

ARTICLE

A STAG2-PAXIP1/PAGR1 axis suppresses lung tumorigenesis

Emily L. Ashkin¹ , Yuning J. Tang^{2*} , Haiqing Xu^{3*} , King L. Hung^{1,5} , Julia A. Belk^{5,6} , Hongchen Cai² , Steven S. Lopez² , Deniz Nesli Dolcen² , Jess D. Hebert² , Rui Li^{5,6} , Paloma A. Ruiz² , Tula Keal⁴ , Laura Andrejka² , Howard Y. Chang^{5,6} , Dmitri A. Petrov^{1,3,7} , Jesse R. Dixon⁴ , Zhichao Xu⁴ , and Monte M. Winslow^{1,2,8}

The cohesin complex is a critical regulator of gene expression. STAG2 is the most frequently mutated cohesin subunit across several cancer types and is a key tumor suppressor in lung cancer. Here, we coupled somatic CRISPR-Cas9 genome editing and tumor barcoding with an autochthonous oncogenic KRAS-driven lung cancer model and showed that STAG2 is uniquely tumor-suppressive among all core and auxiliary cohesin components. The heterodimeric complex components PAXIP1 and PAGR1 have highly correlated effects with STAG2 in human lung cancer cell lines, are tumor suppressors *in vivo*, and are epistatic to STAG2 in oncogenic KRAS-driven lung tumorigenesis *in vivo*. STAG2 inactivation elicits changes in gene expression, chromatin accessibility, and 3D genome conformation that impact the cancer cell state. Gene expression and chromatin accessibility similarities between STAG2- and PAXIP1-deficient neoplastic cells further relate STAG2-cohesin to PAXIP1/PAGR1. These findings reveal a STAG2-PAXIP1/PAGR1 tumor-suppressive axis and uncover novel PAXIP1-dependent and PAXIP1-independent STAG2-cohesin-mediated mechanisms of lung tumor suppression.

Introduction

The cohesin complex is a large multisubunit complex that plays important roles in regulating genome conformation, gene expression, and cell fate (Ren et al., 2017). There are two classes of cohesin complexes distinguished by whether they contain the paralogs STAG1 or STAG2 (Alonso-Gil and Losada, 2023). STAG1- and STAG2-cohesin are thought to control different gene expression programs (Arruda et al., 2020; van der Lelij et al., 2017; Cuadrado et al., 2019). STAG1-cohesin colocalizes with CTCF to form topologically associating domain boundaries that drive long-range chromatin interactions for gene transcription (Cuadrado et al., 2019; Viny et al., 2019). Conversely, STAG2-cohesin has been shown to drive mid-range chromatin interactions (Cuadrado et al., 2019; Richart et al., 2021); however, it remains unknown whether additional proteins contribute specifically to STAG2-cohesin-regulated gene expression (Bailey et al., 2021; De Koninck et al., 2020).

Notably, STAG2 is the most frequently mutated cohesin complex member across many cancer types, including bladder cancer, Ewing's sarcoma, acute myeloid leukemia (AML), and lung adenocarcinoma (De Koninck et al., 2020; Solomon et al.,

2013; Taylor et al., 2014; Aquila et al., 2018; Kim et al., 2012; Adane et al., 2021; Evers et al., 2014; Bailey et al., 2014; Ball et al., 2014; Crompton et al., 2014). Using *in vivo* CRISPR/Cas9 screens within genetically engineered mouse models of lung cancer, we and others identified STAG2 as an important tumor suppressor in lung adenocarcinoma (Cai et al., 2021; Blair et al., 2023). STAG2 mutations occur in ~4% of human lung adenocarcinomas, and STAG2 protein expression is low or absent in ~20% of these tumors (Cai et al., 2021). In lung cancer, STAG2-mediated tumor suppression is not through an impact on chromosomal instability or hindering MAPK signaling, as previously suggested for other cancer types (Solomon et al., 2013; Cai et al., 2021; Kong et al., 2014; Mondal et al., 2019; Ding et al., 2018; Shen et al., 2016; Kim et al., 2019a). Thus, the mechanisms through which STAG2 inactivation drives lung tumor growth remain largely unknown and crucial for understanding cellular processes that contribute to tumor suppression.

Here, using CRISPR/Cas9-mediated somatic genome editing within genetically engineered mouse models of lung adenocarcinoma, we investigated the tumor-suppressive capacity of

¹Cancer Biology Program, Stanford University School of Medicine, Stanford, CA, USA; ²Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA; ³Department of Biology, Stanford University, Stanford, CA, USA; ⁴Gene Expression Laboratory, Salk Institute for Biological Studies, La Jolla, CA, USA; ⁵Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA; ⁶Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA; ⁷Chan Zuckerberg Biohub, San Francisco, CA, USA; ⁸Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.

*Y.J. Tang and H. Xu contributed equally to this paper. Correspondence to Monte M. Winslow: mwinslow@stanford.edu.

© 2024 Ashkin et al. This article is distributed under the terms of an Attribution–Noncommercial–Share Alike–No Mirror Sites license for the first six months after the publication date (see <http://www.rupress.org/terms/>). After six months it is available under a Creative Commons License (Attribution–Noncommercial–Share Alike 4.0 International license, as described at <https://creativecommons.org/licenses/by-nc-sa/4.0/>).

STAG2, every other component of the cohesin complex, and potential co-operators of STAG2-cohesin in lung cancer. We identified PAXIP1 and PAGR1, which form an obligatory heterodimer that has been suggested to be involved in DNA damage repair, cell cycle control, and transcription regulation (Liu and Li, 2022; Gong et al., 2009; Grimley and Dressler, 2018), as potent lung tumor suppressors and provide multiple complementary lines of evidence for a STAG2-PAXIP1/PAGR1 tumor-suppressive axis.

Results

Lung tumor suppression by STAG2 is unique among cohesin subunits

STAG2 is a tumor suppressor in oncogenic KRAS-driven lung cancer, and other cohesin subunits are frequently mutated in KRAS-driven lung cancer as well as in several other cancer types (De Koninck et al., 2020; Solomon et al., 2013; Taylor et al., 2014; Aquila et al., 2018; Kim et al., 2012; Adane et al., 2021; Cai et al., 2021; Solomon et al., 2011; Gao et al., 2013; Cerami et al., 2012; de Bruijn et al., 2023). To determine whether the inactivation of other cohesin subunits would also increase autochthonous lung tumor growth, we used tumor barcoding and somatic CRISPR/Cas9-mediated genome editing (Cai et al., 2021; Rogers et al., 2017). We used a recently optimized version of tumor barcoding coupled with high-throughput BC (barcode) sequencing (Tuba-seq) in which lentiviral vectors have a diverse BC integrated within the U6 promoter directly 5' of the single-guide RNA (sgRNA) (Tuba-seq^{Ultra}; U6 BC labeling with per-tumor resolution analysis; Fig. 1, a and b and Materials and methods) (Tang et al., 2024, Preprint). Tuba-seq^{Ultra} enables the quantification of the size of each clonal tumor and the number of tumors with each sgRNA based on the amplification of the BC-sgRNA region from bulk tumor-bearing lungs followed by high-throughput sequencing.

We generated a lentiviral pool that contained barcoded lentiviral-sgRNA/Cre Tuba-seq^{Ultra} vectors targeting each cohesin gene, an essential gene, and safe-cutting “inert” sgRNAs (Lenti-U6^{BC}sgCohesin/Cre with three sgRNAs/gene; Fig. 1 b, Materials and methods, and Table S1). We initiated tumors with Lenti-U6^{BC}sgCohesin/Cre in *Kras*^{LSL-G12D/+};R26^{LSL-Tomato};H11^{LSL-Cas9} (KT; H11^{LSL-Cas9}) mice and Cas9-negative *Kras*^{LSL-G12D/+};R26^{LSL-Tomato} (KT) mice (Materials and methods). After 15 wk of tumor growth, we extracted DNA from bulk tumor-bearing lungs, PCR amplified the BC-sgRNA region, and high-throughput sequenced the amplicon. By tallying the number of reads from each BC-sgRNA and normalizing to “spike-in” control cells with a known BC-sgRNA that was added at a defined cell number, we quantified the number of neoplastic cells in each clonal tumor with each sgRNA (Materials and methods).

