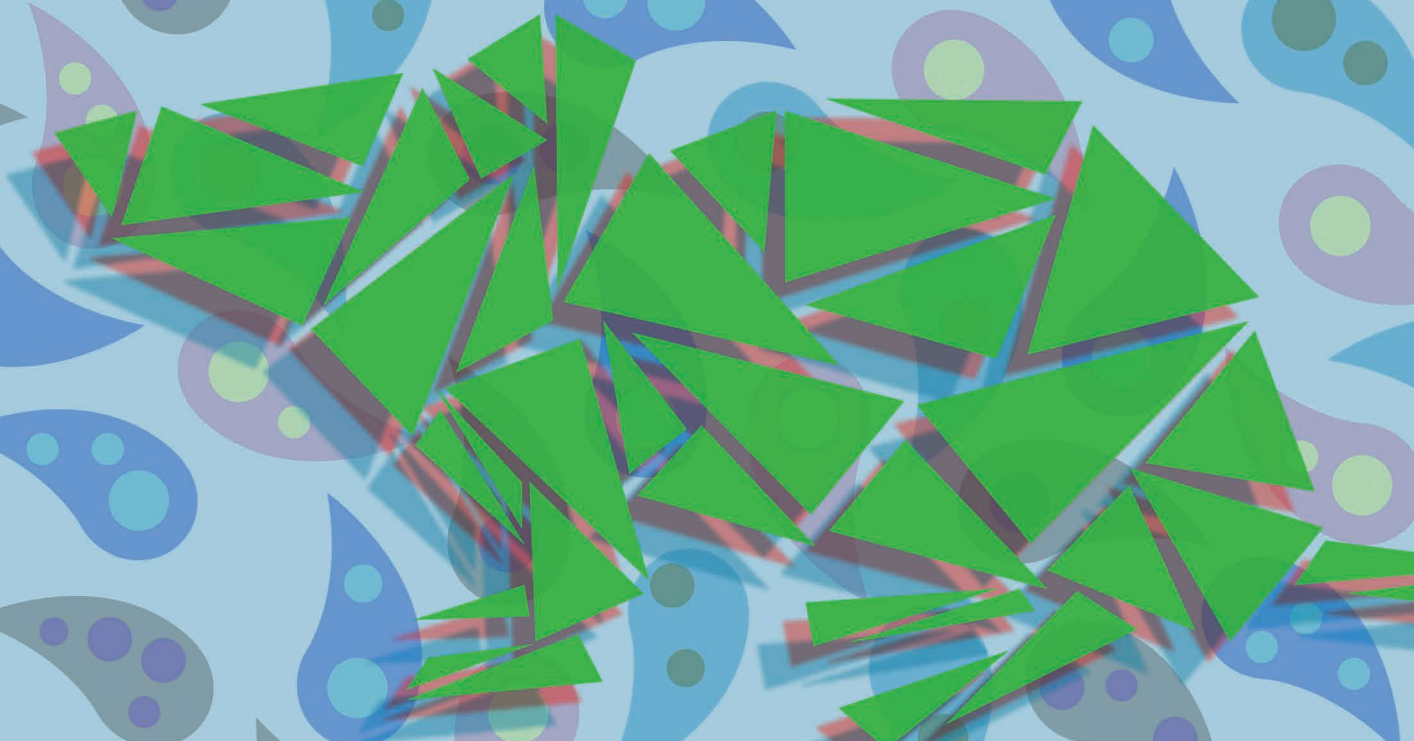
Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

RESEARCH ARTICLE

A Functional Taxonomy of Tumor Suppression in Oncogenic KRAS–Driven Lung Cancer

Hongchen Cai[1], Su Kit Chew[2], Chuan Li[3], Min K. Tsai[1], Laura Andrejka[1], Christopher W. Murray[4], Nicholas W. Hughes[1], Emily G. Shuldiner[3], Emily L. Ashkin[4], Rui Tang[1], King L. Hung[4], Leo C. Chen[1], Shi Ya C. Lee[2], Maryam Yousefi[1], Wen-Yang Lin[1], Christian A. Kunder[5], Le Cong[1,5], Christopher D. McFarland[3], Dmitri A. Petrov[3,4], Charles Swanton[2,6], and Monte M. Winslow[1,4,5]



Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

ABSTRACT Cancer genotyping has identified a large number of putative tumor suppressor genes. Carcinogenesis is a multistep process, but the importance and specific roles of many of these genes during tumor initiation, growth, and progression remain unknown. Here we use a multiplexed mouse model of oncogenic KRAS–driven lung cancer to quantify the impact of 48 known and putative tumor suppressor genes on diverse aspects of carcinogenesis at an unprecedented scale and resolution. We uncover many previously understudied functional tumor suppressors that constrain cancer *in vivo*. Inactivation of some genes substantially increased growth, whereas the inactivation of others increases tumor initiation and/or the emergence of exceptionally large tumors. These functional *in vivo* analyses revealed an unexpectedly complex landscape of tumor suppression that has implications for understanding cancer evolution, interpreting clinical cancer genome sequencing data, and directing approaches to limit tumor initiation and progression.

SIGNIFICANCE: Our high-throughput and high-resolution analysis of tumor suppression uncovered novel genetic determinants of oncogenic KRAS–driven lung cancer initiation, overall growth, and exceptional growth. This taxonomy is consistent with changing constraints during the life history of cancer and highlights the value of quantitative *in vivo* genetic analyses in autochthonous cancer models.

INTRODUCTION

Cancer initiation and development is a multistep process driven in large part by cancer cell–intrinsic alterations (1). Over the past several decades, cancer genome sequencing has contributed to our understanding of the genetic drivers of cancer and identified a large number of putative tumor suppressor genes (2–8). However, genome-sequencing data are insufficient to determine the importance of these genes during various stages of carcinogenesis (9). The nature and frequency of genomic alterations also provide limited insight into the modes of action of putative tumor suppressor genes, underscoring the importance of functional genomics in elucidating gene function (10, 11).

Tumor suppressors regulate many different pathways and cellular processes. Assessing their impact on tumor initiation and each step of cancer development not only distinguishes driver from passenger genes but also highlights different pathways and processes that constrain carcinogenesis across the course of the disease (12, 13). Thus, *in vivo* functional genomic approaches are critical for understanding cancer evolution (14–16), interpreting clinical cancer genome sequencing data (17, 18), and directing precision medicine approaches (19, 20).

In vivo cancer models in which tumor initiation and growth occur entirely within the autochthonous environment are uniquely tractable systems to uncover gene function (21). The integration of CRISPR/Cas9 somatic genome editing into genetically engineered mouse models of human cancer has facilitated the rapid analysis of gene function *in vivo* (22–25). Recently, the combination of somatic CRISPR-based genome editing with tumor barcoding and high-throughput barcode sequencing (Tuba-seq) has greatly increased the scale and precision of these *in vivo* approaches (26, 27). These types of approaches can quantify the impact of many engineered genomic alterations on cancer growth *in vivo* in a multiplexed manner (12, 26–28).

Here we integrate multiple critical advances in our Tuba-seq pipeline and quantify the roles of a broad range of diverse putative tumor suppressors across multiple facets of carcinogenesis. By uncovering the extent to which different tumor suppressors govern tumor initiation, growth, and acquisition of altered phenotypes across time, we uncover an unexpectedly complex taxonomy of tumor suppression across the life history of oncogenic KRAS–driven lung cancer.

RESULTS

Prioritization of Candidate Tumor Suppressor Genes

To characterize the functional landscape of tumor suppression, we selected 48 known and putative tumor suppressor genes to investigate using Tuba-seq in a model of oncogenic KRAS–driven lung cancer (Fig. 1A; Methods). These genes

[1]Department of Genetics, Stanford University School of Medicine, Stanford, California. [2]Cancer Evolution and Genome Instability Laboratory, University College London Cancer Institute, London, United Kingdom. [3]Department of Biology, Stanford University, Stanford, California. [4]Cancer Biology Program, Stanford University School of Medicine, Stanford, California. [5]Department of Pathology, Stanford University School of Medicine, Stanford, California. [6]Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, United Kingdom.

Note: Supplementary data for this article are available at Cancer Discovery Online (http://cancerdiscovery.aacrjournals.org/).

H. Cai, S.K. Chew, and C. Li contributed equally to this article.

Corresponding Authors: Monte M. Winslow, Stanford University School of Medicine, 279 Campus Drive, Beckman Center B256, Stanford, CA 94305. Phone: 650-725-8696; Fax: 650-725-1534; E-mail: mwinslow@stanford.edu; Charles Swanton, Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, 1 Midland Road, London, NW1 1AT, United Kingdom. Phone: 44-203-796-2047; E-mail: Charles.Swanton@crick.ac.uk; and Dmitri A. Petrov, Biology Department, Stanford University, Bass Biology Building, 327 Campus Drive, Stanford, CA 94305. Phone: 650-736-1169; Fax: 650-736-6132; E-mail: dpetrov@stanford.edu

Cancer Discov 2021;11:1754–73

doi: 10.1158/2159-8290.CD-20-1325

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

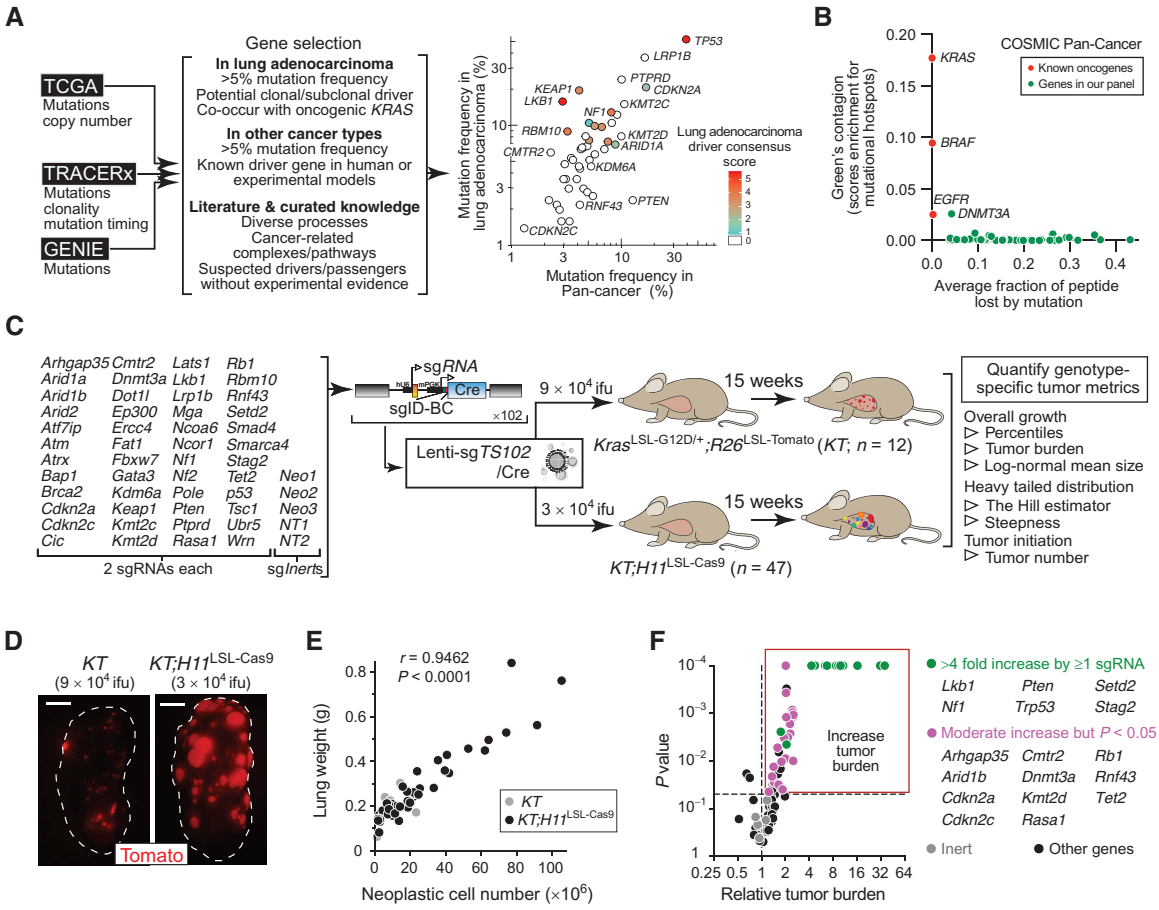


Figure 1. An *in vivo* screen for tumor suppressor genes in autochthonous oncogenic KRAS–driven lung tumors. **A,** Candidate tumor suppressor genes were chosen based on multiple criteria, including their frequency and known/predicted biological functions. The plot shows the mutation frequencies of these 48 genes across pan-cancer and in lung adenocarcinoma (data from TCGA). Color denotes lung adenocarcinoma driver consensus score derived from multiple prediction tools. Several genes that are mutated at high frequency in lung adenocarcinoma or pan-cancer are labeled. **B,** Features of the mutations in each gene are consistent with tumor suppressor function. Green's contagion is a measure of mutational hotspots, which characterize oncogenes. Larger values indicate that mutations are enriched in particular residues of the protein. This measure of overdispersion is normalized to not scale with sample size and to be zero when mutations are randomly scattered across the transcript. Average fraction of protein lost by mutation combines the nonsense/frameshift mutation rate and location of the mutations in each gene [(percent of protein transcript altering mutations that are nonsense or frameshift) * (Average fraction of protein lost by nonsense or frameshift mutations)]. **C,** Schematic of tumor initiation with our pool of 102 barcoded Lenti-sgRNA/Cre vectors (Lenti-sg*TS102*/Cre). Each gene is targeted with two sgRNAs, except *Trp53*, which is targeted by three sgRNAs. Five inert sgRNAs are either nontargeting (NT) or have an active targeting but inert sgRNAs (which target *Neo*^{R} in the *R26*$^{LSL\text{-}Tomato}$ allele). Barcoded lentiviral vectors contain an sgRNA, Cre, and a two-component barcode that includes an sgRNA identifier (sgID) and random barcode (BC). This allows inactivation of multiple target genes in parallel followed by quantification of the number of neoplastic cells by high-throughput sgID-BC sequencing. Mouse genotype, mouse number, and titer of lentiviral vectors are indicated. Tuba-seq was performed on each tumor-bearing lung 15 weeks after initiation, followed by analyses to quantify the indicated metrics. ifu, infectious units. **D,** Fluorescence images of lungs from representative mice at 15 weeks after tumor initiation. Lung lobes are outlined with a dashed white line. Scale bars, 2 mm. **E,** Pearson correlation coefficient (*r*) and *P* value (two-tailed) suggest strong correlation between neoplastic cell number (an indicator of tumor burden) and lung weight. Each dot represents a mouse. When taking into account that tumors were initiated in *KT;H11*$^{LSL\text{-}Cas9}$ mice with 3-fold less Lenti-sg*TS102*/Cre vectors, the total neoplastic cell number is ~10-fold greater in *KT;H11*$^{LSL\text{-}Cas9}$ mice than in *KT* mice. **F,** Volcano plot of the impact of inactivating each putative tumor suppressor gene on relative tumor burden. Each dot represents an sgRNA. Inert sgRNAs are in gray. Tumor suppressor genes are colored pink when both sgRNAs induce a moderate but significant increase and green when one sgRNA induces a >4-fold increase and the other induces a moderate but significant increase. Data are aggregated from 47 *KT;H11*$^{LSL\text{-}Cas9}$ and 12 *KT* mice.

were chosen based on multiple criteria, including their mutational frequency in lung adenocarcinoma from The Cancer Genome Atlas (TCGA), GENIE, and TRACERx data sets; their mutational frequency in pan-cancer genomic data; and the consistency of their mutational profiles with tumor suppressor activity (Fig. 1A and B; Supplementary Fig. S1A–S1E; Supplementary Table S1; refs. 2, 4–7). We also considered their putative tumor-suppressive function in other cancer types as well as their molecular functions (Supplementary Fig. S2A and S2B; refs. 8, 29, 30). Our candidate genes vary greatly in their mutation frequency and co-occurrence with oncogenic *KRAS* alterations (Supplementary Fig. S1C–S1E). Importantly, these genes include well-studied tumor suppressors as well as genes for which there is very limited evidence supporting a role in constraining any aspect of carcinogenesis (Supplementary Fig. S3A and S3B).

