# A functional taxonomy of tumor suppression in oncogenic KRAS-driven lung cancer

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

[1] Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA

[2] Cancer Biology Program, Stanford University School of Medicine, Stanford, CA, USA

[3] Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA

[4] Department of Biology, Stanford University, Stanford, CA, USA

[5] Cancer Evolution and Genome Instability Laboratory, University College London Cancer Institute, London, UK.

[6] Cancer Evolution and Genome Instability Laboratory, The Francis Crick Institute, London, UK.

* These authors contributed equally

[#] 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 | Email: Charles.Swanton@crick.ac.uk

Dmitri A. Petrov, Biology Department, Stanford University, Bass Biology Building, 327 Campus Drive Stanford 94305. Phone: 650-736-1169| Fax: 650-7366132 | E-mail: dpetrov@stanford.edu

## ABSTRACT

Cancer genotyping has identified a large number of putative tumor suppressor genes. Carcinogenesis is a multi-step process, however 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 forty-eight 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, while 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.


## STATEMENT OF SIGNIFICANCE

Our high-throughput and high-resolution analysis of tumor suppression uncovered genetic determinants of lung cancer initiation, overall growth, and exceptional growth. This taxonomy revealed key novel regulators of each phase of oncogenic KRAS-driven lung carcinogenesis, 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 multi-step 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 is 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 occurs 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 were chosen based on multiple criteria including their mutational frequency in lung adenocarcinoma from TCGA, GENIE, and TRACERx datasets, their mutational frequency in pan-cancer genomic data, and the consistency of their mutational profiles with tumor suppressor activity (**Fig. 1A-B; Supplementary Fig. S1** and **Table S1**)(2,4-7). We also considered their putative tumor-suppressive function in other cancer types as well as their molecular functions (**Supplementary Fig. S2** and **S3**)(8,29,30). Our candidate genes vary greatly in their mutation frequency and co-occurrence with oncogenic KRAS alterations (**Supplementary Fig. S2)**. 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. S3**).

**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-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 initiate 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 individual tumor, of each genotype, in each mouse.

*KT;*$H11^{LSL\text{-}Cas9}$ mice had ~10-fold higher total neoplastic cell number and proportionally increased total 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 genotype) 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 where 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, S4C** and **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** and **Supplementary Fig. S4**). 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 *Lkb1, Pten, Setd2*, and *Nf1* in tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice greatly increased tumor growth (**Fig. 2; Supplementary Fig. S5**). Unexpectedly, inactivation of STAG2, a cohesin complex component, increased tumor growth to a comparable extent as inactivation of those well-established tumor suppressors (**Fig. 2**; **Supplementary Fig. S5**).

Inactivation of 14 other genes, including *Cdkn2c, Cmtr2, Rb1, Rnf43, Tsc1*, and *Rbm10,* significantly increased tumor growth (**Fig. 2A and 2B**; **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. S3**). 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, where *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 novel functional 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. S6 and 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;H11*$^{LSL\text{-}Cas9}$ and *KT* 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-E**). Inactivation of *Stag2* in lung tumors in *KT;H11*$^{LSL\text{-}Cas9}$ mice also significantly reduced long-term 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 proficient lung tumors were atypical adenomatous hyperplasias, adenomas, and early adenocarcinomas 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 and H**). 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. S8**). Thus, further work will be necessary to determine the molecular mechanisms of tumor suppression driven by STAG2.

Finally, to further characterize the expression of STAG2 in lung cancer, we perform 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 tumors were often more poorly differentiated and advanced human lung adenocarcinomas (**Fig. 3I**).

**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 timepoint 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, p53, 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 viral titer that would allow them to survive for 26 weeks while maximizing tumor number to achieve reasonable statistical power (**Fig. 4A; Supplementary Fig. S9A** and **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**).

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* also still increased tumor sizes at defined percentiles of the distribution as well as the log-normal mean tumor size at this later time point (**Fig. 4C and 4D**; **Supplementary Fig. S9B**). These results confirm the tumor-suppressive function of these genes (**Supplementary Fig. S9B**). 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 impacted 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 driven were disproportionately large (**Supplementary Fig. S10B**). For example, *p53* 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 95% percentile tumor. 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 C**). These disproportionate increases in relative tumor burden could be driven by genotype-specific increases in tumor number and/or the sizes of the very largest tumors, neither of which are captured very well by log-normal mean or tumor sizes at defined percentile 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 exact 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 tumors 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-quarters of epithelial cells expressing oncogenic Kras$^{G12D}$ fail to expand beyond a very small size 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 *p53* 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*)(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 8.9-32.5% of human lung adenocarcinoma (**Fig. 5E and F**)(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. 5C; Supplementary Fig. S11B and C**). The effects of inactivating each tumor suppressor gene on relative tumor numbers were highly correlated across all three datasets (**Fig. 5D; Supplementary Fig. S11D and E**). Several genes including *Cdkn2a, Dnmt3a, Kdm6a* and *Ncoa6* that initially only increased tumor number also increased overall growth fitness at the later time point. This observation suggests some link between the cellular changes that enable normal epithelial cells to break through the constraints early in the hyperplastic growth and the greater fitness in the resulting tumors (**Fig. 4F and 5D; 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 *p53* deficiency is the generation of such exceptionally large tumors (26,27). Using metrics such as the Hill's estimator (a measure of the heavy-tailedness of a distribution)(52), we quantified the extent to which *p53* inactivation enables the emergence of infrequent but exceptionally large tumors after 15 weeks of tumor growth (**Fig. 6A and 6B; Supplementary**

**Fig. S12A**). The effect of *p53* inactivation is consistent with many previous reports documenting the emergence of larger lung tumors in $Kras^{LSL\text{-}G12D/+};p53^{flox/flox}$ mice (32,53-55). These analyses also show that inactivation of *Cdkn2a* and the DNA methyltransferase *Dnmt3a*, might also allow some tumors to grow to disproportionately large sizes (**Fig. 6A and 6B; 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 of the distributions of tumor sizes specifically highlighted the development of exceptionally large *Dnmt3a* and *Cdkn2a*-targeted tumors (**Fig. 6C-E; Supplementary Fig. S12B-D**). Both sgRNAs targeting *Cdkn2a* are anticipated to inactivate both INK4A and ARF, therefore the effect of *Cdkn2a* inactivation could reflect the combined reduction of the Rb and p53-pathways, consistent with our observation that p53 inactivation generates a heavy-tailed distribution (**Fig. 6; Supplementary Fig. S12**)(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 the vast majority of 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 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 influences of overall tumor burden on 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 would change with tumor density (**Supplementary Fig. S13B-E**).

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. S14**). Thus, tumor suppressor effects in lung cancer are not dramatically impacted by differences in the host environment driven by sex. This was particularly illuminating for *Kdm6a,* which is an X chromosome gene that has both H3K27me3 demethylase and non-enzymatic functions (58). Its non-enzymatic 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 its enzymatic function (**Supplementary Fig. S14**).

## Evaluation of sensitivity and specificity of Tuba-seq

To better estimate the impact of false negatives and false positives on our data, we used all of our datasets to estimate the true positive rate (**Methods**). Within all of our datasets, sgRNAs targeting the same gene showed strong concordance across multiple metrics, consistent with on-target effects (**Fig. 2 and 4; Supplementary Fig. S15A-F**). For instance, in our initial experiment using Lenti-sg*TS102*/Cre pool, when one sgRNA showed a significant tumor suppressive effect (nominal $P < 0.05$), the probability to re-detect the significant effect using the other guide was above 89% for all metrics assessed (**Fig. 2A; 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-I**). 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. S16**).

## Human mutational data, cell line studies, and *in vivo* functional studies are complementary in defining a catalog of tumor suppression

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 considered whether effects on tumorigenesis within the autochthonous environment could be predicted by either 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. S17**). 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. S17**). 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 gene length and mutation rate 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 tumor suppressor gene importance 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 spot, epistatic interactions and biological context (9,62) rather than by the magnitude of their inhibitory function (**Supplementary Fig. S17**). Thus, while experiments using model organisms could be impacted 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 impacted by different aspects of the *in vivo* environment. Taking advantage of 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. 7**). We show that tumor suppression is unexpectedly complex and multi-faceted, 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. Furthermore, while some genes affect only a single feature of carcinogenesis, others affect multiple facets of tumor evolution to varying extents (**Fig. 7**). The relative importance of these genes can also change during the course of carcinogenesis. 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.

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, while *Lkb1, Pten, Kdm6a, Dnmt3a* and *p53* inactivation each increase overall tumor burden, our quantitative, multiplexed design and computational platform uniquely enabled the deconvolution of different aspects of tumor suppression (**Fig. 7**).

We show that the inactivation of many understudied genes has major effects on tumor growth. Identifying additional genes that are fundamentally important in suppressing carcinogenesis, including those that are infrequently mutated in human lung adenocarcinoma, 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 suggest not only the importance of certain genes but also more broadly uncover under-appreciated cellular processes that limit cancer development.

One key approach used to implicate the context-dependency of tumor suppressor function is the analysis of mutual exclusivity in human data, which may indicate functionally redundant genomic alterations (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 *KRAS*-driven lung cancer (**Supplementary Fig. S17**). 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. 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 extrinsic processes that contribute to the etiology and evolution of lung cancer.

**ACKNOWLEDGEMENTS**

We thank the Stanford Shared FACS Facility, the Veterinary Service Center, Human Pathology/Histology Service Center, 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; D. Feldser, J. Sage, and members of the Winslow, Petrov, and Swanton laboratories for helpful comments. H.C. was supported by a Tobacco-Related Disease Research Program (TRDRP) Postdoctoral Fellowship (28FT-0019). S.K.C. was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) Consolidator Grant (THESEUS). C.L. is the Connie and Bob Lurie Fellow of the Damon Runyon Cancer Research Foundation (DRG-2331). C.W.M. was supported by the NSF Graduate Research Fellowship Program and an Anne T. and Robert M. Bass Stanford Graduate Fellowship. N.W.H. was supported by the NSF Graduate Research Fellowship Program. R.T. was supported by a Stanford University School of Medicine Dean's Postdoctoral Fellowship and a TRDRP Postdoctoral fellowship (27FT-0044). M.Y. 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.M. was supported by NIH K99-CA226506. W-Y.L. was supported by an American Association of Cancer Research Postdoctoral fellowship (17-40-18-LIN). C.S. is Royal Society Napier Research Professor and was supported by the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013) Consolidator Grant (THESEUS), and the European Union's Horizon 2020 Research and Innovation Programme Advanced Grant (PROTEUS). This work was supported by the Francis Crick Institute that receives its core funding from Cancer Research UK (to C.S.), the

UK Medical Research Council (to C.S.), and the Wellcome Trust (FC001169, FC001202 to C.S.). This work was supported by NIH R01-CA207133 (to M.M.W and D.A.P.), NIH R01-CA231253 (to M.M.W and D.A.P), NIH R01-CA234349 (to M.M.W and D.A.P.), and in part by the Stanford Cancer Institute support grant (NIH P30-CA124435).