We calculated tumor number (the number of BCs associated with each sgRNA normalized to the expected number from Cas9-negative KT mice) and several metrics of tumor size (log-normal mean [LN mean] tumor size and tumor sizes at defined percentiles within the tumor size distribution; Materials and methods). Consistent with previous observations, *Stag2* inactivation greatly increased tumor size and modestly increased

tumor number (Fig. 1, c and d) (Cai et al., 2021; Blair et al., 2023). Inactivation of most of the other cohesin complex components greatly reduced tumor number and tumor size, although not to the same extent as inactivation of the known essential gene *Pcna* (Fig. 1, c–e). All three sgRNAs targeting *Stag1* only modestly reduced tumor number and tumor size. Collectively, these data are consistent with cohesin in general having an essential function, and the well-described ability of STAG2 can compensate for STAG1 in this essential function (Arruda et al., 2020; Viny et al., 2019; Kong et al., 2014; Canudas and Smith, 2009) (Fig. 1, c–e). Importantly, these results provided no indication that STAG1 is a tumor suppressor and are consistent with distinct roles of STAG1- and STAG2-cohesin in lung cancer.

Heterozygous inactivation of the cohesin subunit *Smc3* increases lung tumorigenesis

While homozygous inactivation of cohesin subunits is detrimental to lung tumorigenesis, mutations in these genes are common in lung cancer, and heterozygous inactivation of core cohesin genes increases tumorigenesis in other contexts (Rivas et al., 2021a, 2021b). Thus, we hypothesized that heterozygous inactivation of core cohesin subunits could increase lung tumor growth. We chose to focus on the core cohesion component SMC3 due to the availability of mice with a *Smc3* floxed allele (Viny et al., 2015). To generate tumors with inactivation on one allele of *Smc3*, we initiated lung tumors in *KT;Smc3*^{fl/fl} and KT control mice. After 14 wk of tumor growth, *KT;Smc3*^{fl/fl} mice had significantly greater tumor burden and higher tumor number compared with KT mice as assessed by total lung weight, fluorescent imaging, and histology (Fig. 1, f–i). These genetic data support a model in which the net effect of reducing cohesin is increased lung tumor growth potentially driven by trade-offs between a reduction in STAG2 cohesin-mediated tumor suppression and the essential function of cohesin in general.

Dependency Map (DepMap) and *in vivo* analyses reveal cooperation between STAG2 and the PAXIP1/PAGR1 complex in KRAS-driven lung tumors

To understand mechanistically how STAG2-cohesin regulates tumorigenesis, we next sought to find genes that cooperate with STAG2 to mediate these effects. We analyzed the CRISPR/Cas9 human cancer cell line DepMap data for genes whose effects on cell growth (referred to as their gene effect) were most similar to that of STAG2 (Materials and methods) (Kim et al., 2019a; Pan et al., 2018). We initially assessed the impact of inactivating each gene across all cell lines (excluding those with STAG2 mutations, $N = 1,018$) and calculated the correlation (Pearson's r) with the STAG2 gene effect. Consistent with previous analyses, *PAXIP1* and its obligate heterodimer partner *PAGR1* had the most highly correlated gene effects with STAG2 ($r = 0.62$ and 0.61 , respectively) (Bailey et al., 2021; Mayayo-Peralta et al., 2023). Their correlation with STAG2 was approximately twofold higher than any other gene, including core and auxiliary cohesin subunits (Fig. 2 a and Fig. S1, a–c). Importantly, in oncogenic KRAS^{G12/G13}-driven cell lines ($N = 120$), as well as in oncogenic KRAS^{G12/G13}-driven lung adenocarcinoma cell lines ($N = 22$), the gene effects

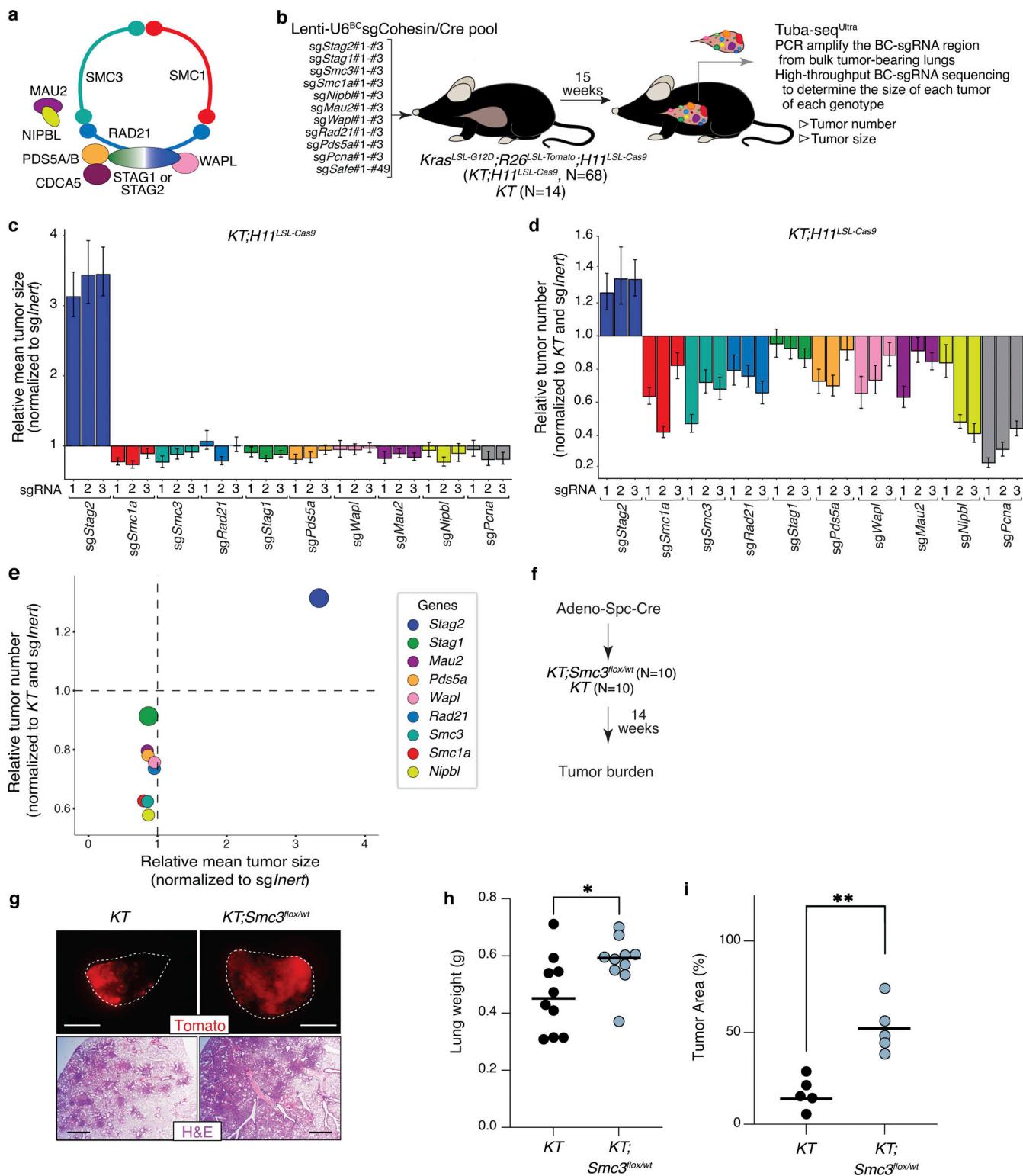


Figure 1. Inactivation of *Stag2* and heterozygous inactivation of the cohesin subunit *Smc3* uniquely increase lung tumor growth. (a) Schematic of the cohesin complex with subunits labeled. (b) Tumor initiation with a pool of Lenti-U6^{BC}sgRNA/Cre vectors. Genotype and number of mice are indicated. Tuba-seq^{Ultra} was performed on each tumor-bearing lung 15 wk after tumor initiation, followed by analyses to quantify tumorigenesis. (c) Relative mean tumor size (normalized to sgInert). Mean \pm 95% confidence intervals are shown. (d) Relative tumor number (normalized to KT and sgInert). Mean \pm 95% confidence intervals are shown. (e) Relative tumor number (normalized to KT and sgInert) versus relative mean tumor size (normalized to sgInert). (f) Tumor initiation with Adeno-Spc-Cre in KT and KT;Smc3^{flx/wt} mice. Mouse number is indicated. (g) Representative fluorescence and histology images of lung lobes from the indicated genotypes of mice. Top scale bars, 5 mm, and bottom scale bars, 1 mm. (h) Lung weights from KT and KT;Smc3^{flx/wt} mice. Each dot represents a mouse and the bar is the mean. Data are representative of two independent experiments. (i) Percentage of Tomato-positive tumor area detected via histology. Each dot represents a mouse, and the bar is the mean. *P value <0.05 , **P value <0.01 , via unpaired t test. Raw values and significance of each effect are shown in Table S1.

