447(7147):966–971.
- Sweet-Cordero A, et al. (2006) Comparison of gene expression and DNA copy number changes in a murine model of lung cancer. *Genes Chromosomes Cancer* 45(4):338–348.
- To MD, et al. (2011) Progressive genomic instability in the FVB/*Kras*(LA2) mouse model of lung cancer. *Mol Cancer Res* 9(10):1339–1345.
- Varela I, et al. (2010) Somatic structural rearrangements in genetically engineered mouse mammary tumors. *Genome Biol* 11(10):R100.
- Nassar D, Latil M, Boeckx B, Lambrechts D, Blanpain C (2015) Genomic landscape of carcinogen-induced and genetically induced mouse skin squamous cell carcinoma. *Nat Med* 21(8):946–954.
- McCreery MQ, et al. (2015) Evolution of metastasis revealed by mutational landscapes of chemically induced skin cancers. *Nat Med* 21(12):1514–1520.
- Calbó J, Meuwissen R, van Montfort E, van Tellingen O, Berns A (2005) Genotype-phenotype relationships in a mouse model for human small-cell lung cancer. *Cold Spring Harb Symp Quant Biol* 70:225–232.
- Dooley AL, et al. (2011) Nuclear factor κ B is an oncogene in small cell lung cancer. *Genes Dev* 25(14):1470–1475.
- McFadden DG, et al. (2014) Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* 156(6):1298–1311.
- Westcott PM, et al. (2015) The mutational landscapes of genetic and chemical models of *Kras*-driven lung cancer. *Nature* 517(7535):489–492.
- Jackson EL, et al. (2005) The differential effects of mutant p53 alleles on advanced murine lung cancer. *Cancer Res* 65(22):10280–10288.
- Tran PT, et al. (2008) Combined inactivation of MYC and K-Ras oncogenes reverses tumorigenesis in lung adenocarcinomas and lymphomas. *PLoS One* 3(5):e2125.
- Podsypanina K, Politi K, Beverly LJ, Varmus HE (2008) Oncogene cooperation in tumor maintenance and tumor recurrence in mouse mammary tumors induced by *Myc* and mutant *Kras*. *Proc Natl Acad Sci USA* 105(13):5242–5247.
- Kim SY, Speed TP (2013) Comparing somatic mutation-callers: Beyond Venn diagrams. *BMC Bioinformatics* 14:189.
- Cibulskis K, et al. (2013) Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31(3):213–219.
- Ewing AD, et al.; ICGC-TCGA DREAM Somatic Mutation Calling Challenge participants (2015) Combining tumor genome simulation with crowdsourcing to benchmark somatic single-nucleotide-variant detection. *Nat Methods* 12(7):623–630.
- DuPage M, Dooley AL, Jacks T (2009) Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat Protoc* 4(7):1064–1072.
- Johnson L, et al. (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* 410(6832):1111–1116.
- Francis SP, et al. (2015) A short splice form of Xin-actin binding repeat containing 2 (XIRP2) lacking the Xin repeats is required for maintenance of stereocilia morphology and hearing function. *J Neurosci* 35(5):1999–2014.
- Scheffer DI, et al. (2015) XIRP2, an actin-binding protein essential for inner ear hair-cell stereocilia. *Cell Reports* 10(11):1811–1818.
- Wang Q, et al. (2010) Essential roles of an intercalated disc protein, mXinbeta, in postnatal heart growth and survival. *Circ Res* 106(9):1468–1478.
- Wu C, et al. (2009) BioGPS: An extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol* 10(11):R130.

37. Fujimoto A, et al. (2015) Whole-genome mutational landscape of liver cancers displaying biliary phenotype reveals hepatitis impact and molecular diversity. *Nat Commun* 6:6120.
38. Unni AM, Lockwood WW, Zejnullahu K, Lee-Lin SQ, Varmus H (2015) Evidence that synthetic lethality underlies the mutual exclusivity of oncogenic KRAS and EGFR mutations in lung adenocarcinoma. *eLife* 4:e06907.
39. Ota S, Zhou ZQ, Link JM, Hurlin PJ (2009) The role of senescence and pro-survival signaling in controlling the oncogenic activity of FGFR2 mutants associated with cancer and birth defects. *Hum Mol Genet* 18(14):2609–2621.
40. Tchaicha JH, et al. (2014) Kinase domain activation of FGFR2 yields high-grade lung adenocarcinoma sensitive to a Pan-FGFR inhibitor in a mouse model of NSCLC. *Cancer Res* 74(17):4676–4684.
41. Olshen AB, Venkatraman ES, Lucito R, Wigler M (2004) Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* 5(4):557–572.
42. Taylor BS, et al. (2008) Functional copy-number alterations in cancer. *PLoS One* 3(9):e3179.
43. Feldser DM, et al. (2010) Stage-specific sensitivity to p53 restoration during lung cancer progression. *Nature* 468(7323):572–575.
44. Soucek L, et al. (2013) Inhibition of Myc family proteins eradicates KRas-driven lung cancer in mice. *Genes Dev* 27(5):504–513.
45. Weir BA, et al. (2007) Characterizing the cancer genome in lung adenocarcinoma. *Nature* 450(7171):893–898.
46. Maddalo D, et al. (2014) In vivo engineering of oncogenic chromosomal rearrangements with the CRISPR/Cas9 system. *Nature* 516(7531):423–427.
47. Sánchez-Rivera FJ, Jacks T (2015) Applications of the CRISPR-Cas9 system in cancer biology. *Nat Rev Cancer* 15(7):387–395.
48. Sánchez-Rivera FJ, et al. (2014) Rapid modelling of cooperating genetic events in cancer through somatic genome editing. *Nature* 516(7531):428–431.
49. Rizvi NA, et al. (2015) Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348(6230):124–128.
50. DuPage M, et al. (2011) Endogenous T cell responses to antigens expressed in lung adenocarcinomas delay malignant tumor progression. *Cancer Cell* 19(1):72–85.
51. Rodrigue S, et al. (2010) Unlocking short read sequencing for metagenomics. *PLoS One* 5(7):e11840.