## CONTRIBUTIONS

H.C., S.K.C., C.L., M.M.W., C.S., and D.A.P. designed the project. H.C. and S.K.C. generated the lentiviral vector pool and initiated lung tumors in mice. M.K.T. generated the mice. H.C., S.K.C., C.L., M.K.T., C.W.M., R.T., K.L.H., L.C.C. and M.Y. collected lung samples. L.A., E.L.A. and K.L.H. performed immunohistochemical staining. H.C., S.K.C., L.A., C.W.M. and W.Y.L. generated the barcode sequencing library. C.L. and E.G.S. analyzed the Tuba-seq data. N.W.H. and L.C. analyzed DepMap and indel data. S.K.C., L.C.C., S.Y.C.L. and C.D.M. analyzed the human datasets. C.A.K. analyzed the tumor histology. H.C., S.K.C., C.L., D.A.P. and M.M.W. wrote the manuscript with comments from all authors.

## CONFLICT OF INTERESTS

S.K.C. receives grant support from Ono Pharma. C.S. receives grant support from Ono, Pfizer, AstraZeneca, BMS, Roche-Ventana and Boehringer-Ingelheim. C.S. has consulted for Pfizer, Novartis, GlaxoSmithKline, MSD, BMS, Celgene, AstraZeneca, Illumina, Genentech, Roche-Ventana, GRAIL, Medicxi, and the Sarah Cannon Research Institute. C.S. is a shareholder of Apogen Biotechnologies, Epic Bioscience, GRAIL, and has stock options in and is co-founder

of Achilles Therapeutics. D.A.P. and M.M.W. are founders of, and hold equity in, D2G Oncology Inc.

## METHODS

### Selection of candidate tumor suppressor genes for this study

In order to select candidate genes that allowed assessment of how *in vivo* quantitative validation using tuba-seq can complement genomics and cell biology approaches, a highly human-curated approach integrating different considerations was used.

Known lung adenocarcinoma driver tumor suppressors genes at >5% mutational frequency (such as *P53*, *LKB1*, *CDKN2A*, *KEAP1*) in from The Cancer Genome Atlas (TCGA), AACR Project Genomics, Evidence, Neoplasia, Information, Exchange (GENIE), and TRAcking Cancer Evolution through therapy (Rx) (TRACERx) datasets and previously assessed by Tuba-seq were included as positive controls, with representation from both genes that tend to co-occur with *KRAS* mutations and those that do not. Genes that have been categorized as tumor suppressor genes in other cancer types with >5% mutational frequency in lung (such as *PTEN*, *KDM6A*, and *FAT1*), even if not highly mutated or predicted to be involved in lung adenocarcinoma, were included to assess whether these genes drive lung cancer onset or progression (**Fig. 1A; Supplementary Fig. S1** and **Table S1**).

We also considered the profile of mutation distribution within genes (Fig. 1B), including low mutation frequency genes (<5%) that show potential clonal or subclonal bias from TRACERx data (**Supplementary Table S1**), genes with discrepancies in scoring of potential driver activity (**Supplementary Fig. S2**), as well as biological processes or functions commonly associated with oncogenesis (**Supplementary Fig. S3**). From a curated survey of literature,

503 candidate genes that have been discussed as likely or unlikely driver cancer genes without many
504 functional data available were also included (**Supplementary Fig. S4**).
505
506 **Analysis of human lung adenocarcinoma cancer genome sequencing data**
507 Mutation frequencies and other information for the 48-gene panel of putative candidate
508 tumor suppressor genes are available from multiple cancer datasets and their analyses such as
509 TRACERx (6), GENIE (2) and TCGA (7,69,70). Oncogenes are characterized by missense point
510 mutations arising in mutational hotspots. In contrast, TSGs are characterized by protein
511 truncating mutations (nonsense and frameshifts) that are more dispersed across the transcript.
512 Moreover, when nonsense and frameshift mutations arise in oncogenes, they tend to truncate C-
513 terminal domains and occur towards the end of the transcript. To identify putative TSGs, we
514 characterized all genes in this survey by these two genetic features: mutational hotspots and the
515 fraction of protein truncated per mutation. We used all point mutations and short insertion and
516 deletions found within the TCGA lung adenocarcinoma (7) and Catalogue Of Somatic Mutations
517 In Cancer (COSMIC)(71) databases. The extent of mutational hotspots within a gene was
518 determined using a normalized measure of dispersion (Green's Contagion) of the number of
519 missense mutations observed within all five residue rolling windows in each gene:
520 $(\sigma^2/\mu-1)/(\mu N-1)$ , where $\mu$ is the mean number of missense mutations observed within each
521 window, $\sigma^2$ is the unbiased estimator of the variance, and N is the number of missense mutations.
522 Green's Contagion and the five-residue window size and were chosen because they maximized
523 the accuracy of classification of known oncogenes and tumor suppressors. Larger values of
524 Green's Contagion suggest that mutations are clumping at a few residues within the protein and
525 that a gene is oncogenic. This measure has a value of zero when mutations are randomly

dispersed throughout the gene and can be negative when mutations are under-dispersed. The fraction of protein truncated per mutation is the mean number of amino acids lost per nonsynonymous mutation. It is calculated by simply averaging together the fraction of a transcript lost by 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)(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's Exact Test) were available on cBioPortal.org. 566 lung adenocarcinoma cases from TCGA Pan-cancer Atlas and 8522 lung adenocarcinoma samples from GENIE 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 simple linear regression between measured phenotypes and observed clinical parameters, the data from mutation timing and clonality in lung adenocarcinomas were previously described (6,70).

**Analysis of publications suggesting tumor suppressive function of each putative tumor suppressor gene in lung cancer**

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 keyword to refine the search.

**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 utilized 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 guide RNA sequences targeting the putative tumor suppressor genes were designed using Desktop Genetic's Guide Picker (72) (https://www.deskgen.com/guide-picker) to prioritize on-target activity (score of >0.6)(73), specificity (score of >0.6)(74), likelihood of generating frameshift indels (score of >0.6)(75), targeting of maximal number of transcript isoforms, no homopolymer runs in the guide, and no extremes in GC-content of guide (0.4-0.75), as detailed in **Supplementary Table S2**.

The derivation of the Lenti-U6-sgRNA-sgID-barcode-Pgk-Cre vectors 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 is annealed and ligated into the BaeI-linearised pLenti-BaeI-Pgk-Cre vector using T4 DNA ligase. The barcode oligo primer contains the 8nt sgID sequence and 20 degenerate nucleotides (**Supplementary Table S2**). The generation of the barcode fragment and subsequent ligation into the vector is as previously described (26).

Lenti-sgRNA/Cre vectors were individually co-transfected 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 0.45μm 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 (MEFs) (a gift from Dr. Alejandro Sweet-Cordero/UCSF). All lentiviral vector aliquots were stored at -80°C and were 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 (31,48,78,79). They were on a C57BL/6:129 mixed background. The *Stag2*$^{tm1c(EUCOMM)Wtsi}$/J (*Stag2*$^{flox}$) mice were initially generated by Viny *et al.*(42) and obtained from the Jackson Laboratory (RRID:IMSR_JAX:030902). Tumors were initiated by intratracheal delivery of 60 µl of lentiviral pools.

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 $9x10^4$ 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 $3x10^4$ 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 sub-pool of 85 Lenti-sgRNA/Cre vectors (Lenti-sgTS85/Cre), which excluded the vectors targeting *Cdkn2c, Lkb1, Nf1, p53, Pten, Rb1, Setd2*, and *Stag2*. Tumors were initiated in *KT* mice with $2.5x10^5$ ifu/mouse (6 mice analyzed at 15 weeks after tumor initiation), *KT;H11*$^{LSL\text{-}Cas9}$ mice with $6x10^4$ ifu/mouse (10 mice analyzed at 15 weeks after tumor initiation), and *KT;H11*$^{LSL\text{-}Cas9}$ mice with $1.5x10^4$ 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