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

Quantitative Analysis Uncovers Diverse Tumor Suppressors with Distinct Abilities to Constrain Tumor Growth *In Vivo*

To determine the impact of inactivating each candidate tumor suppressor gene on carcinogenesis *in vivo*, we used Tuba-seq to quantify the tumor size profiles after inactivation of each gene (Supplementary Fig. S4A). We generated at least two Lenti-single-guide RNA (sgRNA)/Cre vectors with distinct sgRNAs targeting each gene and five Lenti-sg*Inert*/Cre negative control vectors (102 total vectors; Fig. 1C; Supplementary Table S2). Each vector contains a two-component sgID-BC, where the sgID uniquely identifies the sgRNA and the diverse random 20-nucleotide barcode (BC) uniquely labels each clonal tumor. We generated each lentiviral vector separately and pooled them to generate a highly multiplexed vector pool (Lenti-sg*TS102*/Cre; Fig. 1C; Methods). We initiated lung tumors with this pool in *Kras*$^{LSL\text{-}G12D/+}$;*R26*$^{LSL\text{-}Tom}$;*H11*$^{LSL\text{-}Cas9}$ (*KT;H11*$^{LSL\text{-}Cas9}$) mice and Cas9-negative control *Kras*$^{LSL\text{-}G12D/+}$;*R26*$^{LSL\text{-}Tom}$ (*KT*) mice. These Cas9-negative mice are necessary to confirm that all vectors have little impact on tumor growth in the absence of Cas9 and to calculate genotype-specific effects on tumor number (see below). Fifteen weeks after tumor initiation, *KT;H11*$^{LSL\text{-}Cas9}$ mice had visibly larger tumors than *KT* mice (Fig. 1D). We extracted DNA from bulk tumor–bearing lungs and used Tuba-seq to quantify overall tumor burden and the sizes of each tumor, of each genotype, in each mouse.

KT;H11$^{LSL\text{-}Cas9}$ mice had ~10-fold higher total neoplastic cell number and proportionally increased lung weight (Fig. 1E). Initial analysis of the impact of each sgRNA on tumor burden (a metric of the relative number of neoplastic cells in all tumors of the same sgRNA) highlighted many genes as functional tumor suppressors. Even this relatively crude metric, which does not incorporate the per-tumor resolution of Tuba-seq, uncovered genes in which both sgRNAs increased tumor burden (Fig. 1F). To investigate which aspects of carcinogenesis are regulated by putative tumor suppressor genes, we calculated multiple summary statistics. We applied our experimental design to identify tumor suppressor genes that normally limit overall tumor growth, tumor initiation, and the emergence of exceptionally large tumors (Fig. 1C; Supplementary Fig. S4B and S4C; Methods).

Many Diverse Tumor Suppressor Genes Increased Overall Tumor Growth

The ability of Tuba-seq to quantify the number of neoplastic cells in thousands of tumors of each genotype allowed us to precisely assess their impact on tumor growth with greater precision than previous approaches. We calculated two metrics of tumor growth from the distribution of tumor sizes to uncover the effect of inactivating each tumor suppressor on overall tumor growth (tumor sizes at defined percentiles within the tumor size distribution and log-normal mean; Methods; Supplementary Fig. S4B). As expected, tumors initiated with each Lenti-sgRNA/Cre vector in control Cas9-negative *KT* mice had very similar tumor size profiles, suggesting that our pipeline is free from bias and false-positive signals (Supplementary Fig. S5A). Consistent with previous Cre/*lox* and CRISPR/Cas9-based mouse models (22, 26, 31–34), inactivation of *Stk11/Lkb1, Pten, Setd2*, and *Nf1* in tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice greatly increased tumor growth (Fig. 2A–C; Supplementary Fig. S5B). Importantly, inactivation of STAG2, a cohesin complex component, increased tumor growth to a comparable extent as inactivation of those well-established tumor suppressors (Fig. 2A–C; Supplementary Fig. S5B).

Inactivation of 14 other genes, including *Cdkn2c, Cmtr2, Rb1, Rnf43, Tsc1*, and *Rbm10*, significantly increased tumor growth (Fig. 2A–C; Supplementary Fig. S5). These 14 genes include not only well-established tumor suppressors, such as *Rb1* and *Cdkn2a*, but also many genes that have not been previously considered functional tumor suppressors in lung adenocarcinoma or cancer in general. For example, the effects of inactivating *Cmtr2* and *Rnf43* were particularly dramatic and unexpected (Fig. 2B). CMTR2 is the sole cap2 2′-O-ribose methylase that modifies the 5′-cap of mRNAs and small nuclear RNAs and is mutated in ~2.2% of lung adenocarcinomas and 1.4% of all cancers (7, 35) (Supplementary Table S1). No previous studies have investigated its function in cancer, and no commercial or academic cancer gene sequencing panels include *CMTR2* (Supplementary Fig. S3A and S3B). RNF43 is a transmembrane E3 ubiquitin ligase that targets WNT receptors for lysosomal degradation (36). *RNF43* is frequently mutated across multiple cancer types, including in colorectal and pancreatic adenocarcinoma, in which *RNF43* deficiency has been shown to sensitize cancer cells to porcupine inhibitors (37, 38). Thus, our broad survey pinpointed multiple novel functional tumor suppressors in oncogenic KRAS–driven lung cancer and revealed commonality among cancer subtypes.

STAG2 Is a Functional Lung Tumor Suppressor

From our initial analysis of overall tumor growth suppression, STAG2 emerged as a particularly interesting and novel suppressor of lung tumor growth. *STAG2* is mutated in ~4% of lung adenocarcinomas, and cohesin complex components are altered in ~10% of lung adenocarcinomas (Supplementary Fig. S6A and S6B; Supplementary Table S1). *STAG2* has been implicated as a tumor suppressor in bladder cancer, regulates lineage-specific genes in acute myeloid leukemia, and is mutated across diverse cancer types (39–42). However, no previous studies have suggested STAG2 as a critical suppressor of lung cancer growth. To further investigate the tumor-suppressive effect of STAG2, we initiated lung tumors in *KT* and *KT;H11*$^{LSL\text{-}Cas9}$ mice with individual Lenti-sg*Inert*/Cre and Lenti-sg*Stag2*/Cre vectors (Supplementary Fig. S7A). Relative to control cohorts, *Stag2* inactivation dramatically increased tumor burden (Supplementary Fig. S7B–S7E). Inactivation of *Stag2* in lung tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice also significantly reduced survival, consistent with its tumor growth–suppressive function (Supplementary Fig. S7F).

To further characterize STAG2-mediated lung tumor growth suppression, we assessed tumor growth in *KT* mice with Cre/*lox*-mediated inactivation of *Stag2* (Fig. 3A). *Stag2* is located on the X chromosome; thus, both heterozygous and homozygous *Stag2* deletion in female mice and hemizygous *Stag2* deletion in male mice generated tumors that lacked STAG2 protein (Fig. 3B and C). *Stag2* inactivation dramatically increased lung tumor burden, and mice with *Stag2*-deficient tumors had markedly shorter overall survival (Fig. 3D–G). *Stag2*-deficient and *Stag2*-proficient lung tumors were atypical adenomatous hyperplasias, adenomas, and early adenocarcinomas

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

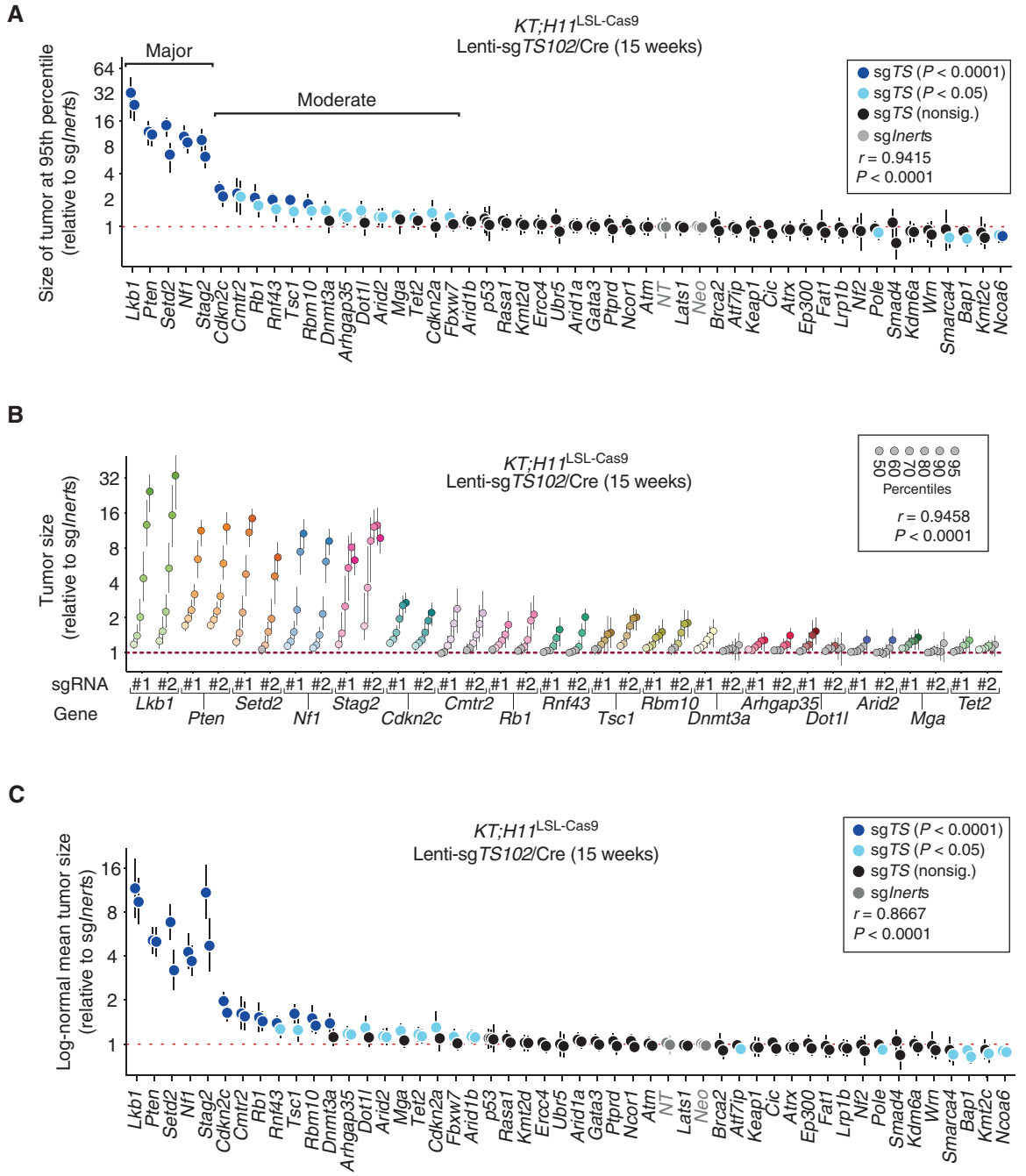


Figure 2. *In vivo* lung tumor growth is suppressed by diverse tumor suppressor genes. **A,** The 95th percentile tumor size (normalized to tumors with sg*Inert*s) for each putative tumor suppressor targeting sgRNA in *KT;H11*$^{LSL\text{-}Cas9}$ mice. Error bars indicate 95% confidence intervals. The 95% confidence intervals and *P* values were calculated by bootstrapping. sgRNAs that significantly increase or decrease tumor size are colored as indicated. sg*Inert*s are in gray and the dotted line indicates no effect. Genes are ordered based on the average of the 95th percentile tumor sizes from all sgRNAs targeting that gene, with individual sgRNAs targeting each gene ranked by effect for clarity. Pearson correlation coefficient (*r*) and *P* value (two-tailed) suggest that sgRNAs targeting the same putative tumor suppressor elicit consistent and similar changes in size at the 95th percentile. **B,** Tumor sizes at the indicated percentiles for the top 17 tumor suppressor genes (relative to the average of sg*Inert*-containing tumors) in *KT;H11*$^{LSL\text{-}Cas9}$ mice. Error bars indicate 95% confidence intervals. Dotted line indicates no effect. Percentiles that are significantly different from the average of sg*Inert*s are in color. Data for all genes are shown in Supplementary Fig. S5B. Pearson correlation coefficient (*r*) and *P* value (two-tailed) for all sgRNA across all indicated percentiles are shown. **C,** The log-normal mean tumor size (normalized to tumors with sg*Inert*s) for each putative tumor suppressor targeting sgRNA in *KT;H11*$^{LSL\text{-}Cas9}$ mice. Error bars indicate 95% confidence intervals. The 95% confidence intervals and *P* values were calculated by bootstrapping. sgRNAs that significantly increase or decrease tumor size are colored as indicated. sg*Inert*s are in gray and the dotted line indicates no effect. Genes and sgRNAs are ordered as in **A**. The high Pearson correlation coefficient suggests that sgRNAs targeting the same putative tumor suppressor elicit consistent and similar changes in log-normal mean tumor size. All plots represent aggregated data from 47 *KT;H11*$^{LSL\text{-}Cas9}$ mice.

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

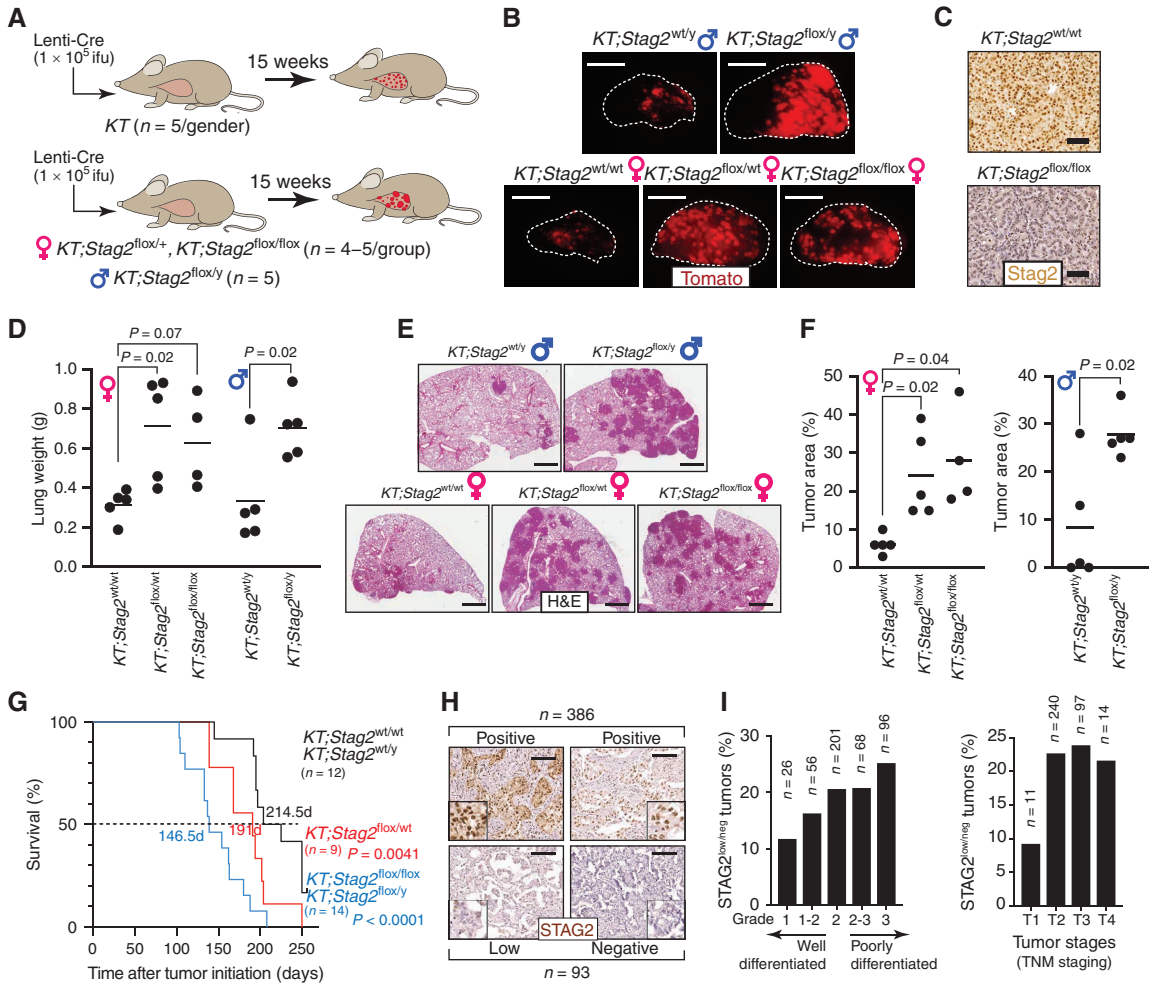


Figure 3. *Stag2*, inactivation of which increases tumor burden and reduces survival, is frequently low-expressed in human lung adenocarcinoma. **A,** Cre/*lox*-mediated *Stag2* inactivation promotes KRAS^{G12D}-driven lung tumor growth. Lung tumors were initiated in the indicated genotypes of mice with Lenti-Cre and allowed to grow for 15 weeks. **B,** Representative fluorescence images of lung lobes from the indicated genotypes and genders of mice are shown. Scale bars, 5 mm. **C,** Lenti-Cre initiated tumors in *KT;Stag2*$^{flox/flox}$ mice lack STAG2 protein. Scale bar, 50 μm. **D,** Lung weight from indicated genotypes of mice 15 weeks after tumor initiation with Lenti-Cre. Each dot represents a mouse and the bar is the mean. *P* values were calculated by Student *t* test. **E,** Inactivation of *Stag2* increases lung tumor growth *in vivo*. Representative histology is shown. Genotype and gender are indicated. Scale bars, 1 mm. **F,** Quantification of tumor area (%) (tumor area/total lung area × 100) on hematoxylin and eosin–stained sections of mouse lungs 15 weeks after tumor initiation. Each dot represents a mouse and the bar is the mean. *P* values were calculated by Student *t* test. **G,** Survival curve of mice with KRAS^{G12D}-driven lung tumors that are Stag2 wild-type (*KT;Stag2*$^{wt/wt}$ female and *KT;Stag2*$^{wt/y}$ male mice), Stag2 heterozygous (*KT;Stag2*$^{flox/wt}$), or Stag2 deficient (*KT;Stag2*$^{flox/flox}$ female and *KT;Stag2*$^{flox/y}$ male mice). Mouse number, *P* value, and median survival (in days) are indicated. *P* values were calculated by comparing each cohort to the *Stag2* wild-type cohort (Mantel–Haenszel test). **H,** Representative STAG2 IHC on human lung adenocarcinomas expressing high (positive) or low (low and negative) STAG2 protein. Scale bars, 100 μm. **I,** Quantification of STAG2 expression in 479 human lung adenocarcinomas. Data are grouped by tumor grade (left, with lower grade indicating well-differentiated tumors and higher grade indicating poorly differentiated tumors) or by tumor stage (right, classified by TNM staging system). A higher percentage of *Stag2*$^{low/neg}$ tumors are poorly differentiated (left) and more advanced (right) tumors.