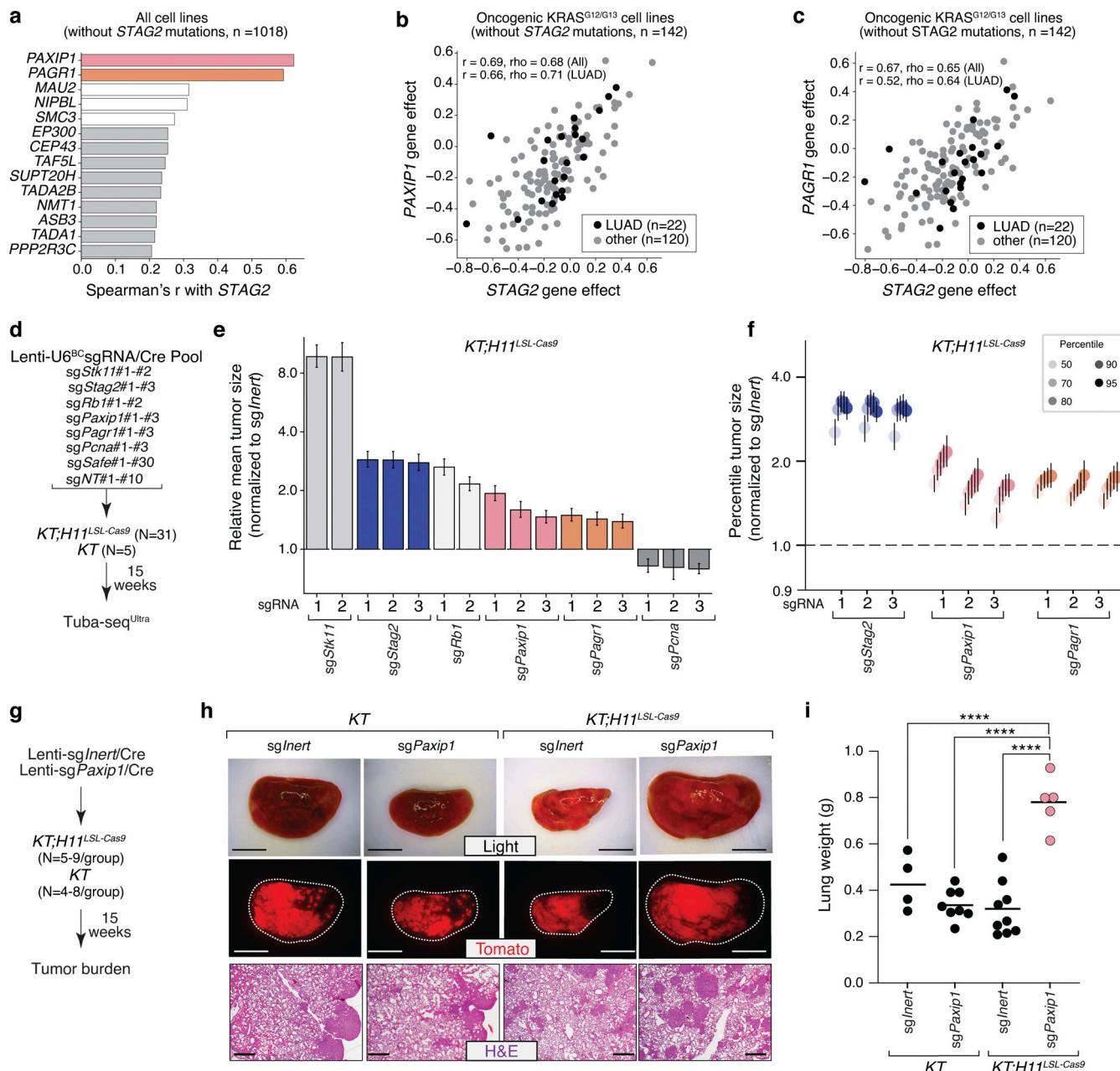


Figure 2. STAG2, PAXIP1, and PAGR1 effects are highly correlated in human cancer cell lines and PAXIP1/PAGR1 are tumor suppressive in KRAS-driven lung tumors in vivo. (a) Genes with the most correlated effects with *STAG2* inactivation from the DepMap. Cell lines with *STAG2* mutations were excluded. *PAXIP1* and *PAGR1* are colored bars. Core cohesin complex genes are white bars. (b and c) Gene effects for *PAXIP1* and *STAG2* inactivation (b) and for *PAGR1* and *STAG2* inactivation (c). Each dot represents a cell line with an oncogenic mutation at codons 12 or 13 of KRAS. Cell lines with *STAG2* mutations were excluded. Lung adenocarcinoma (LUAD) cell lines are shown as black dots. Spearman's r and Pearson rho for all cell lines and for LUAD cell lines are indicated. (d) Tumor initiation with a pool of barcoded Lenti-U6^{BC}sgRNA/Cre Tuba-seq^{Ultra} vectors. Genotype and number of mice are indicated. Tuba-seq^{Ultra} was performed on each tumor-bearing lung 15 wk after tumor initiation, followed by analyses to quantify tumorigenesis. (e) Relative mean tumor size (normalized to sgInert). Mean \pm 95% confidence intervals are shown. Dotted line indicates no effect. (f) Tumor sizes at the indicated percentiles for tumors with sgRNA targeting *Stag2*, *Paxip1*, or *Pagr1* (normalized to sgInert) in *KT;H11*^{LSL-Cas9} mice. Each gene was targeted with three sgRNAs. Error bars indicate 95% confidence intervals. Dotted line indicates no effect. Data represents one replicate of two independent experiments. (g) Tumor initiation with the indicated Lenti-sgRNA/Cre vectors in separate cohorts of *KT* and *KT;H11*^{LSL-Cas9} mice. Number of mice in each group is indicated. Tumor burden was quantified 15 wk after tumor initiation. (h) Representative brightfield and fluorescence images of lung lobes and histology from the indicated groups of mice. Top scale bars and middle scale bars, 5 mm. Lower scale bars, 500 μ m. (i) Lung weights of mice in each group. Each dot represents a mouse and the bar is the mean. ****P value <0.0001 by unpaired t test. Data represent one replicate of two independent experiments. Raw values and significance of each effect are shown in Table S2.

of *PAXIP1* and *PAGR1* robustly correlated with the gene effect of *STAG2* (Fig. 2, b and c; and Fig. S1, d and e). Finally, the positive correlations between *STAG2* and *PAXIP1/PAGR1* gene effects extended to other cancer types where *STAG2* is frequently mutated, including AML, bladder cancer, and Ewing's sarcoma (Fig. S3, f-i) (De Koninck et al., 2020; Solomon et al., 2013; Taylor et al., 2014; Aquila et al., 2018; Kim et al., 2012; Adane et al., 2021).

To determine whether *PAXIP1* and *PAGR1* are tumor suppressors in lung cancer, we used Tuba-seq^{Ultra} to quantify the impact of inactivating *STAG2*, *PAXIP1*, or *PAGR1* on oncogenic KRAS-driven lung tumorigenesis (Fig. 2 d). We initiated tumors *KT;H11^{LSL-Cas9}* and *KT* mice with a pool of Lenti-U6^{BC}-sgRNA/Cre vectors that contained sgRNAs targeting each gene of interest (Fig. 2 d and Table S2). After 15 wk of tumor growth, we quantified the number of neoplastic cells in each tumor using Tuba-seq^{Ultra} (Fig. 2 d). Inactivation of *Paxip1* or *Pagr1* significantly increased tumor size, on par with the effect of inactivating the canonical tumor suppressor *Rbl* (Fig. 2 e). Inactivation of *Paxip1* and *Pagr1* showed modest effects on tumor number, on par with inactivation of *Stkll*, *Stag2*, and *Rbl* (Fig. S1 j). The tumor-suppressive effects of *PAXIP1* and *PAGR1* were less than those of *STAG2* (Fig. 2, e and f). No vectors impacted tumorigenesis in the absence of Cas9 (Table S2).

To validate the tumor-suppressive effects of *PAXIP1*, we initiated lung tumors with Lenti-sgInert/Cre or Lenti-sgPaxip1/Cre in separate cohorts of *KT;H11^{LSL-Cas9}* and *KT* mice (Fig. 2 g). *PAXIP1* inactivation was confirmed by western blotting on sorted neoplastic cells (Fig. S1 k). Overall tumor burden as assessed by lung weight was significantly higher in *KT;H11^{LSL-Cas9}* sgPaxip1 mice compared with each of the three control cohorts (*KT;H11^{LSL-Cas9}* sgInert, *KT* sgPaxip1, and *KT* sgInert) (Fig. 2, h and i). Collectively, these quantitative in vivo data and DepMap findings from human cancer cell lines show that *STAG2*-cohesin effects correlate with *PAXIP1/PAGR1* effects in lung tumorigenesis (Fig. 2, a-f).