614 individual barcoded Lenti-sgRNA/Cre vectors in *KT* mice with $1x10^5$ ifu/mouse (3 mice per
615 vector analyzed at 15 weeks after tumor initiation), and *KT;H11*$^{LSL\text{-}Cas9}$ mice with $1x10^5$
616 ifu/mouse (5-6 mice per vector analyzed at 15 weeks after tumor initiation).
617 For the survival experiments using Lenti-sgRNA/Cre-mediated gene inactivation, tumors
618 were allowed to develop until humane endpoints. Tumors were initiated in *KT;H11*$^{LSL\text{-}Cas9}$ mice
619 with individual barcoded Lenti-sg*Inert*/Cre vectors at $2x10^4$ ifu/mouse and with individual
620 barcoded Lenti-sg*Stag2*/Cre vectors at $1x10^4$ ifu/mouse (7 mice per vector analyzed).
621 For Stag2 validation experiments using the *Stag2*$^{floxed}$ allele, tumors were initiated with
622 Lenti-sg*Inert*/Cre in *KT, KT;Stag2*$^{flox/+}$, *KT;Stag2*$^{flox/flox}$ and *KT;Stag2*$^{flox/y}$ mice with $1x10^5$
623 ifu/mouse (4-5 mice per group analyzed) and allowed to develop for 15 weeks, and *KT,*
624 *KT;Stag2*$^{flox/+}$, *KT;Stag2*$^{flox/flox}$ and *KT;Stag2*$^{flox/y}$ mice with $1x10^5$ ifu/mouse (6-7 mice per
625 genotype analyzed) and allowed to develop until humane endpoints.
626
627 **Tuba-seq library generation**
628 Genomic DNA was isolated from bulk tumor-bearing lung tissue from each mouse as
629 previously described (26). Briefly, benchmark control cell lines were generated from LSL-YFP
630 MEFs transduced by a barcoded Lenti-sgNT3/Cre vector (NT3: an inert sgRNA with a distinct
631 sgID sequence) and purified by sorting $YFP^+$ cells. For mice initiated with Lenti-sg*TS102*/Cre
632 pool, twelve benchmark control cell lines (3 cell lines of 500,000 cells each, 3 cell lines of
633 50,000 cells, 3 cell lines of 5,000 cells, and 3 cell lines of 500 cells) were added to each mouse
634 lung sample prior to lysis to enable the calculation of the absolute number of neoplastic cells in
635 each tumor from the number of sgID-BC reads. Because the standard curve was highly linear, we
636 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 2x Master Mix (New England Biolabs, M0494X) was used to amplify the sgID-BC region from 32 μg of genomic DNA. The unique dual-indexed primers used were *Forward:* AATGATACGGCGACCACCGAGATCTACAC (8 nucleotides for i5 index) ACACTCTTTCCCTACACGACGCTCTTCCGATCT (6-9 random nucleotides for increased diversity) GCGCACGTCTGCCGCGCTG and *Reverse:* CAAGCAGAAGACGGCATACGAGAT (6 nucleotides for i7 index) GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT (9-6 random nucleotides for increased diversity) CAGGTTCTTGCGAACCTCAT. 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 2x150bp) on the Illumina HiSeq 2500 platform (Admera Health).

**Code and data availability**

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

The data sets generated during and/or analyzed during 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 8-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 with roughly equal representation of A, T, G and C. 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 forward 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 from the clonal expansion of a cell transfected by a lentiviral vector are expected to carry the same BC sequence. To minimize the effects of sequencing errors on calling the BC, we require the forward and reverse read to agree completely within the random nucleotide sequence to be further processed. In our pipeline, any “tumor” within a Hamming distance of two from a larger tumor is assigned as “spurious tumors”, which is likely to be resulting from sequencing or PCR error, and is removed from subsequent analysis. Reads with the same sgID and barcode are assigned to be the same tumor. The tumor size (number of neoplastic cells) is calculated by normalizing the number of reads to the benchmark control cell lines added to each sample prior to lysis of the lung and DNA extraction step. The minimum

sequencing depth is ~1 read per 43 cells. We have high statistical power in identifying tumors with over 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 in a later section), 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.

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

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**).

$$\text{Relative log} - \text{normal mean} = \frac{\text{log normal mean for sg}TS \text{ tumors}}{\text{log normal mean for sg}Inert \text{ tumors}}$$

**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's estimator and relative steepness to characterize the heavy-tailedness of the tumor size distribution.

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

$$H = \frac{1}{k}\sum_{i=0}^{k} \ln\left(\frac{X_i}{X_k}\right)$$

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

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

The steepness (99$^{th}$ percentile / 95$^{th}$ percentile) is calculated as the ratio of the 99$^{th}$ percentile over the 95$^{th}$ 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*.

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

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

For both relative Hill's estimator and relative steepness, values higher than one indicate that the gene inactivation leads to heavier tail and value smaller than one indicate gene inactivation leads to 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*$^{LSL\text{-}Cas9}$ 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 will normalize it to that that in *KT* mice to account for the viral titer differences among different sg*TS*.

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

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

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*$^{LSL\text{-}Cas9}$ 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*:

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

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

743 The relative tumor burden is determined mostly by the largest tumors. For instance, the
744 top 1% of tumor cells accounts for over 50% of total tumor burden in *KT;H11*$^{LSL\text{-}Cas9}$ mice at 11
745 weeks. Both *TS* inactivation that leads to faster overall growth, rare but exceptionally large
746 tumors and tumor initiation rate will result in an increase in relative tumor burden
747 (**Supplementary Fig. S4**).
748
749 **Bootstrapping the tumors**
750 In the calculation of confidence intervals and *P*-values, we need to generate distributions
751 of the statistic consider both variation of tumor sizes across mice and within the same mice. We
752 adopted a two-step bootstrap resampling process. We first bootstrap resampled mice to generate
753 a pseudogroup of mice and then within each group of resampled mice, we bootstrap resampled
754 all observed tumors carrying each sgID.
755
756 **Calculation of confidence intervals and *P*-values for size metrics**
757 We have four size metrics that describe the overall growth (relative log-normal mean,
758 relative percentiles) and the heavy tailedness (relative Hill's estimator and relative steepness) of
759 the tumor size distribution. For each of these metrics, we bootstrapped tumors 10,000 times and
760 calculate 10,000 values of each statistic for these bootstrap resampling. The 95% confidence
761 interval is calculated as the 2.5$^{th}$ percentile and the 97.5$^{th}$ percentile of these bootstrapped results,
762 while the *P*-value is calculated the proportion of bootstrapped results that are not in the same
763 direction as the observed score compared with the baseline of 1.
764
765 **Calculation of *P*-values for tumor burden and tumor number**

We bootstrap tumors in both the *KT;H11*$^{LSL\text{-}Cas9}$ and *KT* mice and calculate the relative tumor burden and relative tumor number from these bootstrapped mice. The process was repeated $10^6$ times. The 95% confidence interval is calculated as the 2.5$^{th}$ percentile and the 97.5$^{th}$ percentile of these bootstrapped results, while the *P*-value is 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 divide the 47 *KT;H11*$^{LSL\text{-}Cas9}$ 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 (95$^{th}$ 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 use the ratio to quantify the differences between female mice and male mice. The ratio is calculated as,

$$\text{Ratio} = \frac{\text{X}_{\text{Female}}}{\text{X}_{\text{Male}}}$$

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, (1) Lenti-sg*TS102*/Cre; 15-week experiment, (2) Lenti-sg*TS85*/Cre; 15-week experiment, and (3) Lenti-sg*TS85*/Cre; 26-week experiment. Understanding the true positive rate is important for understanding the probability of identifying functional tumor suppressor genes. Since 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 shows significant tumor suppressor effect, then the test is counted as TRUE (*T*). If the second sgRNA fails to show a significant tumor suppressor effect, then the test is 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 is 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: 95$^{th}$ percentile, log-normal mean, tumor burden, and tumor number. We did not estimate the true positive rate for Hill's 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 target sites in *Rosa26*$^{LSL\text{-}Tomato}$;*H11*$^{LSL\text{-}Cas9}$ mouse embryonic fibroblasts (MEFs) that were generated from E12.5 embryos. $10^5$ 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 2x150 base-pair 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 (76). 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 (IHC)**

Lung lobes were inflated with PBS/4% paraformaldehyde and fixed for 24 hours, stored in 70% ethanol, and paraffin-embedded and sectioned. 4 μm thick sections were collected for Hematoxylin and Eosin (H&E) staining and immunohistochemistry.

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 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 (77). 60,000-100,000 neoplastic cells were collected from each mouse. Genomic DNA and total RNA were purified using Qiagen AllPrep DNA/RNA Micro Kit (Cat# 80284). Genomic DNA was processed with Nextera Flex for karyotyping by low-pass (0.1x coverage) whole genome sequencing. $Log_2$ ratio of reads mapping to each genomic locus versus the average number of reads mapping to all other comparable loci was plotted.

For qRT-PCR total RNA was reverse-transcribed using 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:

*Fos*: 5'-TACTACCATTCCCCAGCCGA-3' and 5'-GCTGTCACCGTGGGGATAAA-3';

*Klf2*: 5'-GAGCCTATCTTGCCGTCCTT-3' and 5'-TTGTTTAGGTCCTCATCCGTG-3';

*Ifnl3*: 5'-GTGCAGTTCCCACCTCATCT-3' and 5'-TGGGAGTGAATGTGGCTCAG-3';

*Ifnb1*: 5'-GTCCTCAACTGCTCTCCACT-3' and 5'-CATCCAGGCGTAGCTGTTGTA-3';

*Mx1*: 5'-ACGGTGCAGACATACCAGAA-3' and 5'-CTGTCTCCCTCTGATACGGT-3';

*Ifi44*: 5'-ATGGCAGCAAGAAAAGTGCC-3' and 5'-AAACTTCTGCACACTCGCCT-3';

*Irf1*: 5'-CCAGAGATTGACAGCCCTCG-3' and 5'-TGCACAAGGAATGGCCTGAA-3';

*Gapdh*: 5'-TGTGAACGGATTTGGCCGTA-3' and 5'-ACTGTGCCGTTGAATTTGCC-3';

*Actb*: 5'-GGCTCCTAGCACCATGAAGA-3' and 5'-GTGTAAAACGCAGCTCAGTAACA-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 non-parametric nested resampling approaches.

For each simulation 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 guides results in fewer tumors per guide 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). 500 simulations were performed for each gene, with a minimum of 16,000 bootstrap samplings per simulation. In each bootstrap, the size of neoplasia at 95$^{th}$ percentile with the focal genotype was compared to the size of neoplasia with sg*Inert*s at 95$^{th}$ 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 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, where the relative abundance of genotypes reflects the make-up 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 model 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 dataset 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, sg*Kdm6a*#2, and sg*Ncoa6*#1, respectively). 500 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 inert tumors) in *KT;H11*$^{LSL\text{-}Cas9}$ mice was compared to 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 (CCLE) 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 dataset of lung adenocarcinoma cell lines and dataset 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.

## REFERENCES


1. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell **2011**;144(5):646-74 doi 10.1016/j.cell.2011.02.013.
2. Consortium APG. AACR Project GENIE: Powering Precision Medicine through an International Consortium. Cancer Discov **2017**;7(8):818-31 doi 10.1158/2159-8290.CD-17-0151.
3. Consortium ITP-CAoWG. Pan-cancer analysis of whole genomes. Nature **2020**;578(7793):82-93 doi 10.1038/s41586-020-1969-6.
4. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. Nature **2014**;511(7511):543-50 doi 10.1038/nature13385.
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(6):703-13 doi 10.1038/nm.4333.

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(22):2109-21 doi 10.1056/NEJMoa1616288.

7. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA*, et al.* The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet **2013**;45(10):1113-20 doi 10.1038/ng.2764.