that were uniformly NKX2-1/TTF1-positive. Interestingly, some *Stag2*-deficient tumors had nuclear palisading and were histologically distinct from the tumors that developed in control *KT* mice (Supplementary Fig. S7G–S7I). STAG2 inactivation in other cancer and cell types is associated with chromosomal instability (43, 44), increased DNA damage (45, 46), and activation of MEK/ERK or cGAS/STING signaling (47, 48). However, immunohistochemistry and analysis of canonical target genes suggest that these mechanisms are unlikely to be major drivers of the increased growth in *Stag2*-deficient lung cancer (Supplementary Fig. S8A–S8E). Thus, further work will be necessary to determine the molecular mechanisms of tumor suppression mediated by STAG2.

Finally, to further characterize the expression of STAG2 in lung cancer, we performed immunohistochemistry for STAG2 on 479 human lung adenocarcinomas. About 20% of tumors were low or negative for STAG2 protein, suggesting that an even larger fraction of lung adenocarcinomas may be driven by alterations in this pathway (Fig. 3H). Interestingly, STAG2-low/negative lung adenocarcinomas were often more poorly differentiated and advanced (Fig. 3I).

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

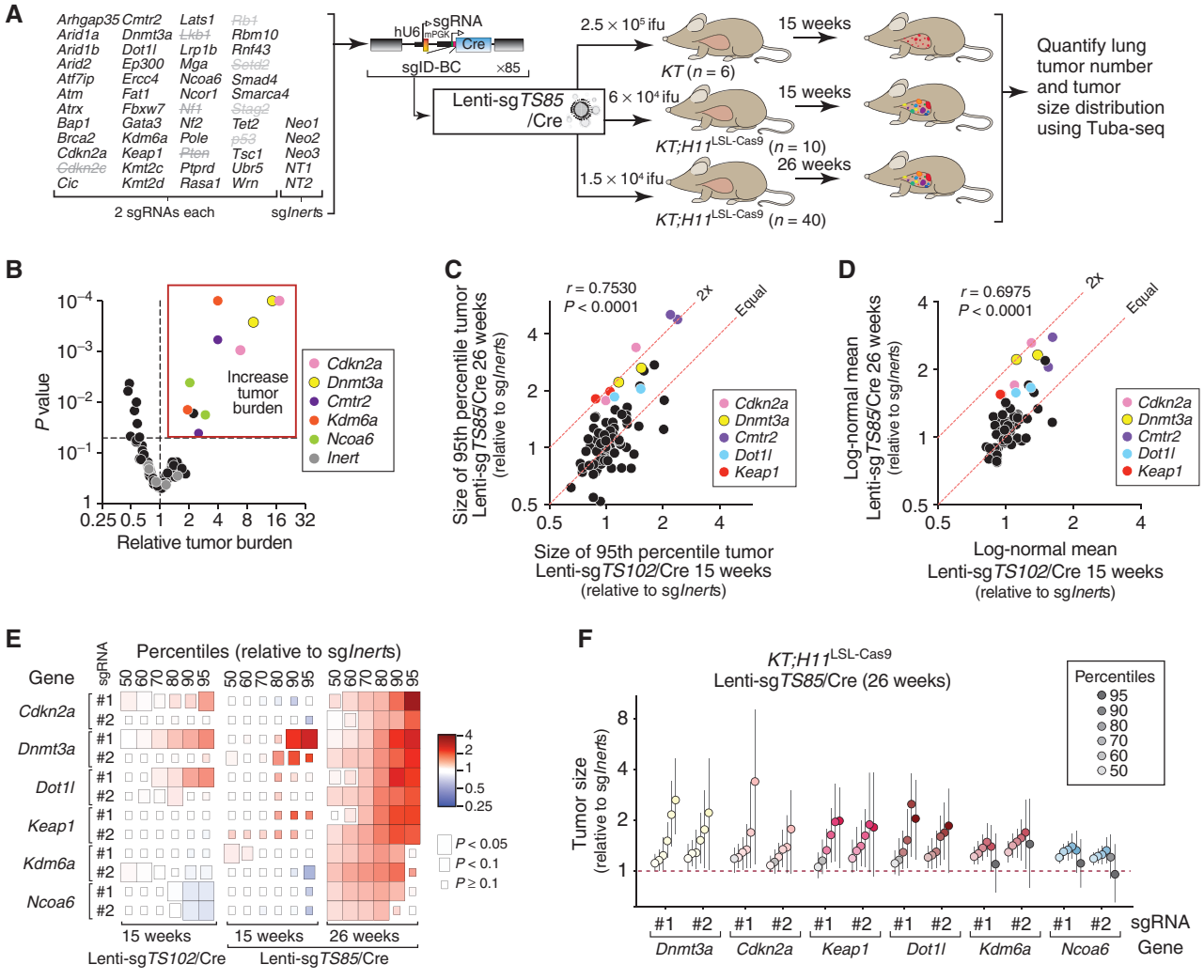


Figure 4. Exaggeration of tumor phenotypes and emergence of more functional tumor suppressors over time. **A,** Schematic of tumor initiation with a pool of 85 barcoded Lenti-sgRNA/Cre vectors (Lenti-sg*TS85*/Cre) that exclude vectors targeting eight tumor suppressor genes (in gray and crossed out) from the Lenti-sg*TS102*/Cre pool that collectively account for ~60% of total tumor burden. Each gene is targeted with two sgRNAs. Mouse genotype, mouse number, and titer of lentiviral vectors delivered to each mouse are indicated. Tuba-seq was performed on each tumor-bearing lung at the indicated time after tumor initiation. **B,** Volcano plot of the impact of inactivating each putative tumor suppressor gene on relative tumor burden. Each dot represents an sgRNA. Genes for which both sgRNA increase tumor burden are colored. **C** and **D,** The impact of inactivating each gene on the size of the 95th percentile tumor (**C**) and log-normal mean (**D**) at 15 weeks (Lenti-sg*TS102*/Cre 15 weeks) and 26 weeks (Lenti-sg*TS85*/Cre 26 weeks) after tumor initiation is shown. Each dot represents an sgRNA. Statistics are calculated from aggregating all tumors from 40 *KT;H11*$^{LSL\text{-}Cas9}$ (26 weeks) and 47 *KT;H11*$^{LSL\text{-}Cas9}$ (15 weeks) mice. **E,** Heat map of the tumor-suppressive effects of six genes that emerge as suppressors of tumor growth at the later time point. Colors indicate the impact of inactivating each gene on tumor size at 15 weeks (Lenti-sg*TS102*/Cre 15 weeks and Lenti-sg*TS85*/Cre 15 weeks) and 26 weeks (Lenti-sg*TS85*/Cre 26 weeks) after tumor initiation, and sizes of the tiles indicate statistical significance levels. **F,** Sizes of tumors at the indicated percentiles for each Lenti-sgRNA/Cre vector relative to that of sg*Inert* tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice. Error bars indicate 95% confidence intervals. Percentiles that are significantly different from the average of sg*Inert*s are in color. Data for all genes are shown in Supplementary Fig. S9B.

Additional Tumor-Suppressive Effects Emerge at Later Time Points

To gain further insights into the dynamics of tumor suppression in lung cancer, we assessed tumor suppressor gene function at a later time point after tumor initiation. We reasoned that allowing tumors to grow for a longer period of time might uncover greater magnitudes of growth suppression for genes that initially had modest effects and could highlight additional tumor suppressors that play more important roles only at later stages of tumor growth. To allow mice to survive for a longer period of time after tumor initiation, we generated a second pool of Lenti-sgRNA/Cre vectors, which excluded those targeting *Lkb1, Pten, Setd2, Nf1, Trp53, Stag2, Cdkn2c,* and *Rb1* that collectively accounted for more than half of the total tumor burden (Lenti-sg*TS85*/Cre; Fig. 4A). We initiated tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice with a titer of Lenti-sg*TS85*/Cre that would allow them to survive for 26 weeks while maximizing tumor number to achieve reasonable statistical power (Fig. 4A; Supplementary Fig. S9A; Methods). As controls, we also initiated tumors with Lenti-sg*TS85*/Cre pool in *KT;H11*$^{LSL\text{-}Cas9}$ and *KT* mice and analyzed them after 15 weeks (Fig. 4A).

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

After 26 weeks of tumor growth, inactivation of *Cdkn2a, Dnmt3a, Cmtr2, Kdm6a*, and *Ncoa6* significantly increased tumor burden (Fig. 4B). Furthermore, inactivation of *Rbm10, Cmtr2, Rnf43*, and *Tsc1* still increased tumor sizes at defined percentiles of the distribution as well as the log-normal mean tumor size at this later time point (Supplementary Fig. S9B). These results confirm the tumor-suppressive function of these genes. Importantly, inactivation of several other genes that had marginal to no effects on tumor sizes after 15 weeks of tumor growth, including *Keap1, Kdm6a, Ncoa6, Cdkn2a, Dnmt3a*, and *Dot1l*, broadly increased tumor sizes after 26 weeks of tumor growth (Fig. 4C–F). Thus, analysis of growth metrics at multiple time points after tumor initiation can provide temporal resolution of tumor suppressor gene effects.

Tuba-seq Captures Additional Aspects of Tumor Suppressor Gene Function

In addition to uncovering tumor suppressor genes that limit overall growth, our methods can quantify other aspects of cancer initiation and progression affected by these genes and pathways. The relative tumor burden induced by each Lenti-sgRNA/Cre vector was mostly consistent with the growth effects uncovered using tumor sizes at defined percentiles (Supplementary Fig. S10A). However, the effects of inactivating some genes on relative tumor burden were disproportionately large (Supplementary Fig. S10A and S10B). For example, *Trp53* was clearly a tumor suppressor based on relative tumor burden, but p53 inactivation did not greatly increase overall tumor growth as assessed by log-normal mean or tumor sizes up to the 95th percentile tumor (Supplementary Fig. S10A). Inactivation of several other genes also had much more significant and dramatic effects on relative tumor burden than on tumor sizes (Supplementary Fig. S10B and S10C). These disproportionate increases in relative tumor burden could result from genotype-specific increases in tumor number and/or the sizes of the very largest tumors, neither of which are captured well by log-normal mean or tumor sizes at defined percentiles of the tumor size distribution.

Many Tumor Suppressors Constrain Tumor Initiation

Our experimental design, in which we initiated tumors in cohorts of *KT;H11*$^{LSL\text{-}Cas9}$ and *KT* mice with the same pool of lentiviral vectors, enabled us for the first time to use Tuba-seq to uncover the impact of each putative tumor suppressor gene on tumor initiation and very early oncogenic KRAS–driven epithelial expansion (Supplementary Fig. S4C and Methods). The genetic alterations that drive the development of very early epithelial expansions are poorly understood, yet these events influence tumor incidence and set the stage for all subsequent events during cancer evolution. *In vivo* mouse models are particularly well suited to study the effects of genetic alterations on these early events.

Fifteen weeks after tumor initiation, inactivation of many genes, including *Lkb1, Setd2*, and *Stag2*, which had some of the most dramatic effects on tumor growth, did not increase tumor number (defined as the number of clonal expansions with more than 200 cells; Fig. 5A; Supplementary Fig. S4C and Methods). However, *Pten* inactivation increased tumor number by ~4-fold, suggesting that at least three fourths of epithelial cells expressing oncogenic KRAS^{G12D} fail to expand beyond 200 cells, if at all (Fig. 5A and B). *Tsc1* inactivation also increases tumor number, albeit to a lesser extent, consistent with TSC1 suppressing mTOR downstream of PI3K (49). Inactivation of *Nf1, Rasa1*, and *Trp53* also increased tumor number, thus implicating several signaling pathways in the earliest stages of lung tumor development (Fig. 5A). Strikingly, inactivation of four members of the COMPASS complex (*Kdm6a, Ncoa6, Kmt2c/Mll4*, and *Kmt2d/Mll3*; refs. 50, 51) all increased tumor number (Fig. 5A). The importance of histone H3K4 methylation mediated by this complex is further substantiated by the mutation of at least one member of this complex in 11.7% to 24.2% of human lung adenocarcinomas (Fig. 5C and D; ref. 2). Importantly, genes that limit tumor initiation and those that constrain tumor growth are often independent, suggesting that these facets of tumor suppression can represent distinct functions (Supplementary Fig. S11A).

Analysis of the effect of each genotype on tumor number in mice with tumors initiated with the Lenti-sg*85*/Cre pool (at both 15 and 26 weeks after tumor initiation) provided us with the opportunity to further validate the effect of tumor suppressor inactivation on tumor initiation and early growth (Fig. 5E; Supplementary Fig. S11B and S11C). The effects of inactivating each tumor suppressor gene on relative tumor number were highly correlated across all three data sets (Fig. 5F; Supplementary Fig. S11D and S11E). Several genes, including *Cdkn2a, Dnmt3a, Kdm6a*, and *Ncoa6*, that initially increased only tumor number also increased overall growth fitness at the later time point. This suggests some link between the cellular changes that enable normal epithelial cells to break through the constraints of early hyperplastic growth and the greater fitness in the resulting tumors (Figs. 4F and 5F; Supplementary Fig. S9B).

Tumor Suppressor Inactivation Allows the Emergence of Rare but Very Large Tumors

Next, we took advantage of the per-tumor resolution of our Tuba-seq data to quantify the impact of inactivating each gene on the generation of exceptionally large tumors. In addition to the effects of tumor suppressor gene inactivation on overall tumor growth and tumor initiation, the development of exceptionally large tumors is suggestive of genotypes that promote or allow additional alterations to drive aggressive tumor growth. We previously found that one major effect of *Trp53* deficiency is the generation of exceptionally large tumors (26, 27). Using metrics such as the Hill estimator (a measure of the heavy-tailedness of a distribution; ref. 52), we quantified the extent to which *Trp53* inactivation enables the emergence of infrequent but exceptionally large tumors after 15 weeks of tumor growth (Fig. 6A and B; Supplementary Fig. S12A). The effect of *Trp53* inactivation is consistent with many previous reports documenting the emergence of large lung tumors in *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{flox/flox}$ mice (32, 53–55). These analyses also showed that inactivation of *Cdkn2a* and the DNA methyltransferase *Dnmt3a* might allow some tumors to grow to disproportionately large sizes (Fig. 6A and B; Supplementary Fig. S12A).