STAG2 inactivation modifies the overall tumor suppressive landscape and is epistatic to *PAXIP1* and *PAGR1*

To determine whether *PAXIP1* and *PAGR1* function in the same pathway as *STAG2*, as well as extend our understanding of how *STAG2* deficiency changes the impact of coincident genetic alterations on KRAS-driven lung tumorigenesis, we quantified the impact of inactivating a curated list of genes on tumor initiation and growth in *Stag2*-proficient and *Stag2*-deficient tumors using Tuba-seq^{Ultra} (Fig. 3, a and b). We generated a pool of Lenti-U6^{BC}-sgRNA/Cre vectors that targeted ~160 genes, including genes that correlate with *STAG2* effects in DepMap (e.g., *Paxip1* and *Pagr1*), several canonical tumor suppressor genes (TSGs), genes with protein–protein interactions with *STAG2*, and genes downstream of *STAG2* based on preliminary molecular analyses (Fig. 3, a and b; Table S3; and Materials and methods). Each gene was targeted with three to six sgRNAs, and this pool contained control Safe-cutting “inert” sgRNAs and sgRNAs targeting several control essential genes (Fig. 3, a and b). We initiated lung tumors in *KT*, *KT;H11^{LSL-Cas9}*, and *KT;H11^{LSL-Cas9};Stag2^{flx}* mice (Fig. 3 b). After 13 wk of tumor growth, we used Tuba-seq^{Ultra} to

quantify the number of neoplastic cells in each tumor with each sgRNA (Fig. 3 b).

Stag2 deficiency in *KT;H11^{LSL-Cas9};Stag2^{flx}* mice significantly changed the impact of inactivating only a small number of genes, including several canonical TSGs and genes identified through DepMap (Fig. 3 c). As expected, sgRNAs targeting *Stag2* increased tumor growth *KT;H11^{LSL-Cas9}* mice but not in *KT;H11^{LSL-Cas9};Stag2^{flx}* mice (Fig. 3 d). Inactivation of *Setd2* increased the growth of *Stag2*-deficient tumors more than *Stag2*-proficient tumors, while inactivation of *Stkll* or *Tscl* increased the growth of *Stag2*-deficient tumors less than *Stag2*-proficient tumors (Fig. 3 e and Fig. S1 l). Thus, *Stag2* inactivation modifies the overall tumor suppressor landscape in lung cancer.

Based on our analysis of DepMap data and the tumor-suppressive effects of *STAG2*, *PAXIP1*, and *PAGR1* in vivo, we hypothesized that they may function together to suppress lung tumorigenesis. Using Tuba-seq^{Ultra}, we found that inactivating *Paxip1* or *Pagr1* increased the growth of *Stag2*-proficient tumors but had no impact on the growth of *Stag2*-deficient tumors (Fig. 3 f and Fig. S1 m). This epistasis suggests that *PAXIP1/PAGR1* and *STAG2*-cohesin suppress lung tumorigenesis through partially overlapping mechanisms.

Finally, *Stag1* was the only cohesin complex component for which inactivation had a significantly different effect in *Stag2*-proficient versus *Stag2*-deficient lung tumors (Fig. 3, c and g). *Stag1* inactivation was significantly more deleterious for *Stag2*-deficient lung tumors with effects on tumor number approaching that of inactivating the known essential gene *Pole* (Fig. 3 g and Fig. S1 n). These results are consistent with the well-established synthetic lethality of *Stag1* and *Stag2* and the essential role of cohesin in general in this tumor type (van der Lelij et al., 2017; van der Lelij et al., 2020). Collectively, these quantitative analyses demonstrate that *STAG2* inactivation impacts the overall tumor suppression landscape of lung cancer, is synthetic lethal with *STAG1*, and is epistatic to the *PAXIP1/PAGR1* complex.

STAG2-deficient cancer cells have gene expression changes related to cell differentiation, metabolic pathways, and response to immune cell infiltration

To characterize the molecular effects of *STAG2* deficiency, we initiated tumors in *KT* and *KT;Stag2^{flx}* mice and used FACS to isolate neoplastic cells (DAP^{neg}, lineage^{neg}, Tomato^{pos} cells) for molecular analyses (Fig. 4 a) (Chuang et al., 2017). To capture any changes in the molecular output of *STAG2* deficiency during tumor development, we collected samples after 8 and 16 wk of tumor growth. We initiated tumors using different viral titers such that mice would have similar overall tumor burden at the time of analysis (Fig. 4 a; and Fig. S2, a and b).

Stag2 inactivation in these samples was validated by mapping reads to the floxed exon of *Stag2* and measuring *Stag2* expression (Fig. S2, c and d). Based on hierarchical clustering and principal component analysis (PCA), samples from *KT* and *KT;Stag2^{flx}* mice formed distinct clusters (Fig. 4 b and Fig. S2 e). Comparisons between neoplastic cells from all *KT* and *KT;Stag2^{flx}* samples uncovered >2,500 differentially expressed genes (absolute value of \log_2 fold change $[\log_2 \text{FC}] > 1$, false discovery rate