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(2):371-85 e18 doi 10.1016/j.cell.2018.02.060.

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(7457):214-8 doi 10.1038/nature12213.

10. Greaves M, Maley CC. Clonal evolution in cancer. Nature **2012**;481(7381):306-13 doi 10.1038/nature10762.

11. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature **2009**;458(7239):719-24 doi 10.1038/nature07943.

12. Winters IP, Murray CW, Winslow MM. Towards quantitative and multiplexed in vivo functional cancer genomics. Nat Rev Genet **2018**;19(12):741-55 doi 10.1038/s41576-018-0053-7.

13. Zahir N, Sun R, Gallahan D, Gatenby RA, Curtis C. Characterizing the ecological and evolutionary dynamics of cancer. Nat Genet **2020** doi 10.1038/s41588-020-0668-4.

14. Ben-David U, Beroukhim R, Golub TR. Genomic evolution of cancer models: perils and opportunities. Nat Rev Cancer **2019**;19(2):97-109 doi 10.1038/s41568-018-0095-3.

15. Graham TA, Sottoriva A. Measuring cancer evolution from the genome. J Pathol **2017**;241(2):183-91 doi 10.1002/path.4821.
16. McGranahan N, Swanton C. Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. Cancer Cell **2015**;27(1):15-26 doi 10.1016/j.ccell.2014.12.001.
17. Garraway LA, Lander ES. Lessons from the cancer genome. Cell **2013**;153(1):17-37 doi 10.1016/j.cell.2013.03.002.
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 doi 10.1101/sqb.2016.81.031070.
19. Friedman AA, Letai A, Fisher DE, Flaherty KT. Precision medicine for cancer with next-generation functional diagnostics. Nat Rev Cancer **2015**;15(12):747-56 doi 10.1038/nrc4015.
20. Weber J, Braun CJ, Saur D, Rad R. In vivo functional screening for systems-level integrative cancer genomics. Nat Rev Cancer **2020** doi 10.1038/s41568-020-0275-9.
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(2):137-53 doi 10.15252/emmm.201606857.
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(7531):428-31 doi 10.1038/nature13906.
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(12):1470-80 doi 10.1101/gad.279190.116.

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(14):1576-85 doi 10.1101/gad.264861.115.

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(7522):380-4 doi 10.1038/nature13589.

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(7):737-42 doi 10.1038/nmeth.4297.

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(4):483-6 doi 10.1038/s41588-018-0083-2.

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(1):2053 doi 10.1038/s41467-017-01519-y.

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(22):3561-8 doi 10.1093/bioinformatics/btv430.

30. The Gene Ontology C. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res **2019**;47(D1):D330-D8 doi 10.1093/nar/gky1055.

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(4):1119-27 doi 10.1158/0008-5472.CAN-07-3117.

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(7155):807-10 doi 10.1038/nature06030.

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(7):1719-29 doi 10.1158/0008-5472.CAN-16-2159.

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(6) doi 10.15252/emmm.201809856.

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(11):4756-68 doi 10.1093/nar/gkr038.

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(7413):665-9 doi 10.1038/nature11308.

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(31):12649-54 doi 10.1073/pnas.1307218110.

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(24):7548-50 doi 10.1073/pnas.1508113112.

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(12):1464-9 doi 10.1038/ng.2799.

40. Romero-Perez L, Surdez D, Brunet E, Delattre O, Grunewald TGP. STAG Mutations in Cancer. Trends Cancer **2019**;5(8):506-20 doi 10.1016/j.trecan.2019.07.001.

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(12):1428-30 doi 10.1038/ng.2800.

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(5):682-96 e8 doi 10.1016/j.stem.2019.08.003.

43. Kleyman M, Kabeche L, Compton DA. STAG2 promotes error correction in mitosis by regulating kinetochore-microtubule attachments. J Cell Sci **2014**;127(Pt 19):4225-33 doi 10.1242/jcs.151613.

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(6045):1039-43 doi 10.1126/science.1203619.

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(4):685-98 doi 10.1128/MCB.01503-13.

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(1):1686 doi 10.1038/s41467-019-09659-z.

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(9):1056-61 doi 10.1038/nm.4155.

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(1):1485 doi 10.1038/s41467-018-03782-z.

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(9):648-57 doi 10.1038/ncb839.

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 doi 10.1146/annurev-biochem-051710-134100.

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(17):E3978-E86 doi 10.1073/pnas.1716589115.

52. Hill BM. A Simple General Approach to Inference About the Tail of a Distribution. The Annals of Statistics **1975**;3(5):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(22):10280-8 doi 10.1158/0008-5472.CAN-05-2193.

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(7323):572-5 doi 10.1038/nature09535.

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(6832):1111-6 doi 10.1038/35074129.

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(44):18061-6 doi 10.1073/pnas.1114946108.

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(6):912-23 doi 10.1158/1541-7786.MCR-13-0620-T.

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(3):512-26 e8 doi 10.1016/j.ccell.2018.02.003.

59. Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS*, et al.* Defining a Cancer Dependency Map. Cell **2017**;170(3):564-76 e16 doi 10.1016/j.cell.2017.06.010.

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(7484):495-501 doi 10.1038/nature12912.

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(7793):102-11 doi 10.1038/s41586-020-1965-x.

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(7793):94-101 doi 10.1038/s41586-020-1943-3.

63. Curry NL, Mino-Kenudson M, Oliver TG, Yilmaz OH, Yilmaz 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(8):908-21 doi 10.1158/2159-8290.CD-12-0507.

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(10):2929-40 doi 10.1172/JCI31854.

65. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. Genome Res **2012**;22(2):398-406 doi 10.1101/gr.125567.111.

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(6):1375-83 doi 10.1016/j.cell.2019.05.005.

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**.

68. Foggetti G, Li C, Cai H, Hellyer JA, Lin WY, Ayeni D*, et al.* Genetic determinants of EGFR-Driven Lung Cancer Growth and Therapeutic Response In Vivo. bioRxiv **2020**.

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(3):271-81 e7 doi 10.1016/j.cels.2018.03.002.

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(283):283ra54 doi 10.1126/scitranslmed.aaa1408.

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(D1):D941-D7 doi 10.1093/nar/gky1015.

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(1):167 doi 10.1186/s12859-017-1581-4.

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(2):184-91 doi 10.1038/nbt.3437.

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(9):827-32 doi 10.1038/nbt.2647.

75. Bae S, Kweon J, Kim HS, Kim JS. Microhomology-based choice of Cas9 nuclease target sites. Nat Methods **2014**;11(7):705-6 doi 10.1038/nmeth.3015.

76. 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(3):224-6 doi 10.1038/s41587-019-0032-3.

77. 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(3):291-300 doi 10.1038/nm.4285.

## FIGURE LEGENDS

**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 p53 which is targeted by three sgRNAs. 5 Inert sgRNAs are either non-targeting (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 2-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 trigger moderate but significant increase and green when one sgRNA triggers >4 fold increase and the other triggers moderate but significant increase. Data is aggregated from 47 *KT;H11*$^{LSL\text{-}Cas9}$ and 12 *KT* mice.

**Figure 2. *In vivo* lung tumor growth is suppressed by diverse tumor suppressor genes.**

(A) The 95$^{th}$ 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. 95% confidence intervals and *P*-values were calculated by bootstrap. 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 95$^{th}$ percentile tumor sizes from

all sgRNAs targeting that gene, individual sgRNAs targeting each gene were 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 95$^{th}$ 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 is shown in **Supplementary Fig. S5**. 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. 95% confidence intervals and *P*-values were calculated by bootstrap. 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 **Fig. 2A**. The high Pearson's 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*$^{LSL\text{-}Cas9}$.

**Figure 3. Stag2, inactivation of which increases tumor burden and reduces survival, is frequently lowly expressed in human lung adenocarcinoma.**

(A) Cre/loxP-mediated Stag2 inactivation promotes *Kras*$^{G12D}$-driven lung tumor growth. Lung tumors were initiated in 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 indicated *KT;Stag2*$^{flox/flox}$ mice lack Stag2 protein expression. Scale bar = 50 mm.

(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's 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 x 100) on H&E-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's t-test.

(G) Survival curve of mice with KrasG12D-driven lung tumors that are either 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.

**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) which excludes 8 tumor suppressor genes (in gray and crossed out) from the Lenti-sg*TS102*/Cre pool whose losses 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,D) The impact of inactivating each gene on the size of the 95$^{th}$ 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) Heatmap of the tumor suppressive effects of six genes that emerge as suppressors of tumor growth at the later timepoint. 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*-targeted 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 is shown in **Supplementary Fig. S9**.

**Figure 5. Tumor initiation is inhibited by diverse tumor suppressor genes independent of their effects on tumor growth.**

(A,C) 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 (A) and 26 weeks after tumor initiation with Lenti-sg*TS85*/Cre in *KT;H11*$^{LSL\text{-}Cas9}$ mice (C) is shown. Error bars indicate 95% confidence intervals. 95% confidence intervals and *P*-values were calculated by bootstrap. 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 95$^{th}$ percentile) and tumor number can be independent aspects of tumor suppression.

(D) 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.

(E,F) 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/IMPACT (E) as well as TCGA and TRACERx (F). Data from GENIE/IMPACT are based on panel sequencing and therefore does not include data on NCOA6. Data from TRACERx are from multi-region sequencing where we report the number of tumors that had any of these four genes mutated in one or more regions.

**Figure 6. Loss of *p53*, *Cdkn2a* and *Dnmt3a* result in 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=1160) are shown for each sgRNA.

(B) Volcano plot of the impact of inactivating each putative tumor suppressor gene on the distribution of tumor sizes (Hill's estimator compares tumors above the 95$^{th}$ percentile to those at the 95$^{th}$ percentile to quantify the relative size of tumors in the tail of the distribution). *P53*- 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. Equal number of tumors (814 tumors randomly sampled) are shown for each sgRNA.

(D) Volcano plot of the impact of inactivating each putative tumor suppressor gene on the developing of infrequent exceptionally large tumors (Hill's 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* generate 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 is aggregated from all tumors from 40 *KT;H11*$^{LSL\text{-}Cas9}$ (26 weeks) mice.