To further investigate the effects of tumor suppressor gene inactivation on the emergence of exceptionally large tumors, we determined which genotypes generate heavy-tailed tumor size distribution after 26 weeks of tumor growth. Analysis

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

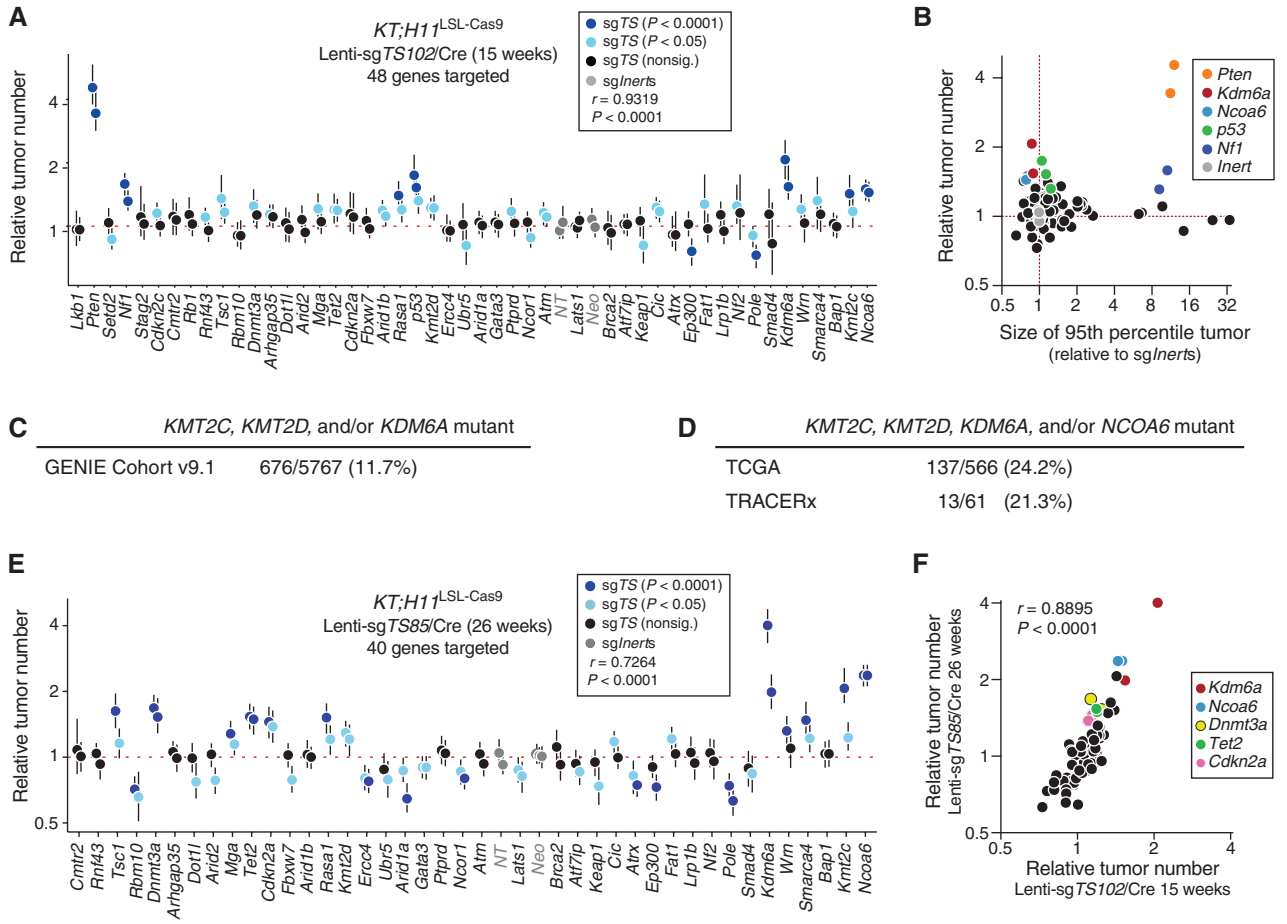


Figure 5. Tumor initiation is inhibited by diverse tumor suppressor genes independent of their effects on tumor growth. **A,** Inactivation of many tumor suppressor genes increases tumor number, highlighting pathways that normally constrain the earliest steps of carcinogenesis. The effect of each sgRNA on tumor number 15 weeks after tumor initiation with Lenti-sg*TS102*/Cre in *KT;H11*$^{LSL\text{-}Cas9}$ mice is shown. Error bars indicate 95% confidence intervals. The 95% confidence intervals and *P* values were calculated by bootstrapping. sgRNAs that significantly increase or decrease tumor number are colored as indicated. sg*Inert*s are in gray and the dotted line indicates no effect. Genes and sgRNAs are ordered as in Fig. 2A. **B,** Genotype-specific effects on growth (represented by the size of the tumor at the 95th percentile) and tumor number can be independent aspects of tumor suppression. **C** and **D,** Mutation frequency of members of the COMPASS complex in human lung adenocarcinoma. Data are shown as the number of patients with mutations in one or more of the COMPASS complex subunits/total patient number from GENIE (**C**) as well as TCGA and TRACERx (**D**). Data from GENIE are based on panel sequencing and therefore do not include data on *NCOA6*. Data from TRACERx are from multiregion sequencing where we report the number of tumors that had any of these four genes mutated in one or more regions. **E,** The effect of each sgRNA on tumor number 26 weeks after tumor initiation with Lenti-sg*TS85*/Cre in *KT;H11*$^{LSL\text{-}Cas9}$ mice is shown. Error bars indicate 95% confidence intervals. The 95% confidence intervals and *P* values were calculated by bootstrapping. sgRNAs that significantly increase or decrease tumor number are colored as indicated. sg*Inert*s are in gray and the dotted line indicates no effect. Genes and sgRNAs are ordered as in **A**. **F,** Effects of tumor suppressor gene inactivation on tumor number are highly reproducible. The impact of inactivating each gene on tumor number at 15 weeks (Lenti-sg*TS102*/Cre 15 weeks) and 26 weeks (Lenti-sg*TS85*/Cre 26 weeks) after tumor initiation is shown. Each dot represents an sgRNA. Statistics are calculated from aggregating all tumors from all mice in each group in each experiment. Pearson correlation coefficient (*r*) shows correlation.

of the distributions of tumor sizes specifically highlighted the development of exceptionally large *Dnmt3a*- and *Cdkn2a*-targeted tumors (Fig. 6C–E; Supplementary Fig. S12B–S12D). Both sgRNAs targeting *Cdkn2a* are anticipated to inactivate both INK4A and ARF; therefore, the effect of *Cdkn2a* inactivation could reflect the combined deregulation of the Rb and p53 pathways, consistent with our observation that p53 inactivation generates a heavy-tailed distribution (Fig. 6A and B; Supplementary Fig. S12A; refs. 26, 27). The emergence of very large *Cdkn2a*- and *Dnmt3a*-deficient tumors is consistent with the increased lung tumor burden in oncogenic *Kras*$^{LSL\text{-}G12D}$-driven tumors with Cre/*lox*-mediated inactivation of these genes (56, 57). However, the per-tumor resolution of our data suggests that the inactivation of INK4A/ARF or the DNA methyltransferase DNMT3A enables the emergence of rare but exceptionally large tumors while having only a modest impact on the growth of most tumors (Fig. 6E; Supplementary Fig. S12C). Therefore, the role of tumor suppressors in preventing the development of exceptionally large tumors can be independent of their roles in regulating tumor initiation and overall growth during cancer evolution.

Limited Effects of Overall Tumor Burden and Sex on Tumor Suppressor Function

Our high-resolution data across multiple facets of tumor suppression in principle allow for quantification of

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

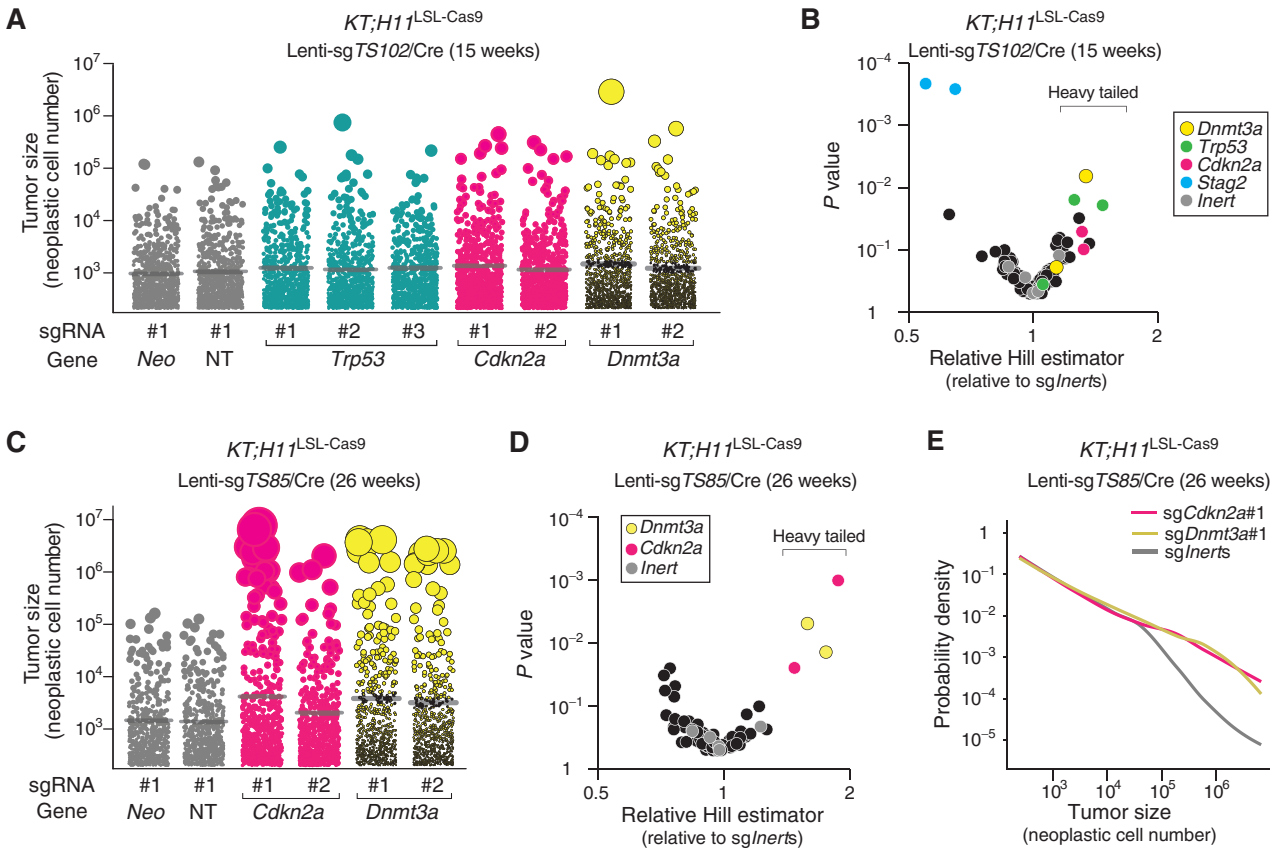


Figure 6. Loss of *Trp53*, *Cdkn2a*, and *Dnmt3a* result in the development of rare yet exceptionally large tumors. **A,** Plot of tumor sizes for each indicated sgRNA in *KT;H11*$^{LSL\text{-}Cas9}$ mice at 15 weeks. Each dot represents a tumor and the area of the dot scales with neoplastic cell number within the tumor. For better visualization, an equal number of tumors ($n = 1{,}160$) are shown for each sgRNA. **B,** Volcano plot of the impact of inactivating each putative tumor suppressor gene on the development of exceptionally large tumors (Hill estimator compares tumors above the 95th percentile with those at the 95th percentile to quantify the relative size of tumors in the tail of the distribution). *Trp53*- and *Dnmt3a*-targeted tumors are heavy-tailed, suggesting that loss of these genes promoted the emergence of exceptionally large tumors. Each dot represents an sgRNA. **C,** Plot of tumor sizes for each indicated sgRNA in *KT;H11*$^{LSL\text{-}Cas9}$ mice at 26 weeks. Each dot indicates a tumor, and the area of the dot indicates neoplastic cell number within the tumor. An equal number of tumors ($n = 814$ tumors randomly sampled) are shown for each sgRNA. **D,** Volcano plot of the impact of inactivating each putative tumor suppressor gene on the development of infrequent exceptionally large tumors (Hill estimator). Each dot represents an sgRNA. Statistics are calculated from aggregating all tumors from 40 *KT;H11*$^{LSL\text{-}Cas9}$ (26 weeks) mice. **E,** Inactivation of *Dnmt3a* and *Cdkn2a* generates tumor size distributions with heavy tails. Probability density plots for tumor sizes show the profile of aggregated tumors with sg*Inert*s as well as individual sgRNAs targeting either *Dnmt3a* or *Cdkn2a*. Data are aggregated from all tumors from 40 *KT;H11*$^{LSL\text{-}Cas9}$ (26 weeks) mice.

the effects of other variables on tumor suppressor effects. Given that overall tumor burden varies across mice and that we initiated tumors in mice of both sexes, we assessed how these variables influence tumor suppressor effects. To uncover whether overall tumor burden influences genotype-specific effects, we divided our *KT;H11*$^{LSL\text{-}Cas9}$ mice with Lenti-sg*TS102*/Cre-initiated tumors into three groups with low, medium, and high tumor burden and reassessed multiple metrics of tumor initiation and growth (Supplementary Fig. S13A). Very few genotype-specific tumor-suppressive effects were influenced by overall tumor burden, suggesting that our results are largely unaffected by potential differences in paracrine or physical interactions that change with tumor density (Supplementary Fig. S13B–S13E).

There is a growing interest in understanding sex-specific effects on all aspects of carcinogenesis. Our data derived from both male and female mice allowed us to investigate sex-specific differences in tumor suppression. Inactivation of most genes, including those on the X chromosome, had similar effects on tumor growth and tumor number in male and female mice (Supplementary Fig. S14A–S14D). Thus, tumor suppressor effects in lung cancer are not dramatically affected by differences in the host environment driven by sex. This was particularly illuminating for *Kdm6a*, which is an X-linked gene that has both H3K27me3 demethylase and nonenzymatic functions (58). Its nonenzymatic function can be compensated for by its paralog UTY on the Y chromosome, and thus different effects in male and female mice have been used to provide insight into the molecular function of KDM6A (58). *Kdm6a* inactivation increased tumor number similarly in male and female mice. The effects were consistent in our data at 15 and 26 weeks after tumor initiation, suggesting that the impact of KDM6A inactivation is most likely driven by loss of its enzymatic function (Supplementary Fig. S14E–S14H).

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

Evaluation of Sensitivity and Specificity

To better estimate the impact of false negatives and false positives on our data, we used all of our data sets to estimate the true-positive rate (Methods). Within all of our data sets, the effects of sgRNAs targeting the same gene were concordant across multiple metrics, consistent with on-target effects (Figs. 2 and 4; Supplementary Fig. S15A–S15F). For instance, in our experiment using Lenti-sg*TS102*/Cre pool, when one sgRNA showed a significant tumor-suppressive effect (nominal $P < 0.05$), the probability to redetect the significant effect using the other guide was above 89% for all metrics assessed (Supplementary Table S3). Thus, the probability that both sgRNAs fail to uncover a functional tumor suppressor that has a similar effect to the tumor suppressors identified in our analysis is below 5% (Supplementary Table S3). Note that for the eight major tumor suppressor genes that were excluded from the Lenti-sg*TS85*/Cre pool, significant effects for both sgRNAs were uncovered in every case. Given these results and the targeting of each putative tumor suppressor gene with two sgRNAs, it is unlikely that functional tumor suppressors were missed for technical reasons. Furthermore, analysis of sgRNA cutting in cells in culture showed comparable efficiency of sgRNAs targeting genes that emerged as tumor suppressors and those that did not (Supplementary Fig. S15G–S15I). Finally, power calculations using our data suggest that an even larger number of genes could be assessed using reasonable numbers of mice using these methods (Supplementary Fig. S16A–S16C).

Human Mutational Data, Cell Line Studies, and *In Vivo* Functional Studies Are Complementary in Defining a Catalog of Tumor Suppressors

The candidate tumor suppressor genes that we assessed were chosen based on existing human mutational data; however, each gene has different levels of correlative data supporting its function as a tumor suppressor (Supplementary Table S1). We explored whether effects on tumorigenesis within the autochthonous environment could be predicted either by human mutation data or through the analysis of human cell lines. Several strong functional tumor suppressors did not stand out based on the human mutational frequency data, and genes such as *STAG2, CMTR2*, and *CDKN2C* were not often predicted to be tumor suppressor genes based on human mutational data (Fig. 2A; Supplementary Fig. S17A–S17G). Thus, computational predictions of tumor suppressor function from mutational data alone (including statistical methods that integrate background mutation rate corrections as well as function- and structure-based impact predictions) nominate some but not all functional tumor suppressors.

Analysis of data from the Dependency Map (59), in which genome-scale knockout screens were performed across diverse cancer cell lines, was also revealing. Inactivation of several top functional tumor suppressors, including *PTEN, CDKN2C, RB1*, and *RNF43*, increased lung adenocarcinoma cell line growth as expected (Supplementary Fig. S17H). However, inactivation of several other major functional tumor suppressors, including *LKB1, SETD2*, and *STAG2*, paradoxically decreased cancer cell growth in culture (Supplementary Fig. S17H). The effects of inactivating several modest tumor suppressors were concordant between the human cell lines and *in vivo* mouse model data, although inactivation of some genes, including *CMTR2, RBM10*, and *KEAP1*, had variable or growth-suppressive effects on cancer cells in culture (Fig. 4B; Supplementary Fig. S17H). Collectively, these results underscore the differences in the fitness landscape in cell lines and indicate that *in vivo* studies can complement these analyses.