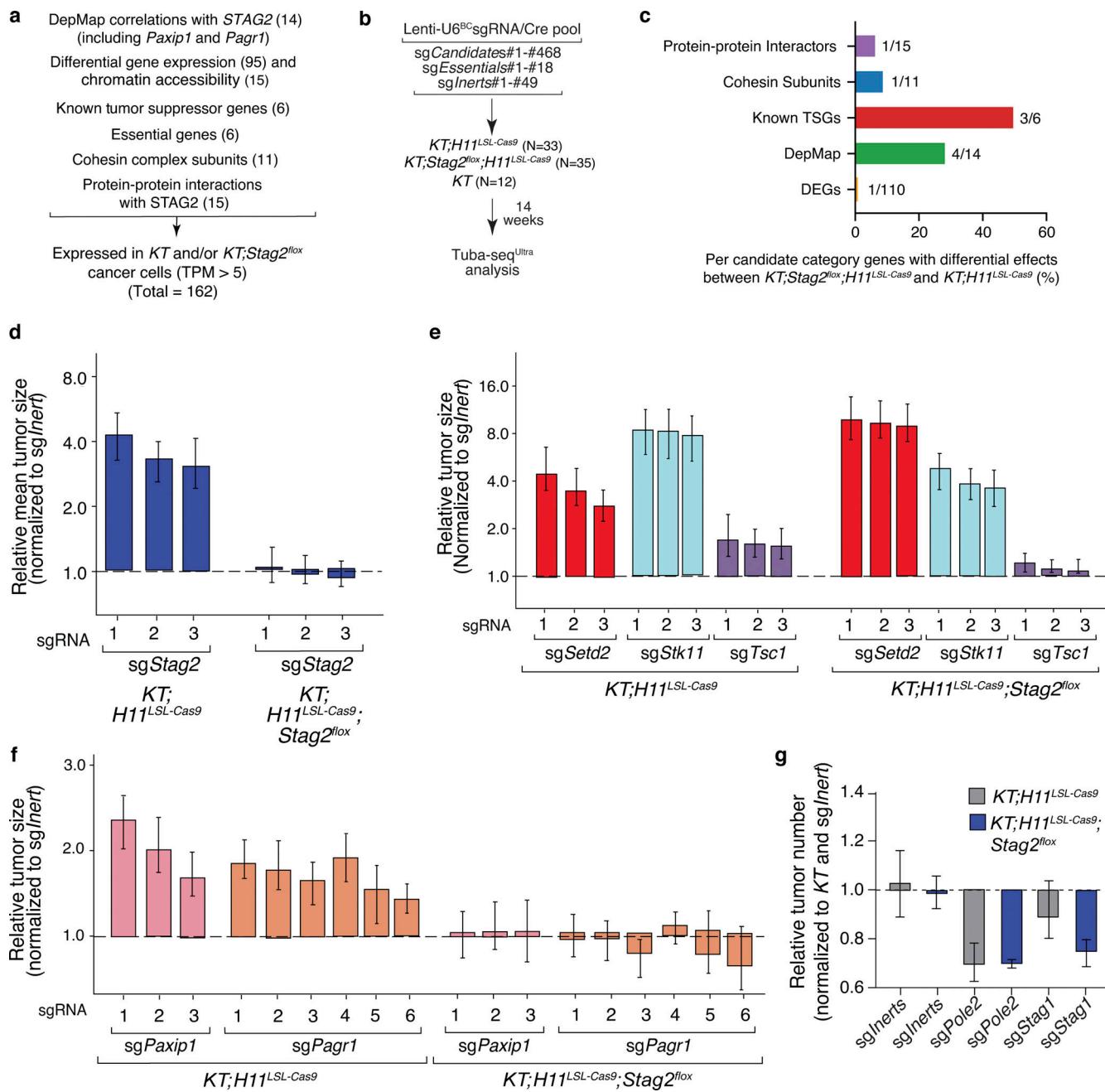


Figure 3. Genetic interactions with STAG2 include modification of overall tumor suppression and functional dependency with PAXIP1 and PAGR1. (a) List of candidate gene criteria for a pool of 468 barcoded Lenti-sgRNA/Cre vectors. (b) Tumor initiation with Lenti-U6^{BC}-sgRNA/Cre vectors in *KT;H11^{LSL-Cas9}*, *KT;H11^{LSL-Cas9};Stag2^{flx}*, and *KT* mice. Number of mice in each group is indicated. Sequencing was performed on each tumor-bearing lung 14 wk after tumor initiation, followed by analysis to quantify tumorigenesis. (c) Bar plot with percent of genes in each candidate category with differential effects between *KT;H11^{LSL-Cas9}* and *KT;H11^{LSL-Cas9};Stag2^{flx}*. (d) Relative mean tumor size of tumors with sgRNAs targeting *Stag2* (normalized to *sgInert*) in *KT;H11^{LSL-Cas9}* mice compared to *KT;H11^{LSL-Cas9};Stag2^{flx}* mice. Mean ± 95% confidence intervals are shown. (e) Relative mean tumor size of tumors with sgRNAs targeting *Setd2*, *Stk11*, or *Tsc1* (normalized to *sgInert*) in *KT;H11^{LSL-Cas9}* mice compared with *KT;H11^{LSL-Cas9};Stag2^{flx}* mice. Mean ± 95% confidence intervals are shown. (f) Relative mean tumor size of tumors with sgRNAs targeting *Paxip1* or *Pagr1* (normalized to *sgInert*) in *KT;H11^{LSL-Cas9}* mice compared with *KT;H11^{LSL-Cas9};Stag2^{flx}* mice. Mean ± 95% confidence intervals are shown. (g) Comparison of average relative tumor number for sgRNAs targeting *Stag1*, *Stag2*, *Pole2* (essential gene), or *Inerts* in *KT;H11^{LSL-Cas9}* mice and *KT;H11^{LSL-Cas9};Stag2^{flx}* mice. Mean ± 95% confidence intervals are shown. Raw values and significance of each effect are shown in Table S4. Data for *KT;H11^{LSL-Cas9}* and *KT* mice represent one replicate of two independent experiments.

[FDR] < 0.01; Fig. 4 b and Table S5). Using gene set enrichment analysis (GSEA) and gene ontology (GO) term enrichment, we uncovered upregulation of pathways related to surfactant protein metabolism (exemplified by significantly increased expression

of the transcription factor *Nkx2-1* and canonical *NKX2-1*-regulated surfactant protein genes), lung development, and several metabolic processes in *Stag2*-deficient neoplastic cells (Fig. 4, c and e; and Fig. S2, f and g). Consistent with this gene expression

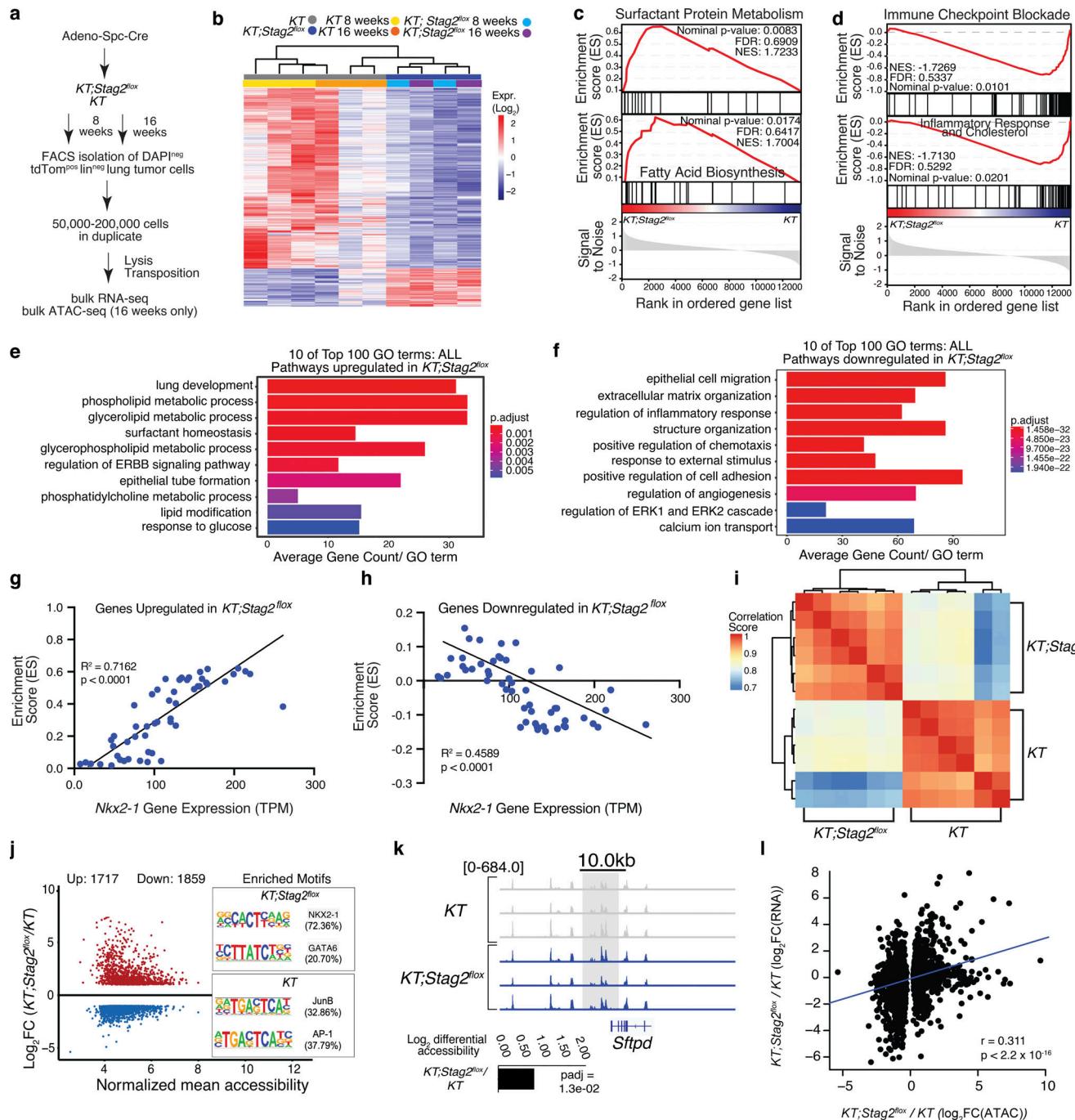


Figure 4. STAG2 inactivation increases tumor-related metabolic processes and cell differentiation. (a) Schematic of tumor initiation in KT and $KT;Stag2^{fl/fl}$ mice. Mice grew tumors until 8 wk to represent large atypical adenomatous hyperplasia and small adenomas and 16 wk to represent larger solid adenomas and early adenocarcinomas (Marjanovic et al., 2020). Outline of tumor cell sorting and sample preparation for bulk RNA-seq and bulk ATAC-seq. (b) Upregulated and downregulated genes ($n = 2,545$ genes) in KT relative to $KT;Stag2^{fl/fl}$ tumors (absolute value of \log_2 fold change ($|\log_2 FC| > 1$, FDR < 0.01). (c and d) GSEA pathways enriched in $KT;Stag2^{fl/fl}$ relative to KT tumors (c) and enriched in KT relative to $KT;Stag2^{fl/fl}$ tumors (d). (e and f) GO Term Gene Count Analysis with ClusterProfiler and EMBL-EBI GO:Term Category Analysis established pathways enriched from upregulated genes for $KT;Stag2^{fl/fl}$ /KT mice (e) and from down-regulated genes for $KT;Stag2^{fl/fl}$ /KT mice (f). (g and h) ES versus $NKX2-1$ gene expression (TPM) from single sample GSEA (ssGSEA) for KPT and KT samples (Chuang et al., 2017) for the gene sets from genes upregulated in $KT;Stag2^{fl/fl}$ versus KT (g) and genes downregulated in $KT;Stag2^{fl/fl}$ versus KT (h). (i) Rank correlation of chromatin accessibility across KT and $KT;Stag2^{fl/fl}$ samples. Samples cluster into two distinct groups. (j) Differential accessibility across 3,576 significant peaks in KT and $KT;Stag2^{fl/fl}$ mice (three mice/group). The x-axis represents the \log_2 mean accessibility per peak and the y-axis represents the \log_2 fold change in accessibility. Colored points are significant ($|\log_2 FC| > 1$, FDR < 0.05). Red points are increased chromatin accessibility in $KT;Stag2^{fl/fl}$ accompanied by transcription factor hypergeometric motif enrichment in $KT;Stag2^{fl/fl}$, and blue points are decreased chromatin accessibility in $KT;Stag2^{fl/fl}$ accompanied by transcription factor hypergeometric motif enrichment in KT tumors. (k) ATAC-seq signal tracks for the *Sftpd* gene locus in neoplastic cells from KT and $KT;Stag2^{fl/fl}$ mice. (l) Comparison of \log_2 fold change for KT and $KT;Stag2^{fl/fl}$ mice for top genes from both bulk RNA-seq and bulk ATAC-seq. $r = 0.311$, $P < 2.2 \times 10^{-16}$.