**Figure 7. Tumor suppressors constrain tumorigenesis at different stages and to different levels.**

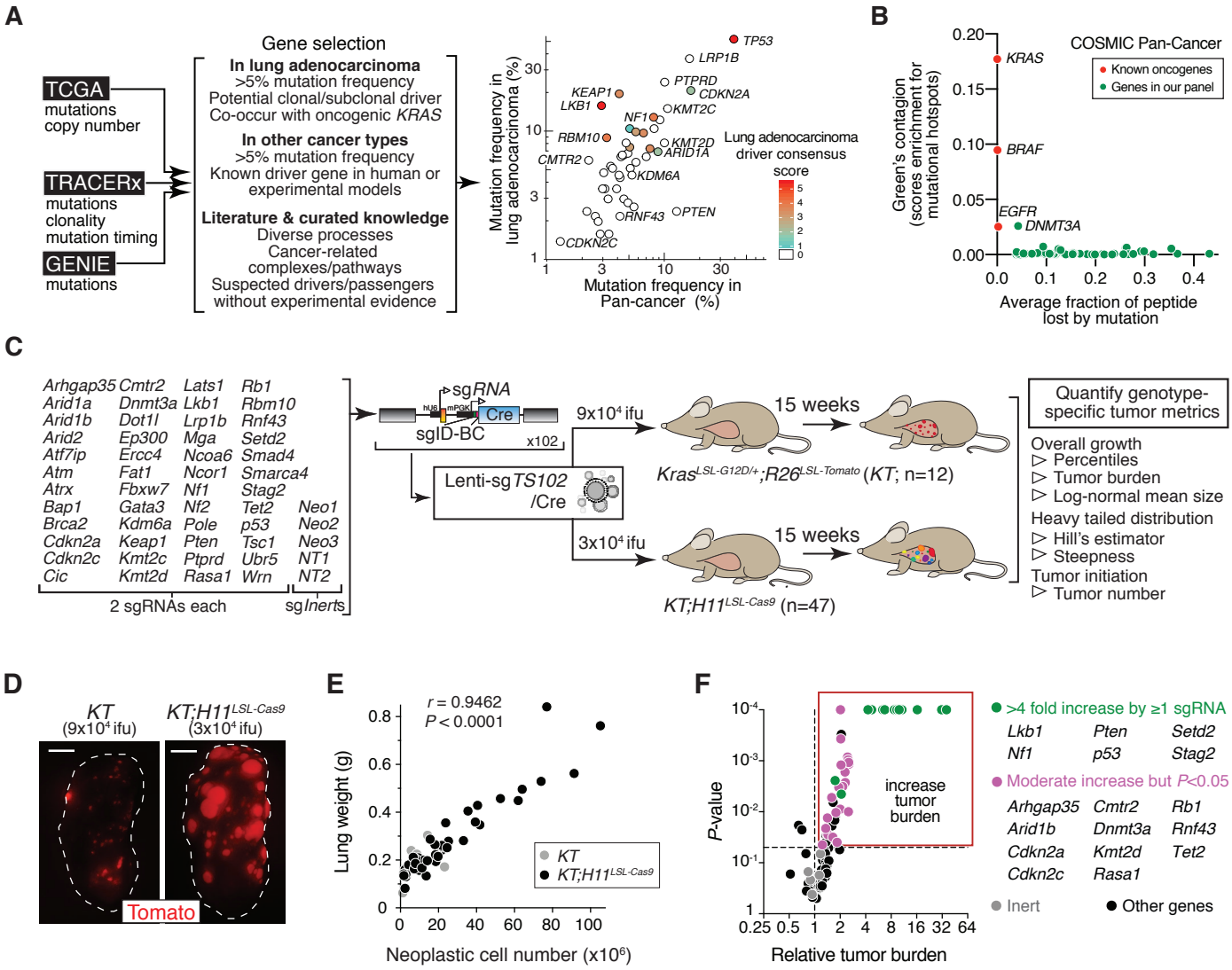
(A) Radar plots of representative genes whose inactivation affects tumor size at the 95$^{th}$ percentile (relative to sg*Inert*s, indicating increased overall growth), tumor number (relative to sg*Inert*s, indicating increased tumor initiation) and Hill's estimator (relative to sg*Inert*s, indicating increased rare large tumors). Tumor suppressors suppress different aspects of tumor development.

(B) Heatmap summarizing the tumor size at the 95$^{th}$ percentile (relative to sg*Inert*s), tumor number (relative to sg*Inert*s) and Hill's 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

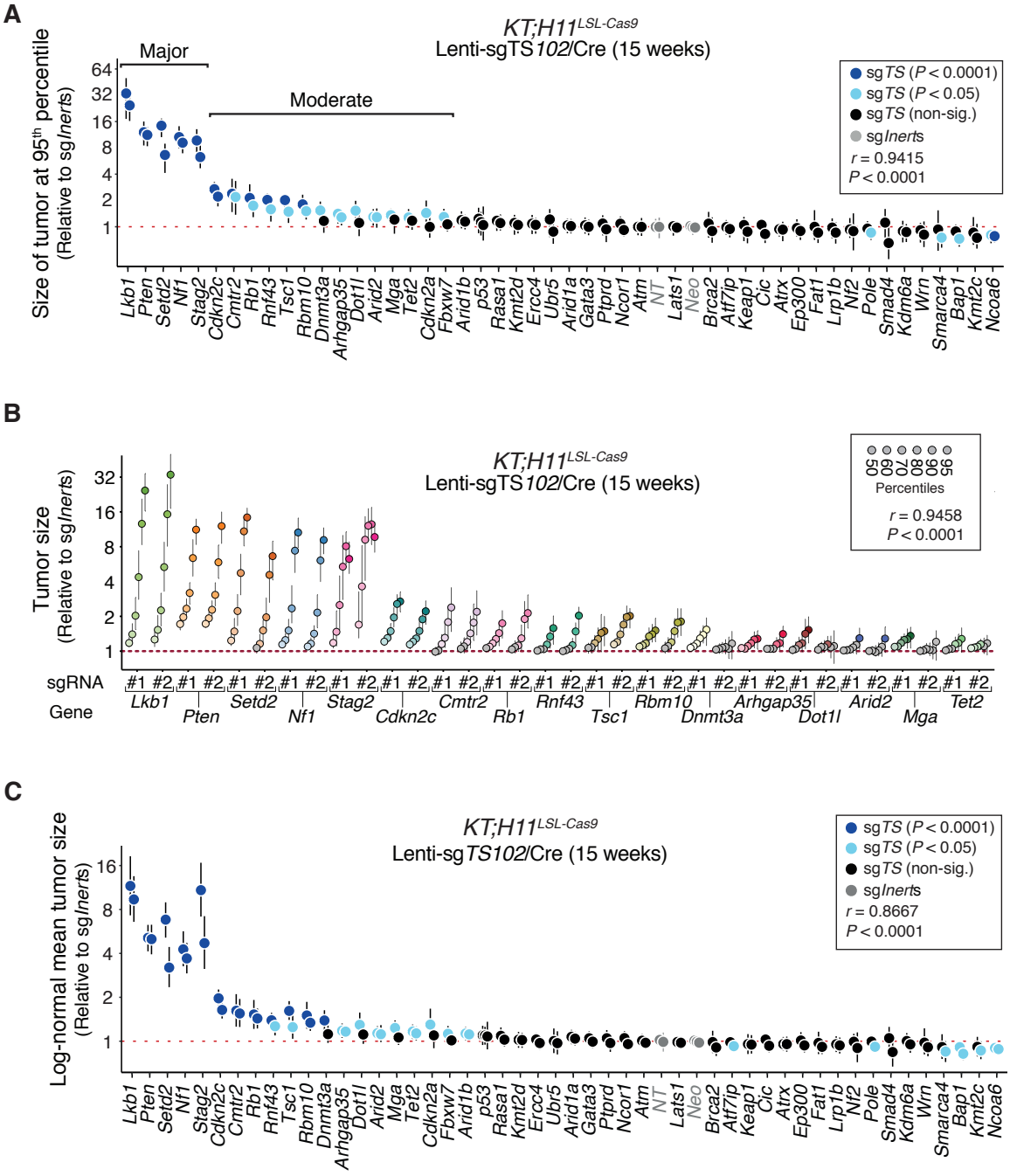
control tumors at 95$^{th}$ percentile, they are significantly greater across multiple percentiles at 26 weeks, and thus they are also considered genes that suppress tumor growth.

(C) Summary schematic of a tumor suppression map in lung adenocarcinoma based on our data.