DISCUSSION

The enormous genomic diversity in cancer, even within tumors of the same subtype, creates a challenge for identifying driver genes and deciphering their roles in tumor development. Given the sample sizes of cancer genome sequencing studies, variation in genomic features such as location, expression, and composition will continue to make computational predictions of tumor suppressor function from mutation data difficult, except for a subset of genes (9, 60, 61). Moreover, mutation frequencies alone cannot easily define the importance of each tumor suppressor gene and even less so be used to glean their mode of action. Indeed, even rarely mutated tumor suppressor genes can have large consequences when inactivated, with the rarity of mutation being driven by mutational cold spots, epistatic interactions, and biological context (9, 62) rather than by the magnitude of their inhibitory function (Supplementary Fig. S17A). Thus, although experiments using model organisms could be affected by species-specific effects, *in vivo* functional studies that include autochthonous tumor initiation, growth, and progression are an important complement to the computational investigation of tumor suppressor inactivation in human tumors (13, 20, 21).

Carcinogenesis is broadly affected by different aspects of the *in vivo* environment. By enhancing the throughput, sensitivity, and precision of Tuba-seq (26, 27), we quantify the effects of inactivating a diverse panel of putative tumor suppressor genes in an autochthonous mouse model of oncogenic KRAS–driven lung cancer. The parallel analysis of ~50 different genotypes not only uncovered previously uncharacterized functional tumor suppressor genes but also provided new insights into the landscape of tumor suppression and multiple modes of action of tumor suppressor genes (Fig. 7A and B). We show that tumor suppression is unexpectedly complex and multifaceted, with some genes suppressing tumor initiation, some constraining overall tumor growth, and others limiting the emergence of a small proportion of unusually fast-growing tumors (Fig. 7A and B). Furthermore, although some genes affect only a single feature of carcinogenesis, others affect multiple facets of tumor evolution to varying extents (Fig. 7C). The relative importance of these genes can also change during the course of carcinogenesis (Fig. 7B and C). Understanding the impact of tumor suppressors that primarily regulate certain aspects of carcinogenesis may have a unique value for cancer prevention, early detection, and therapeutic targeting. The discovery of such functional complexity points to shifting challenges during different stages of carcinogenesis. Thus, tumor suppressors are not simply "brakes" on proliferation but rather contextually and temporally dependent genetic modifiers of different phases of carcinogenesis.

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

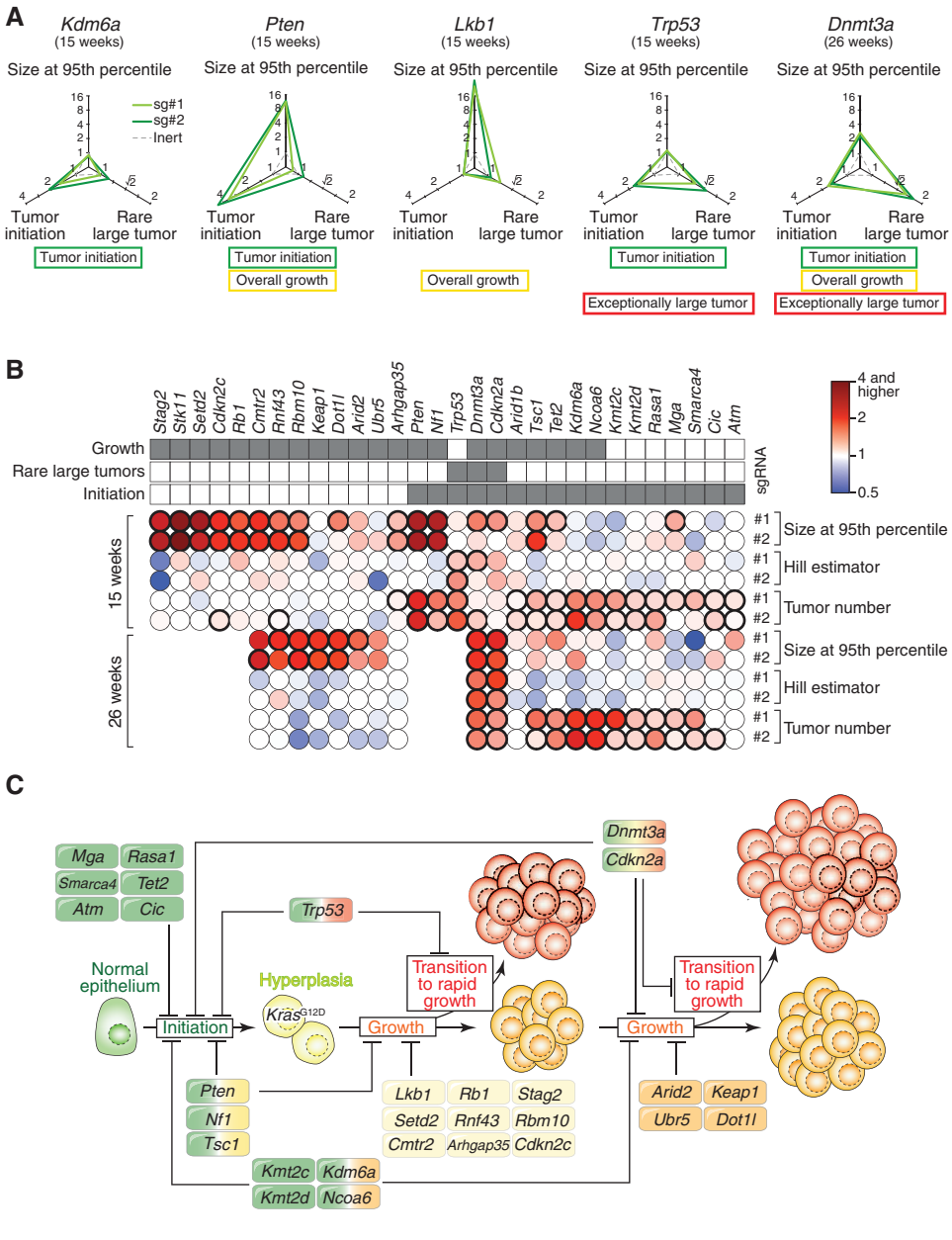


Figure 7. Tumor suppressors constrain tumorigenesis at different stages and to different levels. **A,** Radar plots of representative genes whose inactivation affects tumor size at the 95th percentile (relative to sg*Inert*s, indicating increased overall growth), tumor number (relative to sg*Inert*s, indicating increased tumor initiation), and Hill estimator (relative to sg*Inert*s, indicating increased rare large tumors). Tumor suppressors suppress different aspects of tumor development. **B,** Heat map summarizing the tumor size at the 95th percentile (relative to sg*Inert*s), tumor number (relative to sg*Inert*s), and Hill estimator (relative to sg*Inert*s) of the functional tumor suppressor genes. Color scale is indicated on the side. Bolded circles indicate bootstrap $P < 0.05$. Although the sizes of *Ubr5*-, *Tsc1*-, *Kdm6a*-, and *Ncoa6*-deficient tumors are not significantly different from control tumors at the 95th percentile, they are significantly greater across multiple percentiles at 26 weeks, and thus they are also considered genes that suppress tumor growth. **C,** Summary schematic of a tumor suppression map in lung adenocarcinoma based on our data.

Our results are largely consistent with previous studies that assessed some of these genes individually using similar *in vivo* mouse models of lung cancer (22, 26, 31–34, 51, 63, 64). However, single-gene approaches and quantification of overall tumor burden alone are limited in their ability to uncover the modes of tumor suppression and do not enable direct comparison across many genotypes. For example, although *Lkb1, Pten, Kdm6a, Dnmt3a*, and *Trp53* inactivations each increase overall tumor burden, our quantitative, multiplexed design and computational platform uniquely enabled the deconvolution of different aspects of tumor suppression (Fig. 7A).

We show that the inactivation of many understudied genes has major effects on tumor growth (Fig. 7C; Supplementary Fig. S3). Identifying additional genes that are fundamentally important in suppressing carcinogenesis, including those that are less frequently mutated in human lung adenocarcinoma,

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

can highlight key molecular and cellular processes that are critical in cancer. Furthermore, alterations in *cis*-regulatory elements, epigenetic silencing, and mutations in other members of the same complexes or pathways likely dysregulate these processes in a much higher percentage of tumors. Thus, these types of *in vivo* findings not only suggest the importance of certain genes but also more broadly uncover underappreciated cellular processes that limit cancer development. Our findings nominate several novel genes and key pathways that should be investigated in further mechanistic detail. In particular, the mechanisms by which STAG2 inactivation drives lung cancer growth remain to be elucidated.

One key approach used to implicate the context dependency of tumor suppressor function is the analysis of mutual exclusivity in human data (65). Interestingly, our data demonstrate that genes that trend toward mutual exclusivity with oncogenic *KRAS* mutations, such as *NF1* and *PTEN*, are still important suppressors of oncogenic KRAS–driven lung cancer (Supplementary Fig. S17B). Such statistical trends toward mutual exclusivity should not be misinterpreted as the lack of tumor-suppressive effect of these genes in oncogenic KRAS–driven lung cancer, and more generally, these types of patterns in mutation data should be interpreted with caution (66). Instead, these patterns likely reflect complex epistatic interactions in which context dependence drives frequencies and mutation spectra (9, 62).

Our data, coupled with human lung adenocarcinoma sequencing studies, provide the most comprehensive map of *in vivo* tumor suppressor gene function for cancer (Fig. 7C). Given the quantitative and cost-effective nature of Tuba-seq, even broader studies of many other genes and combinations of genomic alterations may be warranted. Moreover, studies across different genetic and environmental contexts may further elucidate and refine the modality and context dependence of tumor suppressor gene effects (27, 67, 68). This should lead to a more thorough understanding of the interactions between cell-intrinsic and cell-extrinsic processes that contribute to the etiology and evolution of lung cancer.

METHODS

Selection of Candidate Tumor Suppressor Genes for This Study

To select candidate genes to assess *in vivo* using Tuba-seq (and to complement genomics and cell biology approaches), we generated a highly human-curated panel that integrates many different considerations.

Known lung adenocarcinoma driver tumor suppressor genes at >5% mutational frequency (such as *TP53, LKB1, CDKN2A, KEAP1*) from TCGA, AACR Project GENIE, and TRACERx data sets, which were previously assessed by Tuba-seq, were included as positive controls. We included genes that tend to co-occur with oncogenic *KRAS* mutations and those that do not. We also included genes that have been categorized as tumor suppressor genes in other cancer types with >5% mutational frequency in lung (such as *KDM6A* and *FAT1*), even if they are not predicted to be involved in lung adenocarcinoma (Fig. 1A; Supplementary Fig. S1; Supplementary Table S1).

We also considered the distribution of mutations within genes (Fig. 1B), including low-mutation frequency genes (<5%) that show potential clonal or subclonal bias from the TRACERx data set (Supplementary Table S1), genes with discrepancies in scoring of potential driver activity (Supplementary Fig. S2), and genes that represent biological processes or functions commonly associated with carcinogenesis (Supplementary Fig. S3). From a curated survey of literature, candidate genes that have been discussed as cancer driver genes without much or any functional data were also included (Supplementary Fig. S4).

Analysis of Human Lung Adenocarcinoma Cancer Genome Sequencing Data

Mutation frequencies and other information for the 48-gene panel of putative candidate tumor suppressor genes (TSG) are available from multiple cancer data sets and their analyses in TRACERx (6), GENIE (2), and TCGA (7, 69, 70). Oncogenes are characterized by missense point mutations arising in mutational hotspots. In contrast, TSGs are characterized by protein truncating mutations (nonsense and frameshifts) that are more dispersed across the transcript. Moreover, when nonsense and frameshift mutations arise in oncogenes, they tend to truncate C-terminal domains and occur toward the end of the transcript. To identify putative TSGs, we characterized all genes in this survey by these two genetic features: mutational hotspots and the fraction of protein truncated per mutation. We used all point mutations and short insertion and deletions found within the TCGA lung adenocarcinoma (7) and Catalogue of Somatic Mutations in Cancer (71) databases. The extent of mutational hotspots within a gene was determined using a normalized measure of dispersion (Green's contagion) of the number of missense mutations observed within all five residue rolling windows in each gene: $(\sigma^2/\mu - 1)/(\mu N - 1)$, where μ is the mean number of missense mutations observed within each window, σ^2 is the unbiased estimator of the variance, and N is the number of missense mutations. Green's contagion and the five-residue window size were chosen because they maximized the accuracy of classification of known oncogenes and tumor suppressors. Larger values of Green's contagion suggest that mutations are clumping at a few residues within the protein and that the mutant gene is likely oncogenic. This measure has a value of zero when mutations are randomly dispersed throughout the gene and can be negative when mutations are underdispersed. The fraction of protein truncated per mutation is the mean number of amino acids lost per nonsynonymous mutation. It is calculated by simply averaging the fraction of a transcript lost due to each frameshift and nonsense mutation while assigning a value of zero to all missense mutations in this collective average.

To summarize what has previously been described about the biological functions of the candidate genes, we used driver gene scores from attempts to discover cancer driver genes using multiple approaches, such as weighted consensus across multiple tools (8) and prediction by machine learning (29). We also collated the known biological processes and subcellular localization of the 48 genes from the Gene Ontology database (release date 2019–07–01; ref. 30).

For co-occurrence of mutations in *KRAS* and each selected gene, the odds ratio [equals $(N_{\text{neither were mutated}} * N_{\text{Both were mutated}})/(N_{\text{only KRAS is mutated}} * N_{\text{only selected gene is mutated}})$] and P value (one-sided Fisher exact test) were available on cBioPortal.org. In total, 566 lung adenocarcinoma cases from TCGA Pan-Cancer Atlas and 8,522 lung adenocarcinoma samples from GENIE Cohort v8 were analyzed. Note that *NCOA6, ATF7IP, CMTR2*, and *UBR5* are not profiled in any GENIE lung adenocarcinoma cases and hence were excluded from the analysis. For the fitting of a simple linear regression between measured phenotypes and observed clinical parameters, we used data from mutation timing and clonality in lung adenocarcinomas that have been previously described (6, 70).

Analysis of Publications Suggesting Tumor-Suppressive Function of Each Putative TSG in Lung Cancer

A list of articles related to the gene was accessed through the "Bibliography" section of NCBI Gene (https://www.ncbi.nlm.nih.gov/gene/). Subsequently, "lung cancer" and/or "tumor suppressor" were used as the keywords to refine the search.

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

Calculation of Gene Inclusion in Gene-Sequencing Panels

GENIE panel sequencing information was compiled through the GENIE 6.1 Public Release. We first generated a list of panels that provided data from patients with "cancer type detailed" listed as "lung adenocarcinoma," "lung adenocarcinoma in situ," or "lung adenosquamous carcinoma" by filtering the data_clinical_sample.txt file. Then, by parsing the genie_combined.bed file, we generated a list of "screened" genes for each panel, which refers to genes that have "Feature_Type" listed as "exon" and "includeInPanel" listed as "True." This list was then used to categorize our pool of tumor suppressors as either "screened" or "unscreened" by these sequencing panels. Stanford Solid Tumor Actionable Mutation Panel (STAMP) and FoundationOne CDx sequencing panels were obtained from the official websites.

Design, Generation, Barcoding, and Production of Lentiviral Vectors

The sgRNA sequences targeting the putative tumor suppressor genes were designed using Desktop Genetic's Guide Picker (ref. 72; https://genomesunzipped.org/best-genealogy-software/) to prioritize on-target activity (score of >0.6; ref. 73), specificity (score of >0.6; ref. 74), likelihood of generating frameshift indels (score of >0.6; ref. 75), targeting of maximal number of transcript isoforms, no homopolymer runs in the sgRNA, and no extremes in GC content of sgRNA (0.4–0.75), as detailed in Supplementary Table S2.

The Lenti-U6-sgRNA-sgID-barcode-Pgk-Cre vector was modified from our previous work (26) as follows. The sgRNA sequence of the previously described pLenti-sgNT1/Cre (Addgene #66895) vector was replaced with GCGAGGTATTACCGGCGTATCATCCGCG by site-directed mutagenesis to generate pLenti-BaeI-Pgk-Cre. The replacement sequence contains a recognition site for the type IIS restriction endonuclease BaeI, allowing for quick replacement of the sgRNA sequence. To generate each desired vector, forward and reverse single-stranded oligonucleotides containing the sgRNA sequence and complementary overhangs are annealed and ligated into the BaeI-linearized pLenti-BaeI-Pgk-Cre vector using T4 DNA ligase. The barcode oligo primer contains the 8-nucleotide sgID sequence and 20-nucleotide degenerate barcode (Supplementary Table S2). The generation of the barcode fragment and subsequent ligation into the vectors were performed as previously described (26).