data, immunohistochemical staining for NKX2-1 showed that *Stag2*-deficient tumors were significantly more likely to have high expression of NKX2-1 than *Stag2*-proficient tumors ($P < 0.001$; Fig. S2 h). Significantly downregulated gene sets in *Stag2*-deficient neoplastic cells were related to epithelial response to immune cells and extracellular matrix organization (Fig. 4, d and f). Across an independent dataset of lung tumorigenesis in oncogenic KRAS-driven models, *Nkx2-1* expression positively correlated with expression of genes that were upregulated by STAG2 deficiency ($R^2 = 0.7162$, $P < 0.0001$) and negatively correlated with expression of genes that were downregulated by STAG2 inactivation ($R^2 = 0.4589$, $P < 0.0001$), consistent with NKX2-1 driving some of these differences (Fig. 4, g and h) (Hwang et al., 2016).

To assess whether *Stag2* deficiency also impacts the underlying chromatin state, we performed assay for transposase-accessible chromatin with high-throughput sequencing (ATAC-seq) on neoplastic cells isolated from *KT* and *KT;Stag2^{flx}* mice. Neoplastic cells from *KT* and *KT;Stag2^{flx}* mice clustered separately and had comparable library quality (Fig. 4 i and Fig. S2 i and Materials and methods). *Stag2*-deficient neoplastic cells have several thousand regions of increased and decreased chromatin accessibility relative to *Stag2*-proficient neoplastic cells ($|\log_2\text{FC}| > 1$, $\text{FDR} < 0.05$) (Fig. 4 j, Fig. S2 j, and Table S9). Interestingly, regions with increased accessibility in *Stag2*-deficient neoplastic cells were enriched for NKX2-1 and GATA family transcription factor motifs, which are well-established regulators of lung-specific lineage programs and lung development (Fig. 4 j and Fig. S2 k) (Hwang et al., 2016; Yamaguchi et al., 2013; Yang et al., 2012; Zheng and Blobel, 2010). Additionally, there were regions with increased accessibility proximal to lung lineage genes, such as *Sftpd* (Fig. 4 k). Regions with decreased accessibility in *Stag2*-deficient neoplastic cells were enriched for JUNB, AP-1, FOS, BATF, and FRA1/2 motifs (Fig. 4 j and Fig. S2 l). Notably, the gene expression and chromatin accessibility changes in neoplastic cells from *KT* and *KT;Stag2^{flx}* mice were positively correlated ($R = 0.311$, $P < 2.2 \times 10^{-16}$) (Fig. 4 l). Collectively, these data show that STAG2 inactivation rewires the epigenome and changes gene expression states.

STAG2 inactivation results in differential DNA looping that correlates with increased gene expression

Cohesin regulates 3D genome structure which can facilitate enhancer-promoter interactions and impact gene expression (Alonso-Gil and Losada, 2023; Rao et al., 2017). Thus, to determine whether STAG2 inactivation impacts chromatin looping, we performed chromosome conformation capture (Hi-C) on neoplastic cells isolated from *KT* and *KT;Stag2^{flx}* mice (Fig. 5 a and Fig. S3 a). *Stag2*-deficient neoplastic cells gained and lost DNA loops compared to *Stag2*-proficient cells. Specifically, in addition to >7,500 common loops, there were ~3,000 loops that were lost/reduced in *Stag2*-deficient cells (*KT* unique loops) and >5,700 loops that were gained/stronger in *Stag2*-deficient cells (*KT;Stag2^{flx}* unique loops; Fig. 5, b and c). Furthermore, *KT;Stag2^{flx}* unique loops were on average larger ($P = 5.5 \times 10^{-28}$) than *KT* unique loops (Fig. 5 d and Fig. S3 b). We identified unique and shared loop “anchors” between *KT* and *KT;Stag2^{flx}*

samples. Notably, many of the unique loops share one anchor site with loops in the corresponding *KT* or *KT;Stag2^{flx}* samples (Fig. 5 e). On average, the loop anchors were more distal in *KT;Stag2^{flx}* samples compared with *KT*, suggesting loop extrusion may favor different anchor sites in the absence of STAG2 (Fig. S3 c).

To better understand whether the differences in 3D chromatin looping contribute to the molecular changes caused by *Stag2* inactivation, we assessed whether the anchor sites of unique loops are associated with transcription start sites (TSSs) of differentially expressed genes. By overlapping differentially expressed genes with the coordinates of the anchor sites of unique loops, we found that genes upregulated in *KT;Stag2^{flx}* samples are enriched proximal to the anchor sites of *KT;Stag2^{flx}* unique loops (i.e., the gene TSS is within the anchor site) (85 genes, $P = 5.3 \times 10^{-8}$) (Fig. S3 d). Similarly, genes with higher expression in *KT* samples were enriched proximal to anchor sites of *KT* unique loops (67 genes, $P = 5.6 \times 10^{-7}$) (Fig. S3 e and Table S11). For differentially expressed genes that are associated with unique anchor sites in either *KT* or *KT;Stag2^{flx}* samples, we analyzed whether the corresponding anchor (i.e., the other side of the unique loop) had changes in chromatin accessibility, which might occur if the generation of novel enhancers led to the formation of the differential chromatin interactions. Interestingly, >75% of the corresponding anchor sites did not significantly change in chromatin accessibility. This suggests that the changes in 3D chromatin looping are the result of STAG2 inactivation not because of enhancer reprogramming. Additionally, many of the genes with increased expression and proximity to unique anchor sites in *Stag2*-deficient neoplastic cells, such as EREG and FGFR, are associated with tumor growth pathways (Fig. 5 f). We also found two loci associated with unique loops in *Stag2*-deficient cells, which have multiple genes that are all upregulated, possibly due to the different loop extrusion in the absence of STAG2 (Fig. S3, f and g). These regions include genes implicated in lung lineage development and lung adenocarcinoma and progression (Ingram et al., 2022; He et al., 2023, Preprint; Toba-Ichihashi et al., 2015). Altogether, this genome conformation data strongly suggest that STAG2 inactivation leads to changes in 3D chromatin looping and activation of downstream pathways that could promote tumor growth and alter differentiation.

PAXIP1 and STAG2-cohesin control conserved gene expression programs

As our genetic epistasis data suggest that the tumor-suppressive function of PAXIP1 and STAG2-cohesin are related, we next determined the extent to which the molecular programs driven by PAXIP1 and STAG2-cohesin overlap. To characterize the molecular effects of PAXIP1 deficiency in lung tumors, neoplastic cells were FACS-isolated from tumors initiated in *KT;H2^{LSL-Cas9}* mice with Lenti-sgInert/Cre or Lenti-sgPaxip1/Cre (Fig. 6 a). We performed bulk RNA sequencing (RNA-seq) on three to four samples from each group (hereafter sgInert and sgPaxip1). Samples cluster based on their genotype (Fig. 6 b and Fig. S4 a). Many (~540) genes were differentially expressed between sgPaxip1 and sgInert neoplastic cells ($|\log_2\text{FC}| > 1$ and $\text{FDR} < 0.01$) (Fig. 6 b and Table S6).