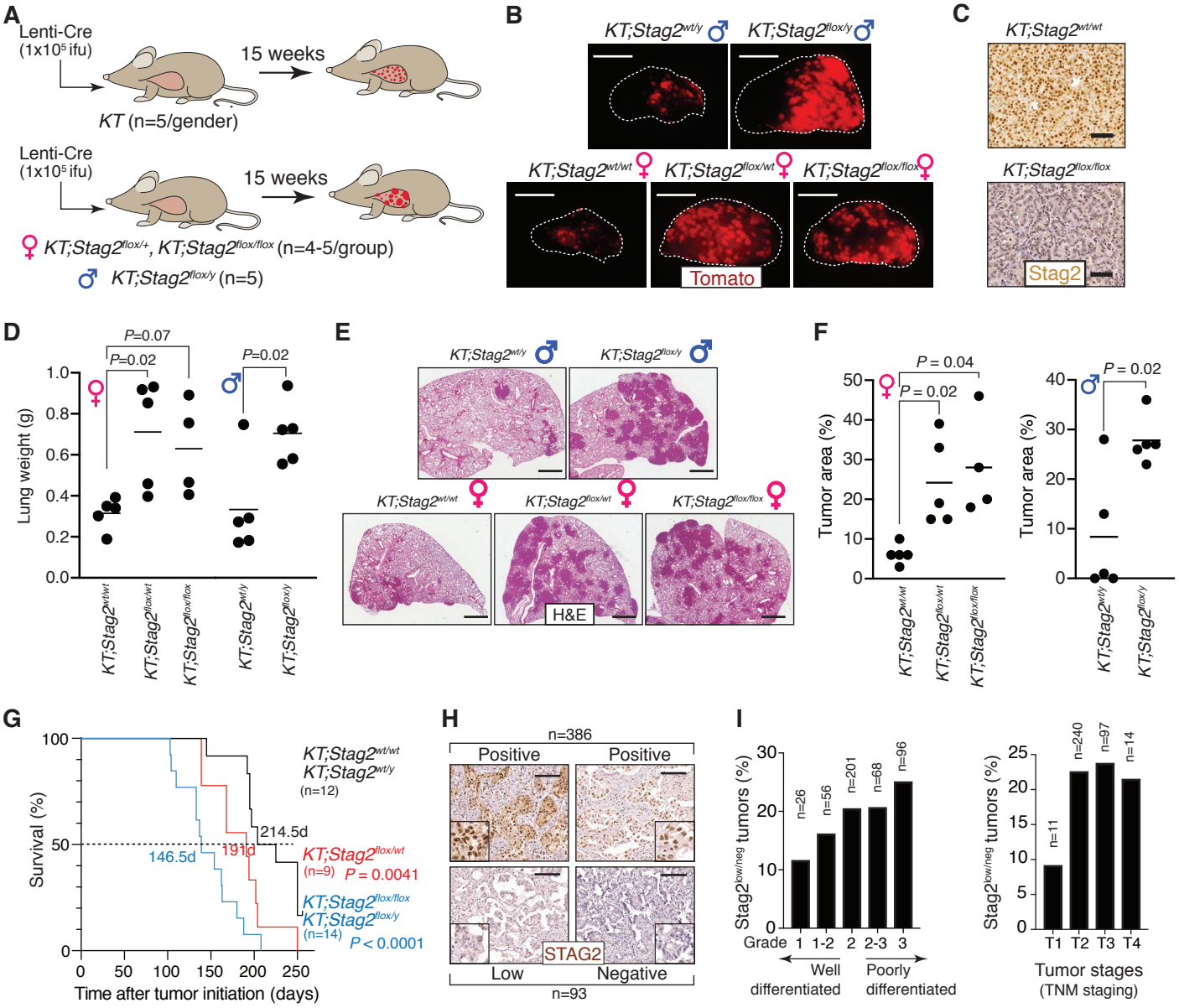
# Figure 1


A
Gene selection
TCGA
mutations
copy number
TRACERx
mutations
clonality
mutation timing
GENIE
mutations
In lung adenocarcinoma
>5% mutation frequency
Potential clonal/subclonal driver
Co-occur with oncogenic KRAS
In other cancer types
>5% mutation frequency
Known driver gene in human or experimental models
Literature & curated knowledge
Diverse processes
Cancer-related complexes/pathways
Suspected drivers/passengers without experimental evidence
Mutation frequency in lung adenocarcinoma (%)
Mutation frequency in Pan-cancer (%)
TP53
LRP1B
PTPRD
CDKN2A
KEAP1
LKB1
KMT2C
NF1
RBM10
KMT2D
ARID1A
CMTR2
KDM6A
RNF43
PTEN
CDKN2C
Lung adenocarcinoma driver consensus score
B
COSMIC Pan-Cancer
Known oncogenes
Genes in our panel
Green's contagion (scores enrichment for mutational hotspots)
KRAS
BRAF
EGFR
DNMT3A
Average fraction of peptide lost by mutation
C
Arhgap35 Cmtr2 Lats1 Rb1
Arid1a Dnmt3a Lkb1 Rbm10
Arid1b Dot1l Lrp1b Rnf43
Arid2 Ep300 Mga Setd2
Atf7ip Ercc4 Ncoa6 Smad4
Atm Fat1 Ncor1 Smarca4
Atrx Fbxw7 Nf1 Stag2
Bap1 Gata3 Nf2 Tet2 Neo1
Brca2 Kdm6a Pole p53 Neo2
Cdkn2a Keap1 Pten Tsc1 Neo3
Cdkn2c Kmt2c Ptprd Ubr5 NT1
Clc Kmt2d Rasa1 Wrn NT2
2 sgRNAs each
sgInerts
sgRNA
sgID-BC
Cre
x102
Lenti-sgTS102 /Cre
9x10⁴ ifu
3x10⁴ ifu
15 weeks
15 weeks
Kras^LSL-G12D/+;R26^LSL-Tomato (KT; n=12)
KT;H11^LSL-Cas9 (n=47)
Quantify genotype-specific tumor metrics
Overall growth
Percentiles
Tumor burden
Log-normal mean size
Heavy tailed distribution
Hill's estimator
Steepness
Tumor initiation
Tumor number
D
KT (9x10⁴ ifu)
KT;H11^LSL-Cas9 (3x10⁴ ifu)
Tomato
E
r = 0.9462
P < 0.0001
Lung weight (g)
KT
KT;H11^LSL-Cas9
Neoplastic cell number (x10⁶)
F
P-value
increase tumor burden
Relative tumor burden
>4 fold increase by ≥1 sgRNA
Lkb1 Pten Setd2
Nf1 p53 Stag2
Moderate increase but P<0.05
Arhgap35 Cmtr2 Rb1
Arid1b Dnmt3a Rnf43
Cdkn2a Kmt2d Tet2
Cdkn2c Rasa1
Inert
Other genes

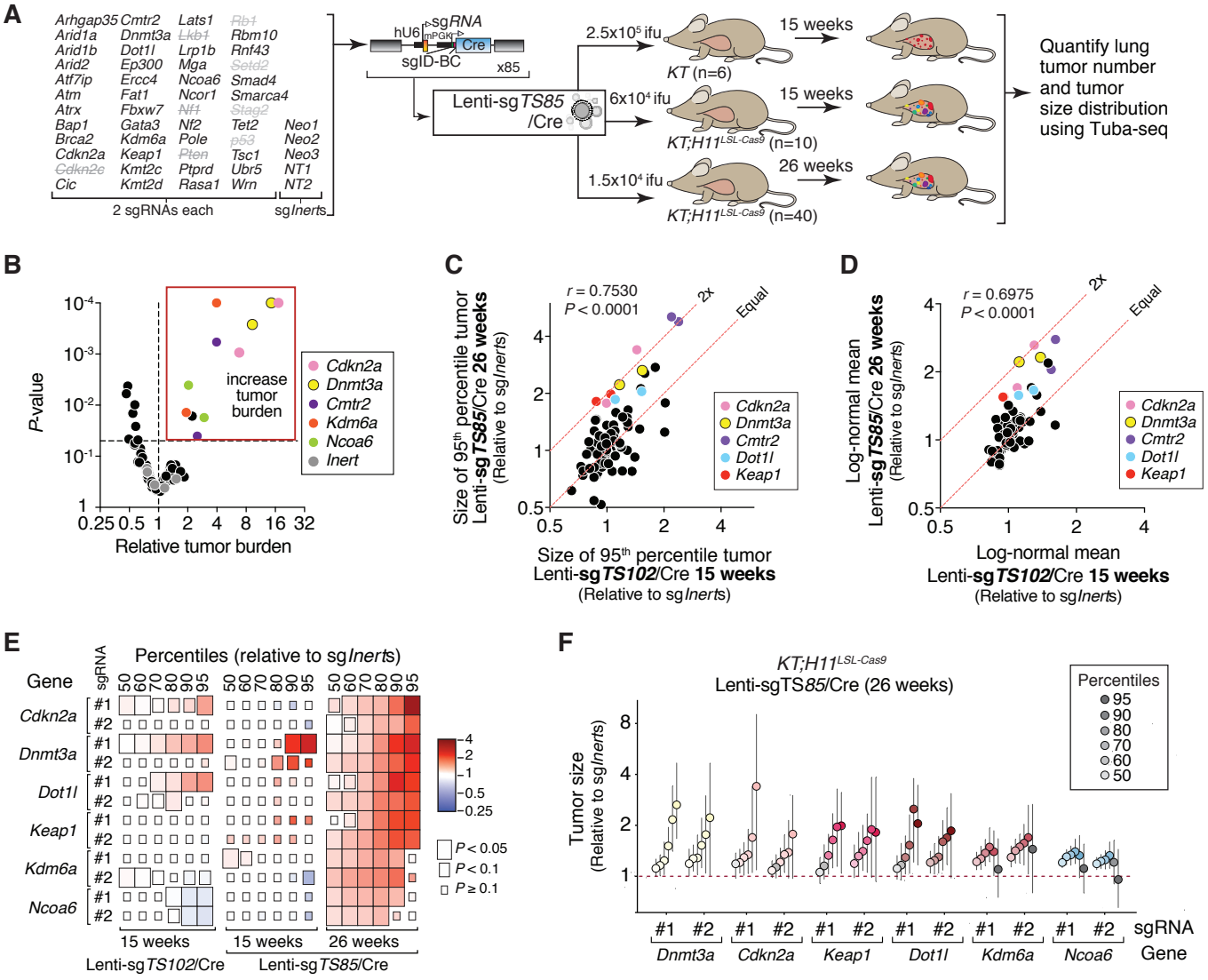
## Figure 2


A
KT;H11LSL-Cas9
Lenti-sgTS102/Cre (15 weeks)
Major
Moderate
Size of tumor at 95th percentile
(Relative to sgInerts)
sgTS (P < 0.0001)
sgTS (P < 0.05)
sgTS (non-sig.)
sgInerts
r = 0.9415
P < 0.0001
Lkb1 Pten Setd2 Nf1 Stag2 Cdkn2c Cmtr2 Rb1 Rnf43 Tsc1 Rbm10 Dnmt3a Arhgap35 Dot1l Arid2 Mga Tet2 Cdkn2a Fbxw7 Arid1b p53 Rasa1 Kmt2d Ercc4 Ubr5 Arid1a Gata3 Ptprd Ncor1 Atm NT Lats1 Neo Brca2 Atf7ip Keap1 Cic Atrx Ep300 Fat1 Lrp1b Nf2 Pole Smad4 Kdm6a Wrn Smarca4 Bap1 Kmt2c Ncoa6
B
KT;H11LSL-Cas9
Lenti-sgTS102/Cre (15 weeks)
Tumor size
(Relative to sgInerts)
05 09 70 08 06 96
Percentiles
r = 0.9458
P < 0.0001
sgRNA
Gene
C
KT;H11LSL-Cas9
Lenti-sgTS102/Cre (15 weeks)
Log-normal mean tumor size
(Relative to sgInerts)
sgTS (P < 0.0001)
sgTS (P < 0.05)
sgTS (non-sig.)
sgInerts
r = 0.8667
P < 0.0001