Lenti-sgRNA/Cre vectors were individually cotransfected into 293T cells with pCMV-VSV-G (Addgene #8454) envelope plasmid and pCMV-dR8.2 dvpr (Addgene #8455) packaging plasmid using polyethylenimine. Supernatants were collected at 48 and 72 hours after transfection, filtered through a 0.45-μm syringe filter unit (Millipore SLHP033RB) to remove cells and debris, concentrated by ultracentrifugation (25,000 × *g* for 1.5 hours at 4°C), and resuspended in PBS. Each virus was titered against a standard of known titer using LSL-YFP mouse embryonic fibroblasts (MEF; a gift from Dr. Alejandro Sweet-Cordero/UCSF). These MEFs and 293T cells were regularly tested with the MycoAlert *Mycoplasma* detection kit (Lonza, cat. LT07–418) to make sure that they were free of *Mycoplasma*. All lentiviral vector aliquots were stored at −80°C and thawed and pooled at equal ratios immediately prior to delivery to mice.

Mice and Tumor Initiation

The use of mice for the current study has been approved by Institutional Animal Care and Use Committee at Stanford University, protocol number 26696.

Kras$^{LSL\text{-}G12D/+}$ (RRID:IMSR_JAX:008179), *R26*$^{LSL\text{-}tdTomato}$ (RRID:IMSR_JAX:007909), and *H11*$^{LSL\text{-}Cas9}$ (RRID:IMSR_JAX:027632) mice have been previously described (24, 76, 77). They were on a C57BL/6:129 mixed background. The *Stag2*$^{tm1c(EUCOMM)Wtsi}$/J (*Stag2*^{flox}) mice were initially generated by Viny and colleagues (42) and obtained from the Jackson Laboratory (RRID:IMSR_JAX:030902). Tumors were initiated by intratracheal delivery of 60 μL of lentiviral vectors dissolved in PBS.

For the initial experiments, tumors were allowed to develop for 15 weeks after viral delivery of a lentiviral pool that contained 102 barcoded Lenti-sgRNA/Cre vectors (Lenti-sg*TS102*/Cre). Tumors were initiated in *Kras*$^{LSL\text{-}G12D}$;*R26*$^{LSL\text{-}Tom/LSL\text{-}Tom}$ (*KT*) mice with 9×10^4 infectious units (ifu)/mouse of the Lenti-sg*TS102*/Cre pool (12 mice analyzed at 15 weeks after tumor initiation) and in *KT;H11*$^{LSL\text{-}Cas9/LSL\text{-}Cas9}$ mice with 3×10^4 ifu/mouse of the Lenti-sg*TS102*/Cre pool (47 mice analyzed at 15 weeks after tumor initiation).

After the detection of the top functional tumor suppressors after 15 weeks of tumor development, tumors were initiated in additional mice using a subpool of 85 Lenti-sgRNA/Cre vectors (Lenti-sg*TS85*/Cre), which excluded the vectors targeting *Cdkn2c, Lkb1, Nf1, Trp53, Pten, Rb1, Setd2*, and *Stag2*. Tumors were initiated in *KT* mice with 2.5×10^5 ifu/mouse (6 mice analyzed at 15 weeks after tumor initiation), *KT;H11*$^{LSL\text{-}Cas9}$ mice with 6×10^4 ifu/mouse (10 mice analyzed at 15 weeks after tumor initiation), and *KT;H11*$^{LSL\text{-}Cas9}$ mice with 1.5×10^4 ifu/mouse (40 mice analyzed at 26 weeks after tumor initiation).

For the validation experiments using Lenti-sgRNA/Cre-mediated gene inactivation, tumors were allowed to develop for 15 weeks after viral delivery. Tumors were initiated with individual barcoded Lenti-sgRNA/Cre vectors in *KT* mice with 1×10^5 ifu/mouse (3 mice per vector analyzed at 15 weeks after tumor initiation) and *KT;H11*$^{LSL\text{-}Cas9}$ mice with 1×10^5 ifu/mouse (5–6 mice per vector analyzed at 15 weeks after tumor initiation).

For the survival experiments using Lenti-sgRNA/Cre-mediated gene inactivation, tumors were allowed to develop until humane endpoints. Tumors were initiated in *KT;H11*$^{LSL\text{-}Cas9}$ mice with individual barcoded Lenti-sg*Inert*/Cre vectors at 2×10^4 ifu/mouse and with individual barcoded Lenti-sg*Stag2*/Cre vectors at 1×10^4 ifu/mouse (7 mice per vector analyzed).

For *Stag2* validation experiments using the *Stag2*^{floxed} allele, tumors were initiated with Lenti-sg*Inert*/Cre in *KT, KT;Stag2*$^{flox/+}$, *KT;Stag2*$^{flox/flox}$, and *KT;Stag2*$^{flox/y}$ mice with 1×10^5 ifu/mouse (4–5 mice per group analyzed) and allowed to develop for 15 weeks, as well as *KT, KT;Stag2*$^{flox/+}$, *KT;Stag2*$^{flox/flox}$, and *KT;Stag2*$^{flox/y}$ mice with 1×10^5 ifu/mouse (6–7 mice per genotype analyzed) and allowed to develop until humane endpoints.

Tuba-seq Library Generation

Genomic DNA was isolated from bulk tumor–bearing lung tissue from each mouse as previously described (26). Briefly, benchmark control cell lines were generated from LSL-YFP MEFs transduced by a barcoded Lenti-sgNT3/Cre vector (NT3: an inert sgRNA with a distinct sgID) and purified by sorting YFP$^+$ cells. For mice initiated with Lenti-sg*TS102*/Cre pool, 12 benchmark control cell lines (3 cell lines of 500,000 cells each, 3 cell lines of 50,000 cells, 3 cell lines of 5,000 cells, and 3 cell lines of 500 cells) were added to each mouse lung sample prior to lysis to enable the calculation of the absolute number of neoplastic cells in each tumor from the number of sgID-BC reads. Because the standard curve was highly linear, we reduced the benchmark controls to three cell lines with 500,000 cells each for the Lenti-sg*TS85*/Cre pool. Following homogenization and overnight protease K digestion, genomic DNA was extracted from the lung lysates using standard phenol-chloroform and ethanol precipitation methods.

Subsequently, Q5 High-Fidelity 2× Master Mix (New England Biolabs, M0494X) was used to amplify the sgID-BC region from 32 μg genomic DNA. The unique dual-indexed primers used were forward: AATGATACGGCGACCACCGAGATCTACAC-8 nucleotides for i5 index-ACACTCTTTCCCTACACGACGCTCTTCCGATCT-6 to 9 random nucleotides for increased diversity-GCGCACGTCTGC CGCGCTG and reverse: CAAGCAGAAGACGGCATACGAGAT-6 nucleotides for i7 index-GTGACTGGAGTTCAGACGTGTGCTCTTC

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

CGATCT-9 to 6 random nucleotides for increased diversity-CAGGT TCTTGCGAACCTCAT. The PCR products were purified with Agencourt AMPure XP beads (Beckman Coulter, A63881) using a double-size selection protocol. The concentration and quality of the purified libraries were determined using Agilent High Sensitivity DNA Kit (Agilent Technologies, 5067–4626) on the Agilent 2100 Bioanalyzer (Agilent Technologies, G2939BA). The libraries were pooled based on lung weight to ensure even reading depth, cleaned up again using AMPure XP beads, and sequenced (read length 2 × 150 bp) on the Illumina HiSeq 2500 or NextSeq 550 platform (Admera Health Biopharma Services).

Code and Data Availability

Python 3.6 and R 3.6 were used for analyzing the data. The codes are available on GitHub (https://github.com/lichuan199010/functional-taxonomy-of-tumor-suppressors).

The data sets generated and analyzed in the current study are available in the NCBI Gene Expression Omnibus database (token: ezsjeksixhkvbqh; link: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE146302).

Process Paired-End Reads to Identify the sgID and Barcode

The FASTQ files were parsed to identify the sgID and barcode for each read. Each read is expected to contain an eight-nucleotide sgID region followed by a random nucleotide barcode region (**GC**NNNNN**TA**NNNNN**GC**NNNNN**TA**NNNNN**GC**), and each of the 20 Ns represents random nucleotides. The sgID region identifies the putative tumor suppressor gene being targeted, for which we require a perfect match between the sequence in the forward read and one of the 102 sgIDs with known sequences. Note that all sgID sequences differ from each other by at least three nucleotides. Therefore, the incorrect assignment of sgID due to PCR or sequencing error is extremely unlikely. All cells in a clonal expansion from a cell transduced by a lentiviral vector carry the same barcode. To minimize the effects of sequencing errors on calling the barcode, we require the forward and reverse reads to agree completely within the random nucleotide sequence to be further processed. In our pipeline, any "tumor" within a Hamming distance of 2 from a larger tumor is assigned as a "spurious tumor," which likely results from sequencing or PCR errors and is removed from subsequent analysis. Reads with the same sgID and barcode are assigned to the same tumor. The tumor size (number of neoplastic cells) is calculated by normalizing the number of reads from an individual tumor to the number of reads from the benchmark control cell lines added to each sample prior to lung lysis and DNA extraction. The minimum sequencing depth was ~1 read per 43 cells. We have high statistical power in identifying tumors with more than 200 cells, which was used as the minimum cell number cutoff for calling tumors.

Summary Statistics for Overall Growth Rate

Three summary statistics, relative sizes at defined percentiles, relative log-normal mean, and relative tumor burden (will be introduced below), were used to describe the overall tumor growth as previously described. Relative sizes at defined percentiles are nonparametric summary statistics for the tumor size distribution. Specifically, the relative sizes at *X*th percentiles are calculated as the *X*th percentile [*X* represents 50% (median), 60%, 70%, 80%, 90%, and 95%] of the tumor size distribution of sg*TS* tumors divided by the corresponding percentile of the tumor size distribution of all sg*Inert* tumors. This ratio represents the growth advantage at various percentiles conferred by the inactivation of the tumor suppressor gene.

$$\text{Relative size of tumors at } X\text{th percentile} = \frac{\text{Neoplastic cell number at the } X\text{th percentile for sg}TS\text{ tumors}}{\text{Neoplastic cell number at the } X\text{th percentile for sg}Inert\text{ tumors}}.$$

Log-normal mean is the maximum likelihood estimator for the mean number of neoplastic cells for sg*TS* tumors assuming a log-normal distribution of tumor sizes. Similarly, we calculate the relative log-normal mean by dividing the log-normal mean of sg*TS* tumors by the log-normal mean of the sg*Inert* tumors (Supplementary Fig. S4).

$$\text{Relative log-normal mean} = \frac{\text{log-normal mean for } sgTS\text{ tumors}}{\text{log-normal mean for } sgInert\text{ tumors}}.$$

Summary Statistics for Heavy-Tailedness of the Tumor Size Distribution

Some tumor suppressor genes may lead to rare cases of exceptionally large tumors, which results in a tumor size distribution with a heavy tail. We used two summary statistics, relative Hill estimator and relative steepness, to characterize the heavy-tailedness of the tumor size distribution.

The Hill estimator is a commonly used tail index to characterize the tail shape of heavy-tailed distributions (52). Suppose $X_1, X_2, ..., X_n$ are sg*TS* tumor sizes, and we order them by size such that $X_1 \geq X_2 \geq ... \geq X_n$. Let X_k be the tumor size at the 95th percentile, and the Hill estimator is calculated as

$$H = \frac{1}{k}\sum_{i=1}^{k} \ln\left(\frac{X_i}{X_k}\right).$$

The relative Hill estimator is calculated by dividing the Hill estimator for tumors with sg*TS* by that of tumors with sg*Inert*.

$$\text{Relative Hill estimator} = \frac{H\text{ for sg}TS\text{ tumors}}{H\text{ for sg}Inert\text{ tumors}}.$$

The steepness (99th percentile/95th percentile) is calculated as the ratio of the 99th percentile over the 95th percentile for the tumor size distribution for each sgID. Large values of these estimators indicate that the tumor size distributions are heavy-tailed. We calculate the relative steepness by dividing the steepness of tumors with sg*TS* by that of tumors with sg*Inert*.

$$\text{Steepness} = \frac{\text{Number of neoplastic cells at the 99th percentile for } sgTS\text{ tumors}}{\text{Number of neoplastic cells at the 95th percentile for sg}Inert\text{ tumors}}.$$

$$\text{Relative steepness} = \frac{\text{Steepness for sg}TS\text{ tumors}}{\text{Steepness for sg}Inert\text{ tumors}}.$$

For both relative Hill estimator and relative steepness, values higher than 1 indicate that the gene inactivation leads to a heavier tail, and values smaller than 1 indicate gene inactivation leads to a lighter tail than expected (Supplementary Fig. S4).

Summary Statistics for Relative Tumor Number and Relative Tumor Burden

The four metrics above compare the tumor size distribution of sg*TS* tumors relative to sg*Inert* tumors and can be calculated for both *KT;H11*$^{\text{LSL-Cas9}}$ mice and *KT* mice, separately. Unlike these size metrics, relative tumor number and relative tumor burden are affected linearly by lentiviral titer. Therefore, when calculating these two metrics, we normalized it to that in *KT* mice to account for the viral titer differences among different Lenti-sg*RNA*/Cre vectors.

We normalized the observed tumor number for sg*TS* tumors in *KT;H11*$^{\text{LSL-Cas9}}$ mice by dividing it by that of sg*TS* tumors in *KT* mice to account for the titer differences for each sg*TS*.

$$\text{Tumor number} = \frac{\sum\text{tumor number in } KT;H11^{\text{LSL-Cas9}}\text{ mice}}{\sum\text{tumor number in } KT\text{ mice}}\text{ for each sg}TS.$$

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

The relative tumor number is calculated as the ratio of tumor number for each sg*TS* relative to sg*Inert*:

$$\text{Relative tumor number} = \frac{\text{Tumor number for sg}TS\text{ tumors}}{\text{Tumor number for sg}Inert\text{ tumors}}.$$

The relative tumor number is a metric that reflects the probability of tumor initiation. If the tumor suppressor genes normally constrain tumor initiation, inactivation of the gene will increase the relative tumor number to be larger than 1.

Similarly, we normalized the observed tumor burden for sg*TS* tumors in *KT;H11*$^{\text{LSL-Cas9}}$ mice by dividing it by that of sg*TS* tumors. The relative tumor burden is calculated as the ratio of the tumor burden for each sg*TS* relative to sg*Inert*:

$$\text{Tumor burden} = \frac{\sum\text{neoplastic cell number in } KT;H11^{\text{LSL-Cas9}}\text{ mice}}{\sum\text{neoplastic cell number in } KT\text{ mice}}\text{ for each sg}TS.$$

$$\text{Relative tumor burden} = \frac{\text{Tumor burden for sg}TS\text{ tumors}}{\text{Tumor burden for sg}Inert\text{ tumors}}.$$

The relative tumor burden is determined mostly by the largest tumors. For instance, the top 1% of tumor cells accounts for more than 50% of total tumor burden in *KT;H11*$^{\text{LSL-Cas9}}$ mice at 11 weeks. Both *TS* inactivation that leads to faster overall growth, rare but exceptionally large tumors, and tumor initiation rate will result in an increase in relative tumor burden (Supplementary Fig. S4).

Bootstrapping the Tumors

In the calculation of confidence intervals and *P* values, we needed to generate distributions of the statistic considering both variation of tumor sizes across mice and within the same mice. We adopted a two-step bootstrap resampling process. We first bootstrap resampled mice to generate a pseudogroup of mice, and then within each group of resampled mice, we bootstrap resampled all observed tumors carrying each sgID.

Calculation of Confidence Intervals and P Values for Size Metrics

We have four size metrics that describe the overall growth (relative log-normal mean, relative percentiles) and the heavy-tailedness (relative Hill estimator and relative steepness) of the tumor size distribution. For each of these metrics, we bootstrapped tumors 10,000 times and calculated 10,000 values of each statistic for these bootstrap resampling. The 95% confidence interval was calculated as the 2.5th percentile and the 97.5th percentile of these bootstrapped results, whereas the *P* value was calculated as the proportion of bootstrapped results that are not in the same direction as the observed score compared with the baseline of 1.

Calculation of P Values for Tumor Burden and Tumor Number

We bootstrapped tumors in both the *KT;H11*$^{\text{LSL-Cas9}}$ and *KT* mice and calculated the relative tumor burden and relative tumor number from these bootstrapped mice. The process was repeated 10^6 times. The 95% confidence interval was calculated as the 2.5th percentile and the 97.5th percentile of these bootstrapped results, whereas the *P* value was calculated as the proportion of bootstrapped values that are not in the same direction as the observed score compared with the baseline of 1.