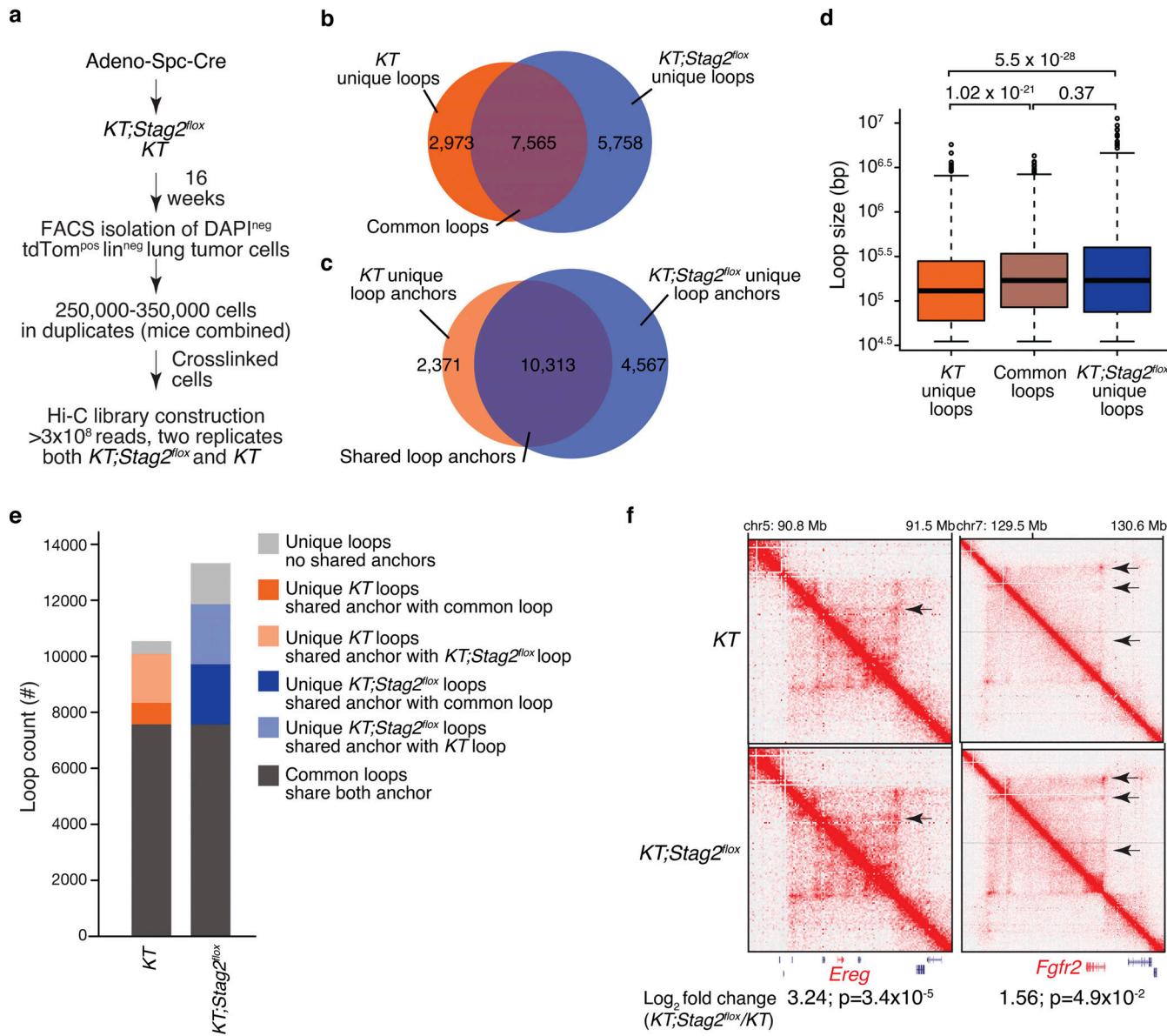


Figure 5. *Stag2* deficiency impacts overall chromatin looping with effects on gene expression. (a) Schematic of tumor initiation with Adeno-Spc-Cre in KT and $KT;Stag2^{fl/fl}$ mice. Outline of tumor cell sorting, crosslinking, and sample preparation for Hi-C. (b) Venn diagram of loops in KT and $KT;Stag2^{fl/fl}$ samples. (c) Venn diagram of loop anchors in KT and $KT;Stag2^{fl/fl}$ samples. (d) Sizes for KT unique loops, common loops, and $KT;Stag2^{fl/fl}$ unique loops. Boxes show median \pm interquartile range. Whiskers show standard error. P values (Wilcoxon rank test) are shown. (e) Loops in KT and $KT;Stag2^{fl/fl}$ samples were compared and sorted into indicated categories based on their unique and shared loop anchors. (f) Differences in chromatin looping and gene expression (from RNA-seq data, \log_2 fold change [$KT;Stag2^{fl/fl}/KT$]; P value) for *Ereg* and *Fgfr2*.

Downloaded from http://upress.org/jem/article-pdf/222/1/e20240765/1936305/jem_20240765.pdf by guest on 09 February 2026

Interestingly, genes that were regulated by both STAG2 and PAXIP1 were almost exclusively downregulated in *Stag2*-deficient and *Paxip1*-deficient neoplastic cells (Fig. 6 c). Genes that were downregulated in *sgPaxip1* neoplastic cells (“PAXIP1 DOWN”) were downregulated in $KT;Stag2^{fl/fl}$ neoplastic cells (P value <0.0036 ; Fig. 6 d). However, genes that were downregulated by *Stag2* deficiency were either equally dependent on PAXIP1 or not dependent on PAXIP1. Pathways enriched in the genes downregulated in both $KT;Stag2^{fl/fl}$ and $KT;H11^{LSL-Cas9}$ *sgPaxip1* tumors include those related to ion transport, metabolic processes, and angiogenesis (Fig. 6 e and f). Genes that were upregulated in $KT;H11^{LSL-Cas9}$ *sgPaxip1* neoplastic cells were

not upregulated in $KT;Stag2^{fl/fl}$ neoplastic cells and vice versa (Fig. S4, b–g).

Given the broad changes in chromatin accessibility induced by *Stag2* deficiency, we performed ATAC-seq on *sgInert* and *sgPaxip1* neoplastic cells (Fig. 6 a). These samples were of similar quality and clustered by genotype (Fig. S4 h). There were $>4,200$ regions with decreased chromatin accessibility and $>13,000$ regions with increased chromatin accessibility in neoplastic cells from *sgInert* versus *sgPaxip1* mice ($|\log_2 FC| > 1$, FDR < 0.05) (Fig. S4, i and j; and Table S9). Motifs for NKX2-1, ZEB1, and GATA were enriched in regions with increased accessibility in *sgPaxip1* neoplastic cells, and motifs for TCF21 and BHLHA15 were