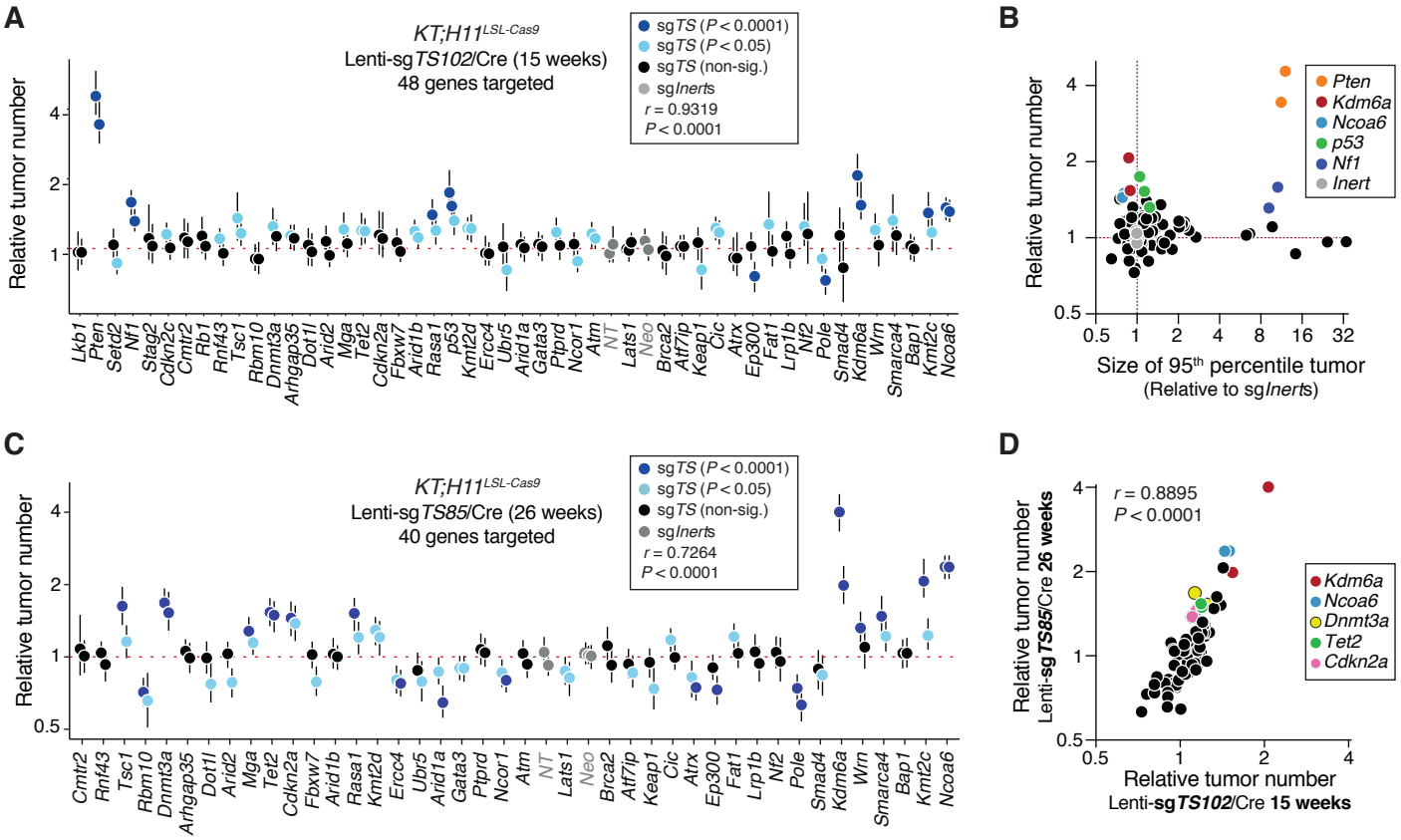
# Figure 3


A
Lenti-Cre
(1x10⁵ ifu)
15 weeks
KT (n=5/gender)
Lenti-Cre
(1x10⁵ ifu)
15 weeks
♀ KT;Stag2flox/+, KT;Stag2flox/flox (n=4-5/group)
♂ KT;Stag2flox/y (n=5)
B
KT;Stag2wt/y ♂
KT;Stag2flox/y ♂
KT;Stag2wt/wt ♀
KT;Stag2flox/wt ♀
KT;Stag2flox/flox ♀
Tomato
C
KT;Stag2wt/wt
KT;Stag2flox/flox
Stag2
D
P=0.07
P=0.02
P=0.02
Lung weight (g)
1.0
0.8
0.6
0.4
0.2
0.0
KT;Stag2wt/wt
KT;Stag2flox/wt
KT;Stag2flox/flox
KT;Stag2wt/y
KT;Stag2flox/y
E
KT;Stag2wt/y ♂
KT;Stag2flox/y ♂
KT;Stag2wt/wt ♀
KT;Stag2flox/wt ♀
KT;Stag2flox/flox ♀
H&E
F
P = 0.04
P = 0.02
Tumor area (%)
50
40
30
20
10
0
KT;Stag2wt/wt
KT;Stag2flox/wt
KT;Stag2flox/flox
P = 0.02
Tumor area (%)
40
30
20
10
0
KT;Stag2wt/y
KT;Stag2flox/y
G
Survival (%)
100
50
0
0
50
100
150
200
250
Time after tumor initiation (days)
KT;Stag2wt/wt
KT;Stag2wt/y
(n=12)
214.5d
146.5d
191d
KT;Stag2flox/wt
(n=9)
P = 0.0041
KT;Stag2flox/flox
KT;Stag2flox/y
(n=14)
P < 0.0001
H
n=386
Positive
Positive
STAG2
Low
Negative
n=93
I
Stag2low/neg tumors (%)
30
20
10
0
n=26
n=56
n=201
n=68
n=96
Grade 1 1-2 2 2-3 3
Well differentiated
Poorly differentiated
Stag2low/neg tumors (%)
25
20
15
10
5
0
n=11
n=240
n=97
n=14
T1 T2 T3 T4
Tumor stages
(TNM staging)

# Figure 4


A
Arhgap35 Arid1a Arid1b Arid2 Atf7ip Atm Atrx Bap1 Brca2 Cdkn2a ~~Cdkn2c~~ Cic
Cmtr2 Dnmt3a Dot1l Ep300 Ercc4 Fat1 Fbxw7 Gata3 Kdm6a Keap1 Kmt2c Kmt2d
Lats1 ~~Lkb1~~ Lrp1b Mga Ncoa6 Ncor1 ~~Nf1~~ Nf2 Pole ~~Pten~~ Ptprd Rasa1
~~Rb1~~ Rbm10 Rnf43 ~~Setd2~~ Smad4 Smarca4 ~~Stag2~~ Tet2 ~~p53~~ Tsc1 Ubr5 Wrn
Neo1 Neo2 Neo3 NT1 NT2
2 sgRNAs each
sgInerts
hU6
sgRNA
mPGK
Cre
sgID-BC
x85
Lenti-sgTS85 /Cre
2.5x10^5 ifu
6x10^4 ifu
1.5x10^4 ifu
KT (n=6)
KT;H11LSL-Cas9 (n=10)
KT;H11LSL-Cas9 (n=40)
15 weeks
15 weeks
26 weeks
Quantify lung tumor number and tumor size distribution using Tuba-seq
B
P-value
increase tumor burden
Cdkn2a
Dnmt3a
Cmtr2
Kdm6a
Ncoa6
Inert
Relative tumor burden
C
r = 0.7530
P < 0.0001
2x
Equal
Size of 95th percentile tumor Lenti-sgTS85/Cre 26 weeks (Relative to sgInerts)
Cdkn2a
Dnmt3a
Cmtr2
Dot1l
Keap1
Size of 95th percentile tumor Lenti-sgTS102/Cre 15 weeks (Relative to sgInerts)
D
r = 0.6975
P < 0.0001
2x
Equal
Log-normal mean Lenti-sgTS85/Cre 26 weeks (Relative to sgInerts)
Cdkn2a
Dnmt3a
Cmtr2
Dot1l
Keap1
Log-normal mean Lenti-sgTS102/Cre 15 weeks (Relative to sgInerts)
E
Gene
sgRNA
Percentiles (relative to sgInerts)
Cdkn2a
Dnmt3a
Dot1l
Keap1
Kdm6a
Ncoa6
P < 0.05
P < 0.1
P ≥ 0.1
15 weeks
15 weeks
26 weeks
Lenti-sgTS102/Cre
Lenti-sgTS85/Cre
F
KT;H11LSL-Cas9
Lenti-sgTS85/Cre (26 weeks)
Percentiles
Tumor size (Relative to sgInerts)
sgRNA
Gene
Dnmt3a
Cdkn2a
Keap1
Dot1l
Kdm6a
Ncoa6