Robustness to Tumor Burden Differences

To investigate whether overall tumor burden has an impact on genotype-specific tumor initiation and growth, we calculated summary statistics for tumor initiation and tumor size distribution on groups of mice with different overall tumor burden. Specifically, we divided the 47 *KT;H11*$^{\text{LSL-Cas9}}$ mice with Lenti-sg*TS102*/Cre-initiated tumors at the 15-week time point into three groups based on the total tumor burden in each mouse, namely, the low tumor burden group (16 mice), the medium tumor burden group (16 mice), and the high tumor burden group (15 mice). We performed calculations separately for each group for four metrics (95th percentile tumor size, log-normal mean, tumor burden, and tumor number) and evaluated whether these metrics show any correlation with tumor burden.

Quantification of Sex Differences

For each statistic, we used the ratio to quantify the differences between female mice and male mice. The ratio was calculated as

$$\text{Ratio} = \frac{X_{\text{Female}}}{X_{\text{Male}}},$$

where X_{Male} and X_{Female} are the statistics quantified in male and female mice, respectively. When calculating the *P* values, we respectively bootstrapped tumors in male and female mice and calculated the proportion of times that the bootstrapped results are not in the same direction as the observed score compared with the baseline of 1.

Empirical Estimation of True-Positive Rates

We estimated the power (true-positive rate) for each of the three experiments: (i) Lenti-sg*TS102*/Cre, 15-week experiment; (ii) Lenti-sg*TS85*/Cre, 15-week experiment; and (iii) Lenti-sg*TS85*/Cre, 26-week experiment. Understanding the true-positive rate is important for understanding the probability of identifying functional tumor suppressor genes. Because we do not have a list for genuine functional tumor suppressor genes, we used each sgRNA that generated a significant tumor suppressor effect (with nominal $P < 0.05$) as a proxy for true tumor suppressor effects.

For each experiment, whenever we detected a significant effect for an sgRNA, we queried whether the other sgRNA targeting that same gene also generated a significant tumor suppressive effect. If the other sgRNA showed significant tumor suppressor effect, then the test was counted as true (*T*). If the second sgRNA failed to show a significant tumor suppressor effect, then the test was false (*F*). Across all sgRNA (including sgRNA#1 and sgRNA#2 for each gene), we tallied the number of true and false discoveries. We used additive smoothing by adding a pseudocount of 0.5 to both *T* and *F* counts to avoid the zero-probability problem in some cases. Therefore, the estimated false-negative rate for a gene targeted with a single sgRNA would be

$$p = \frac{F + 0.5}{(T + 0.5) + (F + 0.5)}.$$

The estimated true-positive rate in our experiment was the probability of failing to identify a functional tumor suppressor gene with both of two sgRNAs. Thus, this is

$$\text{False-negative rate} = p^2.$$

$$\text{True-positive rate} = 1 - \text{False-negative rate} = 1 - p^2.$$

We performed this calculation separately for four metrics: 95th percentile, log-normal mean, tumor burden, and tumor number. We did not estimate the true-positive rate for the Hill estimator because the number of positive findings was too few for robust estimations.

In Vitro Analysis of sgRNA Efficiency

To analyze the relative cutting efficiencies of the sgRNAs, we measured the insertion and deletion (indel) rates at the target sites in *Rosa26*$^{\text{LSL-Tomato}}$*;H11*$^{\text{LSL-Cas9}}$ MEFs that were generated from E12.5 embryos. These MEFs tested negative for *Mycoplasma* contamination

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

using the MycoAlert *Mycoplasma* detection kit (Lonza, cat. LT07–418). The 10^5 MEFs were transduced individually with each Lenti-sg*TS*/Cre vector and cultured for 1 week followed by FACS-based isolation of Tomato-positive transduced cells. Genomic DNA was extracted from sorted cells using the QIAamp DNA Micro Kit (Qiagen 56304) and subjected to PCR-based target enrichment. Two rounds of PCR were performed with Q5 Master Mix (NEB #M0494L). The first round amplified each of the 97 sgRNA targeted regions (see Supplementary Table S2 for target-enrichment primer sequences). The second round added unique dual-indexed Illumina sequencing adaptors to the amplicons.

These libraries were sequenced on an Illumina NextSeq 500 in the 2 × 150-bp paired-ended configuration (Admera Health Biopharma Services). The resulting reads were demultiplexed based on their sample indexes. CRISPRessoPooled was used to quantify on-target indel mutations (78). Briefly, pooled reads were initially demultiplexed into files according to their specific sgRNA and aligned to the reference sequence to identify indel mutations. Substitution events were ignored, and all indels that occurred within 10 nucleotides of the predicted target site (3 nucleotides upstream from the NGG PAM) were counted as on-target indel mutations. Indel percent mutated was calculated as the number of reads with an on-target indel divided by the total number of reads.

Histology and Immunohistochemistry

Lung lobes were inflated with PBS/4% paraformaldehyde and fixed for 24 hours, stored in 70% ethanol, and paraffin-embedded and sectioned. Then, 4 μmol/L-thick sections were used for hematoxylin and eosin staining and immunohistochemistry (IHC).

Primary antibodies used for IHC were anti-STAG2 (1:500, LifeSpan cat. LS-B11284, RRID:AB_2725802), anti-NKX2.1 (1:250, Abcam cat. ab76013, RRID:AB_1310784), anti-Phospho-RPA2 (1:400, Abcam cat. ab87277, RRID:AB_1952482), anti-Phospho-Histone H2A.X (1:400, Cell Signaling Technology cat. 9718, RRID:AB_2118009), and anti-Phospho-ERK1/2 (1:400, Cell Signaling Technology cat. 4370, RRID:AB_2315112). IHC was performed using the Avidin/Biotin Blocking Kit (Vector Laboratories, SP-2001), Avidin-Biotin Complex Kit (Vector Laboratories, PK-4001), and DAB Peroxidase Substrate Kit (Vector Laboratories, SK-4100) following the standard protocols. Human lung adenocarcinoma tissue microarrays were purchased from US Biomax (HLugA120PG01, BC041115e, LC1261, LC706a, NSC155, and NSC157).

Whole-Genome Sequencing and Quantitative RT-PCR

For whole-genome sequencing and qRT-PCR–based gene expression analysis, samples were generated from Lenti-Cre–initiated tumors from three *KT* and three *KT;Stag2*$^{flox/flox}$ mice (a subset of samples from Fig. 3G). Briefly, neoplastic cells were isolated from pooled tumors within two lung lobes of each mouse by FACS for Tomato^{positive} lineage (CD45/CD31/F4–80/Ter119)^{negative} cells (79). In total, 60,000 to 100,000 neoplastic cells were collected from each mouse. Genomic DNA and total RNA were purified using the Qiagen AllPrep DNA/RNA Micro Kit (cat. 80284). Genomic DNA was processed with Nextera Flex for karyotyping by low-pass (0.1× coverage) whole-genome sequencing. Log_2 ratio of reads mapping to each genomic locus compared with the average number of reads mapping to all other comparable loci was plotted.

For qRT-PCR, total RNA was reverse-transcribed using the Reliance Select cDNA Synthesis Kit with oligo(dT) primers (BioRad cat. 12012802). Quantitative PCR was performed with PowerUp SYBR Green Master Mix (Thermo Fisher Scientific cat. A25776) on an Applied Biosystems QuantStudio 3 Real-Time PCR System. PCR primers were as follows:

Fos: 5′-TACTACCATTCCCCAGCCGA-3′ and 5′-GCTGTCAC CGTGGGGATAAA-3′
Klf2: 5′-GAGCCTATCTTGCCGTCCTT-3′ and 5′-TTGTTTA GGTCCTCATCCGTG-3′
Ifnl3: 5′-GTGCAGTTCCCACCTCATCT-3′ and 5′-TGGGAGT GAATGTGGCTCAG-3′
Ifnb1: 5′-GTCCTCAACTGCTCTCCACT-3′ and 5′-CATCCAG GCGTAGCTGTTGTA-3′
Mx1: 5′-ACGGTGCAGACATACCAGAA-3′ and 5′-CTGTCTC CCTCTGATACGGT-3′
Ifi44: 5′-ATGGCAGCAAGAAAAGTGCC-3′ and 5′-AAACTTC TGCACACTCGCCT-3′
Irf1: 5′-CCAGAGATTGACAGCCCTCG-3′ and 5′-TGCACAA GGAATGGCCTGAA-3′
Gapdh: 5′-TGTGAACGGATTTGGCCGTA-3′ and 5′-ACTGT GCCGTTGAATTTGCC-3′
Actb: 5′-GGCTCCTAGCACCATGAAGA-3′ and 5′-GTGTAAA ACGCAGCTCAGTAACA-3′

Power Analyses

Power analyses were used to evaluate the ability of the Tuba-seq platform to identify functional tumor suppressors across a variety of experimental scenarios. The likelihood of detecting a tumor suppressor depends on the strength of its effect, the number of mice assayed, and the number of guides in the viral pool. We explored how these parameters influence statistical power to detect genes affecting tumor growth and initiation through a pair of nonparametric nested resampling approaches.

For each simulation that focused on tumor growth, a pseudo-cohort of mice ($n = 5$, 10, 20, 50, 100, 200) was sampled with replacement from the cohort of 47 *KT;H11*$^{LSL\text{-}Cas9}$ mice analyzed 15 weeks after tumor initiation, and statistical significance was assessed by bootstrap resampling of tumors from the pseudo-cohort. For a given viral titer, a larger number of multiplexed vectors result in fewer tumors with each sgRNA and a resulting loss of power due to less thorough sampling of the underlying distribution of tumor sizes. To model this effect, the number of tumors sampled from each mouse was scaled by the ratio of the number of sgIDs in the underlying data to the simulated number of sgIDs ($n = 10$, 20, 50, 100, 200, 500). To capture differences in power due to effect size, we performed analyses for representative strong, moderate, and weak tumor suppressor-targeting sgRNAs (sg*Nf1*#1, sg*Rb1*#1, and sg*Dot1l*#1, respectively). Five hundred simulations were performed for each gene, with a minimum of 16,000 bootstrap samplings per simulation. In each bootstrap, the size of tumor at the 95th percentile with the focal genotype was compared with the size of tumor with sg*Inert*s at the 95th percentile, and significance in each simulation was assessed by bootstrapped P value <0.05 (two-tailed test, Bonferroni corrected for the simulated number of pooled sgRNAs).

Effects on tumor initiation are inferred through changes in the representation of tumor genotypes in *KT;H11*$^{LSL\text{-}Cas9}$ mice relative to the original proportions of the sgRNAs in the lentiviral vector pool. As a result, identifying genes that influence tumor initiation requires comparison of *KT;H11*$^{LSL\text{-}Cas9}$ mice to *KT* mice, in which the relative abundance of genotypes reflects the makeup of the viral pool. For each simulation, we therefore sampled a cohort of both *KT;H11*$^{LSL\text{-}Cas9}$ and *KT* mice ($n = 5$, 10, 20, 50, 100, 200). For simplicity, we maintained the approximate 4:1 ratio of *KT;H11*$^{LSL\text{-}Cas9}$:*KT* used in this study while ensuring that there was more than 1 *KT* mouse per cohort (e.g., for 50 total mice, we sampled 40 *KT;H11*$^{LSL\text{-}Cas9}$ and 10 *KT* mice). Analogous to the tumor size simulations, we modeled the effect of the number of pooled sgRNAs by scaling the number of tumors sampled from each mouse by the ratio of the number of sgIDs in the underlying data to the simulated number of sgIDs ($n = 10$, 20, 50, 100, 200, 500); the resulting data set was then bootstrapped to assess significance. To capture differences in power due to effect size, analyses were performed for representative strong, moderate, and weak suppressors of tumor initiation (sg*Pten*#2,

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

sg*Kdm6a*#2, and sg*Ncoa6*#1, respectively). Five hundred simulations were performed for each gene, with a minimum of 16,000 bootstrap samplings per simulation. In each bootstrap, the relative tumor number (ratio of number of tumors with focal genotype to number of sg*Inert* tumors) in *KT;H11*$^{LSL\text{-}Cas9}$ mice was compared with the relative tumor number in *KT* mice, and significance in each simulation was assessed by bootstrapped P value <0.05 (two-tailed test, Bonferroni corrected for the simulated number of pooled sgRNAs).

DepMap Data and Filtering

Cancer cell line dependency data (DepMap Public 19Q4) and mutation data (Cancer Cell Line Encyclopedia) were acquired from the Broad Institute DepMap Portal (RRID:SCR_017655) (59). Lung adenocarcinoma cell lines were identified by their Project Achilles identification code. For each gene of interest, the cell lines that contained damaging mutations within the gene were identified and flagged. Damaging mutations were defined as mutations that likely caused loss of gene function. Subsequently, dependency scores for each gene of interest were exported from both the complete data set of lung adenocarcinoma cell lines and data set of cell lines that contains no damaging mutation in the gene of interest. Finally, the distribution of dependency scores across each gene of interest was plotted using GraphPad Prism 8.

Authors' Disclosures

H. Cai reports grants from the University of California during the conduct of the study. S.K. Chew reports grants from Ono Pharmaceuticals outside the submitted work. S.C. Lee reports other support from Agency for Science, Technology and Research (A*STAR) outside the submitted work. M. Yousefi reports grants from the NIH, American Lung Association, and Stanford Dean's Fellowship during the conduct of the study. D.A. Petrov reports personal fees from D2G Oncology outside the submitted work; in addition, D.A. Petrov has patents (10738300 and 10801021) issued to D2G Oncology and patents pending (20190367908 and 20180282720) to D2G Oncology. C. Swanton reports grants and personal fees from Pfizer, Bristol Myers Squibb, Ono Pharmaceuticals, and Roche-Ventana; grants from Boehringer Ingelheim and Archer Dx Inc. (collaboration in minimal residual disease sequencing technologies); personal fees from Novartis, GlaxoSmithKline, MSD, Celgene, Illumina, Amgen, Genentech, GRAIL, Bicycle Therapeutics, Medixci, Sarah Canon Research Institute, Achilles Therapeutics, and AstraZeneca; stock options from GRAIL, Epic Biosciences, Apogen Biotechnologies, and Achilles Therapeutics; also, C. Swanton is an advisory board member and chief investigator for the MeRmaiD1 clinical trial from AstraZeneca and is cofounder of Achilles Therapeutics outside the submitted work; in addition, C. Swanton has a patent for assay technology to detect tumor recurrence (PCT/GB2017/053289), a patent for targeting neoantigens (PCT/EP2016/059401), a patent for response to immune checkpoint blockade (PCT/EP2016/071471), a patent for determining whether HLA LOH is lost in a tumor (PCT/GB2018/052004), a patent for predicting survival rates of cancer patients (PCT/GB2020/050221), a patent for treating cancer by targeting insertion/deletion mutations (PCT/GB2018/051893), a patent for identifying insertion/deletion mutation targets (PCT/GB2018/051892), a patent for detecting tumor mutations (PCT/US2017/028013), and a patent for identifying responders to cancer treatment (PCT/GB2018/051912). M.M. Winslow reports grants from the NIH during the conduct of the study, has filed patents broadly related to this work, and is a founder and adviser of D2G Oncology, Inc. No disclosures were reported by the other authors.

Authors' Contributions

H. Cai: Conceptualization, validation, investigation, visualization, writing–original draft. **S.K. Chew:** Conceptualization, formal analysis, investigation, visualization, methodology, writing–original draft. **C. Li:** Data curation, formal analysis, validation, investigation, visualization, methodology, writing–original draft. **M.K. Tsai:** Investigation. **L. Andrejka:** Investigation. **C.W. Murray:** Investigation, writing–review and editing. **N.W. Hughes:** Formal analysis. **E.G. Shuldiner:** Formal analysis, visualization. **E.L. Ashkin:** Investigation. **R. Tang:** Resources. **K.L. Hung:** Validation, investigation. **L.C. Chen:** Formal analysis, investigation. **S.C. Lee:** Formal analysis. **M. Yousefi:** Investigation. **W. Lin:** Investigation. **C.A. Kunder:** Formal analysis. **L. Cong:** Supervision. **C.D. McFarland:** Formal analysis, methodology. **D.A. Petrov:** Conceptualization, resources, supervision, funding acquisition, project administration, writing–review and editing. **C. Swanton:** Conceptualization, resources, supervision, funding acquisition, writing–review and editing. **M.M. Winslow:** Conceptualization, resources, supervision, funding acquisition, investigation, methodology, project administration, writing–review and editing.