# Figure 5



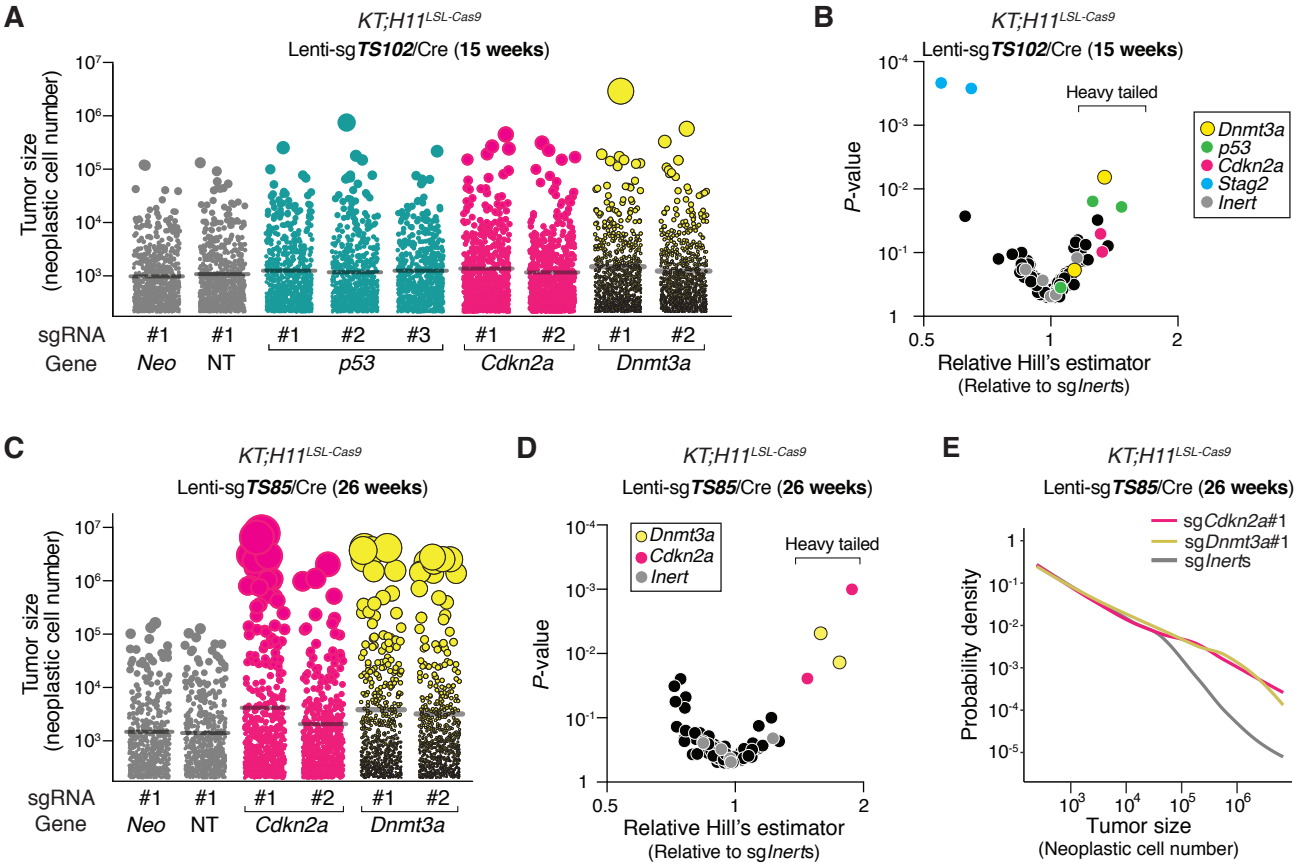

**E**

| | *KMT2C, KMT2D,* and/or *KDM6A* mutant |
|---|---|
| GENIE/IMPACT | 469/5271 (8.9%) |

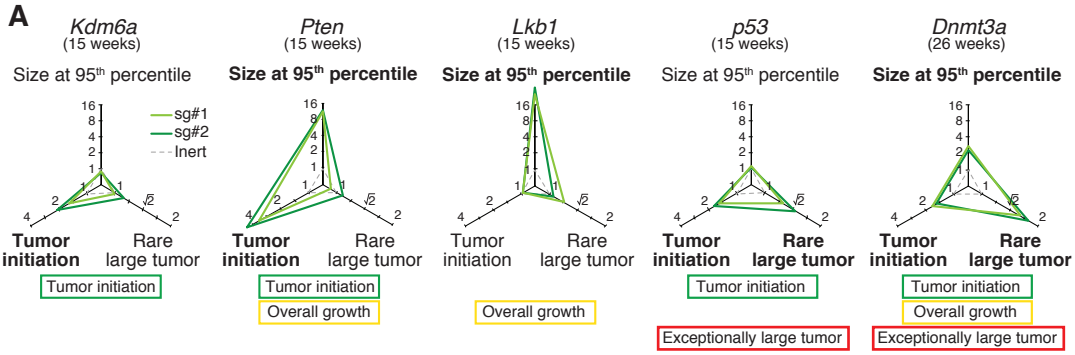
**F**

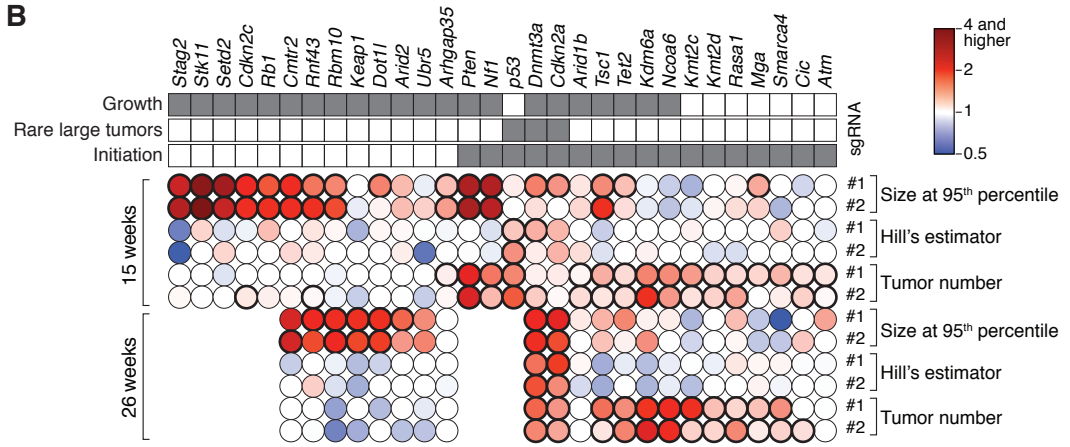
| | *KMT2C, KMT2D, KDM6A* and/or *NCOA6* mutant |
|---|---|
| TCGA | 100/388 (32.5%) |
| TRACERx | 13/61 (21.3%) |

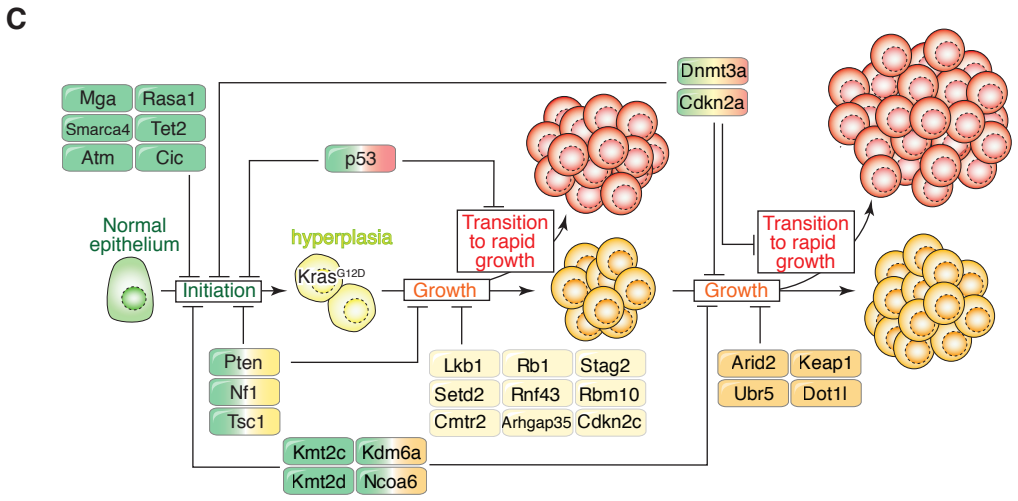
# Figure 6


A
KT;H11LSL-Cas9
Lenti-sgTS102/Cre (15 weeks)
Tumor size (neoplastic cell number)
sgRNA
Gene
#1 #1 #1 #2 #3 #1 #2 #1 #2
Neo NT p53 Cdkn2a Dnmt3a
B
KT;H11LSL-Cas9
Lenti-sgTS102/Cre (15 weeks)
Heavy tailed
Dnmt3a
p53
Cdkn2a
Stag2
Inert
P-value
Relative Hill's estimator
(Relative to sgInerts)
C
KT;H11LSL-Cas9
Lenti-sgTS85/Cre (26 weeks)
Tumor size (neoplastic cell number)
sgRNA
Gene
#1 #1 #1 #2 #1 #2
Neo NT Cdkn2a Dnmt3a
D
KT;H11LSL-Cas9
Lenti-sgTS85/Cre (26 weeks)
Dnmt3a
Cdkn2a
Inert
Heavy tailed
P-value
Relative Hill's estimator
(Relative to sgInerts)
E
KT;H11LSL-Cas9
Lenti-sgTS85/Cre (26 weeks)
sgCdkn2a#1
sgDnmt3a#1
sgInerts
Probability density
Tumor size
(Neoplastic cell number)

# Figure 7


A
Kdm6a
(15 weeks)
Size at 95th percentile
sg#1
sg#2
Inert
Tumor initiation
Rare large tumor
Tumor initiation
Pten
(15 weeks)
Size at 95th percentile
Tumor initiation
Rare large tumor
Tumor initiation
Overall growth
Lkb1
(15 weeks)
Size at 95th percentile
Tumor initiation
Rare large tumor
Overall growth
p53
(15 weeks)
Size at 95th percentile
Tumor initiation
Rare large tumor
Tumor initiation
Exceptionally large tumor
Dnmt3a
(26 weeks)
Size at 95th percentile
Tumor initiation
Rare large tumor
Tumor initiation
Overall growth
Exceptionally large tumor


B
Stag2
Stk11
Setd2
Cdkn2c
Rb1
Cmtr2
Rnf43
Rbm10
Keap1
Dot1l
Arid2
Ubr5
Arhgap35
Pten
Nf1
p53
Dnmt3a
Cdkn2a
Arid1b
Tsc1
Tet2
Kdm6a
Ncoa6
Kmt2c
Kmt2d
Rasa1
Mga
Smarca4
Cic
Atm
Growth
Rare large tumors
Initiation
sgRNA
4 and higher
2
1
0.5
15 weeks
26 weeks
#1
#2
Size at 95th percentile
Hill's estimator
Tumor number
Size at 95th percentile
Hill's estimator
Tumor number


C
Mga
Rasa1
Smarca4
Tet2
Atm
Cic
p53
Dnmt3a
Cdkn2a
Normal epithelium
Initiation
hyperplasia
KrasG12D
Transition to rapid growth
Growth
Transition to rapid growth
Growth
Pten
Nf1
Tsc1
Lkb1
Rb1
Stag2
Setd2
Rnf43
Rbm10
Cmtr2
Arhgap35
Cdkn2c
Arid2
Keap1
Ubr5
Dot1l
Kmt2c
Kdm6a
Kmt2d
Ncoa6