Acknowledgments

We thank the Stanford Shared FACS Facility for flow cytometry and cell sorting services, the Stanford Veterinary Animal Care Staff for expert animal care, Human Pathology/Histology Service Center, Stanford Protein and Nucleic Acid Facility, the Francis Crick Genomics Equipment Park, Advanced Sequencing Facility, Bioinformatics & Biostatistics, and Y. Zhao, D. Maghini, and R. Ma for experimental support; A. Orantes for administrative support; R. Levine's laboratory for making the *Stag2*^{flox} allele available prior to publication; and D. Feldser, J. Sage, and members of the Winslow, Petrov, and Swanton laboratories for helpful comments. H. Cai was supported by a Tobacco-Related Disease Research Program (TRDRP) Postdoctoral Fellowship (28FT-0019). S.K. Chew was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013) Consolidator Grant (THESEUS). C. Li is the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2331). C.W. Murray was supported by the NSF Graduate Research Fellowship Program and an Anne T. and Robert M. Bass Stanford Graduate Fellowship. N.W. Hughes was supported by the NSF Graduate Research Fellowship Program. R. Tang was supported by a Stanford University School of Medicine Dean's Postdoctoral Fellowship and a TRDRP Postdoctoral fellowship (27FT-0044). M. Yousefi was supported by a Stanford University School of Medicine Dean's fellowship, an American Lung Association senior research training grant, and NIH Ruth L. Kirschstein National Research Service Award (F32-CA236311). C.D. McFarland was supported by NIH K99-CA226506. W.-Y. Lin was supported by an American Association for Cancer Research Postdoctoral fellowship (17–40–18-LIN). C. Swanton is Royal Society Napier Research Professor (RP150154). This work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001169), the UK Medical Research Council (FC001169), and the Wellcome Trust (FC001169). This research was funded in whole, or in part, by the Wellcome Trust (FC001169). For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission. C. Swanton is funded by Cancer Research UK (TRACERx, PEACE, and CRUK Cancer Immunotherapy Catalyst Network), Cancer Research UK Lung Cancer Centre of Excellence, the Rosetrees Trust, Butterfield and Stoneygate Trusts, NovoNordisk Foundation (ID16584), Royal Society Professorship Enhancement Award (RP/EA/180007), the National Institute for Health Research (NIHR) Biomedical Research Centre at University College London Hospitals, the CRUK-UCL Centre, Experimental Cancer Medicine Centre, and the Breast Cancer Research Foundation (BCRF). This research is supported by a Stand Up To Cancer–LUNGevity–American Lung Association Lung Cancer Interception Dream Team Translational Research Grant (Grant Number: SU2C-AACR-DT23–17). Stand Up To Cancer (SU2C) is a program of the Entertainment Industry Foundation. Research

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

grants are administered by the American Association for Cancer Research, the Scientific Partner of SU2C. C. Swanton receives funding from the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007–2013) Consolidator Grant (FP7-THESEUS-617844), European Commission ITN (FP7-PloidyNet 607722), an ERC Advanced Grant (PROTEUS) from the European Research Council under the European Union's Horizon 2020 research and innovation program (grant agreement No. 835297), and Chromavision from the European Union's Horizon 2020 research and innovation program (grant agreement 665233). This work was supported by NIH R01-CA207133 (to M.M. Winslow and D.A. Petrov), NIH R01-CA231253 (to M.M. Winslow and D.A. Petrov), NIH R01-CA234349 (to M.M. Winslow and D.A. Petrov), and in part by the Stanford Cancer Institute support grant (NIH P30-CA124435).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received September 14, 2020; revised December 25, 2020; accepted February 12, 2021; published first February 19, 2021.

REFERENCES

1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144:646–74.
2. Consortium APG. AACR Project GENIE: powering precision medicine through an international consortium. Cancer Discov 2017;7:818–31.
3. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature 2020;578:82–93.
4. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature 2014;511:543–50.
5. Zehir A, Benayed R, Shah RH, Syed A, Middha S, Kim HR, et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. Nat Med 2017;23:703–13.
6. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TBK, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. N Engl J Med 2017;376:2109–21.
7. Cancer Genome Atlas Research Network, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet 2013;45:1113–20.
8. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell 2018;173:371–85.
9. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature 2013;499:214–8.
10. Greaves M, Maley CC. Clonal evolution in cancer. Nature 2012;481:306–13.
11. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature 2009;458:719–24.
12. Winters IP, Murray CW, Winslow MM. Towards quantitative and multiplexed in vivo functional cancer genomics. Nat Rev Genet 2018;19:741–55.
13. Zahir N, Sun R, Gallahan D, Gatenby RA, Curtis C. Characterizing the ecological and evolutionary dynamics of cancer. Nat Genet 2020;52:759–67.
14. Ben-David U, Beroukhim R, Golub TR. Genomic evolution of cancer models: perils and opportunities. Nat Rev Cancer 2019;19:97–109.
15. Graham TA, Sottoriva A. Measuring cancer evolution from the genome. J Pathol 2017;241:183–91.
16. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell 2015;27:15–26.
17. Garraway LA, Lander ES. Lessons from the cancer genome. Cell 2013;153:17–37.
18. Howard TP, Vazquez F, Tsherniak A, Hong AL, Rinne M, Aguirre AJ, et al. Functional genomic characterization of cancer genomes. Cold Spring Harb Symp Quant Biol 2016;81:237–46.
19. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. Nat Rev Cancer 2015;15:747–56.
20. Weber J, Braun CJ, Saur D, Rad R. In vivo functional screening for systems-level integrative cancer genomics. Nat Rev Cancer 2020;20:573–93.
21. Kersten K, de Visser KE, van Miltenburg MH, Jonkers J. Genetically engineered mouse models in oncology research and cancer medicine. EMBO Mol Med 2017;9:137–53.
22. Sanchez-Rivera FJ, Papagiannakopoulos T, Romero R, Tammela T, Bauer MR, Bhutkar A, et al. Rapid modelling of cooperating genetic events in cancer through somatic genome editing. Nature 2014;516:428–31.
23. Annunziato S, Kas SM, Nethe M, Yucel H, Del Bravo J, Pritchard C, et al. Modeling invasive lobular breast carcinoma by CRISPR/Cas9-mediated somatic genome editing of the mammary gland. Genes Dev 2016;30:1470–80.
24. Chiou SH, Winters IP, Wang J, Naranjo S, Dudgeon C, Tamburini FB, et al. Pancreatic cancer modeling using retrograde viral vector delivery and in vivo CRISPR/Cas9-mediated somatic genome editing. Genes Dev 2015;29:1576–85.
25. Xue W, Chen S, Yin H, Tammela T, Papagiannakopoulos T, Joshi NS, et al. CRISPR-mediated direct mutation of cancer genes in the mouse liver. Nature 2014;514:380–4.
26. Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang CH, Petrov D, et al. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression in vivo. Nat Methods 2017;14:737–42.
27. Rogers ZN, McFarland CD, Winters IP, Seoane JA, Brady JJ, Yoon S, et al. Mapping the in vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. Nat Genet 2018;50:483–6.
28. Winters IP, Chiou SH, Paulk NK, McFarland CD, Lalgudi PV, Ma RK, et al. Multiplexed in vivo homology-directed repair and tumor barcoding enables parallel quantification of Kras variant oncogenicity. Nat Commun 2017;8:2053.
29. Kumar RD, Searleman AC, Swamidass SJ, Griffith OL, Bose R. Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data. Bioinformatics 2015;31:3561–8.
30. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res 2019;47:D330–D8.
31. Iwanaga K, Yang Y, Raso MG, Ma L, Hanna AE, Thilaganathan N, et al. Pten inactivation accelerates oncogenic K-ras-initiated tumorigenesis in a mouse model of lung cancer. Cancer Res 2008;68:1119–27.
32. Ji H, Ramsey MR, Hayes DN, Fan C, McNamara K, Kozlowski P, et al. LKB1 modulates lung cancer differentiation and metastasis. Nature 2007;448:807–10.
33. Walter DM, Venancio OS, Buza EL, Tobias JW, Deshpande C, Gudiel AA, et al. Systematic in vivo inactivation of chromatin-regulating enzymes identifies Setd2 as a potent tumor suppressor in lung adenocarcinoma. Cancer Res 2017;77:1719–29.
34. Wang X, Min S, Liu H, Wu N, Liu X, Wang T, et al. Nf1 loss promotes Kras-driven lung adenocarcinoma and results in Psat1-mediated glutamate dependence. EMBO Mol Med 2019;11:e9856.
35. Werner M, Purta E, Kaminska KH, Cymerman IA, Campbell DA, Mittra B, et al. 2′-O-ribose methylation of cap2 in human: function and evolution in a horizontally mobile family. Nucleic Acids Res 2011;39:4756–68.
36. Koo BK, Spit M, Jordens I, Low TY, Stange DE, van de Wetering M, et al. Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. Nature 2012;488:665–9.
37. Jiang X, Hao HX, Growney JD, Woolfenden S, Bottiglio C, Ng N, et al. Inactivating mutations of RNF43 confer Wnt dependency in pancreatic ductal adenocarcinoma. Proc Natl Acad Sci U S A 2013;110:12649–54.
38. Koo BK, van Es JH, van den Born M, Clevers H. Porcupine inhibitor suppresses paracrine Wnt-driven growth of Rnf43;Znrf3-mutant neoplasia. Proc Natl Acad Sci U S A 2015;112:7548–50.

Published OnlineFirst February 19, 2021; DOI: 10.1158/2159-8290.CD-20-1325

39. Balbas-Martinez C, Sagrera A, Carrillo-de-Santa-Pau E, Earl J, Marquez M, Vazquez M, et al. Recurrent inactivation of STAG2 in bladder cancer is not associated with aneuploidy. Nat Genet 2013;45:1464–9.
40. Romero-Perez L, Surdez D, Brunet E, Delattre O, Grunewald TGP. STAG mutations in cancer. Trends Cancer 2019;5:506–20.
41. Solomon DA, Kim JS, Bondaruk J, Shariat SF, Wang ZF, Elkahloun AG, et al. Frequent truncating mutations of STAG2 in bladder cancer. Nat Genet 2013;45:1428–30.
42. Viny AD, Bowman RL, Liu Y, Lavallee VP, Eisman SE, Xiao W, et al. Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. Cell Stem Cell 2019;25:682–96.
43. Kleyman M, Kabeche L, Compton DA. STAG2 promotes error correction in mitosis by regulating kinetochore-microtubule attachments. J Cell Sci 2014;127:4225–33.
44. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, et al. Mutational inactivation of STAG2 causes aneuploidy in human cancer. Science 2011;333:1039–43.
45. Kong X, Ball AR Jr, Pham HX, Zeng W, Chen HY, Schmiesing JA, et al. Distinct functions of human cohesin-SA1 and cohesin-SA2 in double-strand break repair. Mol Cell Biol 2014;34:685–98.
46. Mondal G, Stevers M, Goode B, Ashworth A, Solomon DA. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. Nat Commun 2019;10:1686.
47. Shen CH, Kim SH, Trousil S, Frederick DT, Piris A, Yuan P, et al. Loss of cohesin complex components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. Nat Med 2016;22:1056–61.
48. Ding S, Diep J, Feng N, Ren L, Li B, Ooi YS, et al. STAG2 deficiency induces interferon responses via cGAS-STING pathway and restricts virus infection. Nat Commun 2018;9:1485.
49. Inoki K, Li Y, Zhu T, Wu J, Guan KL. TSC2 is phosphorylated and inhibited by Akt and suppresses mTOR signalling. Nat Cell Biol 2002;4:648–57.
50. Shilatifard A. The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. Annu Rev Biochem 2012;81:65–95.
51. Wu Q, Tian Y, Zhang J, Tong X, Huang H, Li S, et al. In vivo CRISPR screening unveils histone demethylase UTX as an important epigenetic regulator in lung tumorigenesis. Proc Natl Acad Sci U S A 2018;115:E3978–E86.
52. Hill BM. A simple general approach to inference about the tail of a distribution. Ann Stat 1975;3:1163–74.
53. Jackson EL, Olive KP, Tuveson DA, Bronson R, Crowley D, Brown M, et al. The differential effects of mutant p53 alleles on advanced murine lung cancer. Cancer Res 2005;65:10280–8.
54. Feldser DM, Kostova KK, Winslow MM, Taylor SE, Cashman C, Whittaker CA, et al. Stage-specific sensitivity to p53 restoration during lung cancer progression. Nature 2010;468:572–5.
55. Johnson L, Mercer K, Greenbaum D, Bronson RT, Crowley D, Tuveson DA, et al. Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. Nature 2001;410:1111–6.
56. Gao Q, Steine EJ, Barrasa MI, Hockemeyer D, Pawlak M, Fu D, et al. Deletion of the de novo DNA methyltransferase Dnmt3a promotes lung tumor progression. Proc Natl Acad Sci U S A 2011;108:18061–6.
57. Schuster K, Venkateswaran N, Rabellino A, Girard L, Pena-Llopis S, Scaglioni PP. Nullifying the CDKN2AB locus promotes mutant K-ras lung tumorigenesis. Mol Cancer Res 2014;12:912–23.
58. Andricovich J, Perkail S, Kai Y, Casasanta N, Peng W, Tzatsos A. Loss of KDM6A activates super-enhancers to induce gender-specific squamous-like pancreatic cancer and confers sensitivity to BET inhibitors. Cancer Cell 2018;33:512–26.
59. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a cancer dependency map. Cell 2017;170:564–76.
60. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, et al. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature 2014;505:495–501.
61. Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. Nature 2020;578:102–11.
62. Alexandrov LB, Kim J, Haradhvala NJ, Huang MN, Tian Ng AW, Wu Y, et al. The repertoire of mutational signatures in human cancer. Nature 2020;578:94–101.
63. Curry NL, Mino-Kenudson M, Oliver TG, Yilmaz OH, Y VO, Moon JY, et al. Pten-null tumors cohabiting the same lung display differential AKT activation and sensitivity to dietary restriction. Cancer Discov 2013;3:908–21.
64. Yanagi S, Kishimoto H, Kawahara K, Sasaki T, Sasaki M, Nishio M, et al. Pten controls lung morphogenesis, bronchioalveolar stem cells, and onset of lung adenocarcinomas in mice. J Clin Invest 2007;117:2929–40.
65. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res 2012;22:398–406.
66. van de Haar J, Canisius S, Yu MK, Voest EE, Wessels LFA, Ideker T. Identifying epistasis in cancer genomes: a delicate affair. Cell 2019;177:1375–83.
67. Li C, Lin WY, Rizvi H, Cai H, McFarland CD, Rogers ZN, et al. Quantitative in vivo analyses reveal a complex pharmacogenomic landscape in lung adenocarcinoma. bioRxiv 2020. Preprint: https://www.biorxiv.org/content/10.1101/2020.01.28.923912v1. Provisionally accepted by *Cancer Research*.
68. Foggetti G, Li C, Cai H, Hellyer JA, Lin W-Y, Ayeni D, et al. Genetic determinants of EGFR-driven lung cancer growth and therapeutic response in vivo. Cancer Discov 2021;11:1736–53.
69. Ellrott K, Bailey MH, Saksena G, Covington KR, Kandoth C, Stewart C, et al. Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst 2018;6:271–81.
70. McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med 2015; 7:283ra54.
71. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the catalogue of somatic mutations in cancer. Nucleic Acids Res 2019;47:D941–D7.
72. Hough SH, Kancleris K, Brody L, Humphryes-Kirilov N, Wolanski J, Dunaway K, et al. Guide Picker is a comprehensive design tool for visualizing and selecting guides for CRISPR experiments. BMC Bioinformatics 2017;18:167.
73. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat Biotechnol 2016;34:184–91.
74. Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat Biotechnol 2013;31:827–32.
75. Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. Nat Methods 2014;11:705–6.
76. Madisen L, Zwingman TA, Sunkin SM, Oh SW, Zariwala HA, Gu H, et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat Neurosci 2010;13:133–40.
77. Jackson EL, Willis N, Mercer K, Bronson RT, Crowley D, Montoya R, et al. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev 2001;15:3243–8.
78. Clement K, Rees H, Canver MC, Gehrke JM, Farouni R, Hsu JY, et al. CRISPResso2 provides accurate and rapid genome editing sequence analysis. Nat Biotechnol 2019;37:224–6.
79. Chuang CH, Greenside PG, Rogers ZN, Brady JJ, Yang D, Ma RK, et al. Molecular definition of a metastatic lung cancer state reveals a targetable CD109-Janus kinase-Stat axis. Nat Med 2017;23:291–300.

AACR American Association for Cancer Research

CANCER DISCOVERY

A Functional Taxonomy of Tumor Suppression in Oncogenic KRAS–Driven Lung Cancer

Hongchen Cai, Su Kit Chew, Chuan Li, et al.

Cancer Discov 2021;11:1754-1773. Published OnlineFirst February 19, 2021.