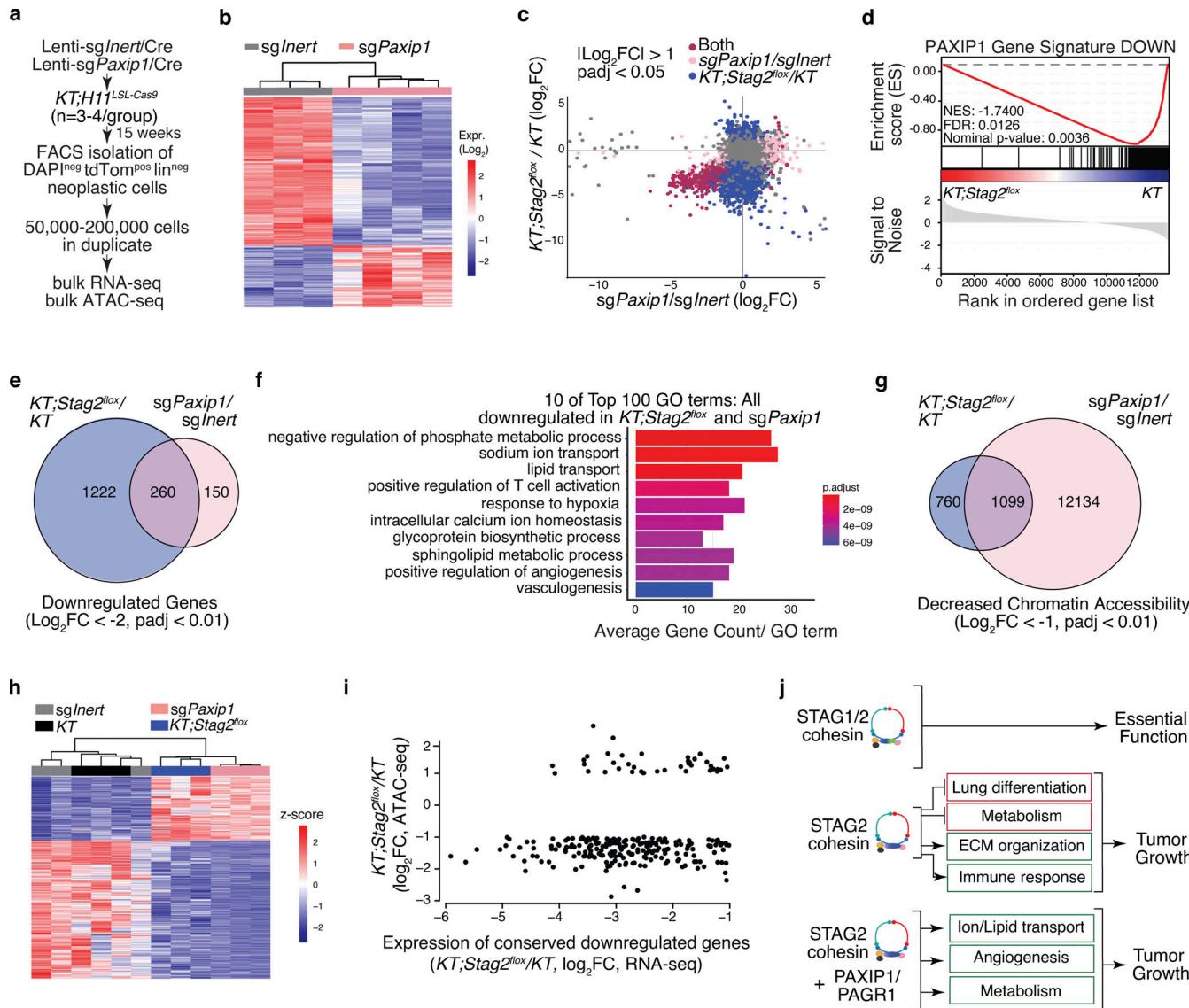


Figure 6. PAXIP1 and STAG2-cohesin mechanisms of tumor suppression are conserved. (a) Schematic of tumor initiation with Lenti-sgPaxip1/Cre or Lenti-sgInert/Cre in KT;H11^{LSL-Cas9} mice (three to four mice/group). Mice grew tumors for 15 wk. Outline of tumor cell sorting and sample preparation for bulk RNA-seq and bulk ATAC-seq. (b) Upregulated and downregulated genes ($n = 540$ genes) in KT;H11^{LSL-Cas9} sgInert relative to KT;H11^{LSL-Cas9} sgPaxip1 tumors ($|\log_2 \text{FC}| > 1$, Padj < 0.01). (c) Positive correlation between KT;Stag2^{flx}/KT ($\log_2 \text{FC}$) and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert ($\log_2 \text{FC}$). Each dot is a gene. Dots $|\log_2 \text{FC}| > 1$, Padj < 0.05 in both are maroon, dots $|\log_2 \text{FC}| > 1$, Padj < 0.05 in KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} only are pink, and dots $|\log_2 \text{FC}| > 1$, Padj < 0.05 in KT;Stag2^{flx}/KT are dark blue. Dots $|\log_2 \text{FC}| > 1$, Padj < 0.05 in neither are grey. (d) Downregulated PAXIP1 gene signature enrichment in rank-ordered gene list for KT versus KT;Stag2^{flx}. (e) Venn diagram of shared downregulated genes ($\log_2 \text{FC} < -2$, Padj < 0.01) between KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert. (f) GO term gene count analysis with ClusterProfiler and EMBL-EBI GO:term category analysis established from conserved downregulated pathways between KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert. (g) Venn diagram of shared regions of decreased chromatin accessibility ($\log_2 \text{FC} < -1$, Padj < 0.01) between KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert. (h) Accessibility z-score for regions with significantly different accessibility in both KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert ATAC-seq samples ($|\log_2 \text{FC}| > 1$, Padj < 0.01) that are also associated with genes with significantly different expression in both KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert neoplastic cells ($|\log_2 \text{FC}| > 1$, Padj < 0.01). Each row is a differentially accessible region. (i) Enrichment for regions with significant decreased accessibility in KT;Stag2^{flx}/KT ($|\log_2 \text{FC}| > 1$, Padj < 0.01) samples associated with genes with significantly decreased expression in both KT;Stag2^{flx}/KT and KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert neoplastic cells ($|\log_2 \text{FC}| < -1$, Padj < 0.01). Changes in accessibility and expression in KT;Stag2^{flx}/KT is shown. Each dot is a differentially accessible region. (j) Model of lung tumor suppression regulated by STAG2-cohesin-PAXIP1/PAGR1 axis.

Downloaded from http://jupress.org/jem/article-pdf/222/1/02040765/1936305/jem_20240765.pdf by guest on 09 February 2024

enriched in regions with decreased accessibility (Fig. S4, k and l). Motifs for DLX1, SMAD2, and HOXA2 were enriched in regions with decreased accessibility that overlap between sgPaxip1 neoplastic cells and STAG2-deficient cells (Fig. S4 m). Although there was motif enrichment for NKX2-1 in regions with

increased accessibility in sgPaxip1 neoplastic cells, neither NKX2-1 nor NKX2-1-regulated genes were upregulated in sgPaxip1 neoplastic cells, suggesting that regulation of cell differentiation through NKX2-1 is a STAG2-specific mechanism (Fig. S4, f and g; and Fig. S2, f and g).

Similar to gene expression, there were more shared regions with decreased rather than increased chromatin accessibility (Fig. 6 g, Fig. S4 j, and Table S9). Interestingly, many down-regulated genes conserved between *KT;Stag2^{flx}* and *sgPaxip1* had decreased accessibility in *Stag2*-deficient and *Paxip1*-deficient neoplastic cells (Fig. 6, h and i; and Table S7). Lastly, there was no overlap between genes regulated by PAXIP and genes regulated by STAG2-dependent chromatin looping (Fig. S4 n). Collectively, these results emphasize that there are distinct PAXIP1-dependent and PAXIP1-independent STAG2-mediated tumor suppressive programs, consistent with the overall tumor-suppressive effects of PAXIP1, PAGR1, and STAG2 in vivo (Fig. 6 j and Fig. S5).

Discussion

Here, we coupled multiplexed and quantitative functional genomics within autochthonous lung cancer models, *in vivo* genetic epistasis experiments, and molecular analyses to uncover mechanisms of STAG2-mediated tumor suppression and link it to PAXIP1/PAGR1. Tumor suppression in lung cancer was specific to STAG2-cohesin rather than STAG1-cohesin or overall cohesin. Cohesin components beyond STAG2 are also frequently mutated in human lung cancer, and our results show that heterozygous inactivation of a core cohesin gene increased tumor growth. We show that SMC3 exhibits both essential and haploinsufficient TSG characteristics, further emphasizing the versatility and power of STAG2-cohesin as a tumor suppressor. Thus, heterozygous or hypomorphic mutations in core or auxiliary cohesin components likely reduce STAG2-cohesin-mediated tumor suppression and make the fraction of human lung tumors that are driven by this mechanism greater than only those with STAG2 mutations (Fig. 1 and Fig. S2).

Using cancer cell line dependency data, multiple studies have revealed a correlation between the dependency scores for STAG2 and the PAXIP1-PAGR1 complex (Bailey et al., 2021; Mayayo-Peralta et al., 2023), which we show extends to cell lines driven by oncogenic KRAS as well as to the subset of those cell lines derived from lung adenocarcinoma (Fig. 2, b and c). Previous studies using cancer cell lines have shown that inactivation of STAG2 and PAXIP1 reduces growth, leading to the suggestion that these genes promote cancer cell growth (Mayayo-Peralta et al., 2023; van Schie et al., 2023). However, our functional data show that STAG2, PAXIP1, and PAGR1 normally constrain lung tumor growth *in vivo*, consistent with many studies that document STAG2 as a tumor suppressor (De Koninck et al., 2020; Taylor et al., 2014; Aquila et al., 2018; Kim et al., 2012; Adane et al., 2021; Cai et al., 2021; Solomon et al., 2011). Thus, while cell lines may correctly uncover correlated gene effects, their optimized growth in culture and/or lack of physiologic context limit their ability to identify functional tumor suppressors.

To identify genes that cooperate with STAG2 to suppress tumor growth *in vivo*, we compared the tumor-suppressive effects of a broad panel of genes in STAG2-proficient and -deficient autochthonous lung tumors (Fig. 3). These data uncovered relatively few genes whose effects were impacted by *Stag2*

deficiency, emphasizing the uniqueness of the interaction between STAG2 and PAXIP1/PAGR1 (Fig. 3 c). In this experiment, we did not address the role of inflammation in KRAS-driven tumor progression, which could serve as a critical dimension for future studies (Serresi et al., 2016, 2018). PAXIP1 and PAGR1 have been implicated in several biological processes, including DNA damage responses (Liu and Li, 2022; Gong et al., 2009; Lechner et al., 2000; Callen et al., 2012). Thus, we initially considered whether the reduced impact of *Paxip1* and *Pagr1* inactivation on tumor growth relative to *Stag2* inactivation could be due to partially offsetting tumor-suppressive and essential DNA damage functions (Fig. 3 f; and Fig. S5, a and b). However, *Paxip1* or *Pagr1* inactivation in the context of *Stag2* deficiency (i.e., in *KT;H1^{LSL}Cas9;Stag2^{flx}* mice) did not reduce tumorigenesis, as would be expected if these two genes had a second essential function (Fig. 3 f; and Fig. S5, a and b). Furthermore, PAXIP1/PAGR1 are required for the expression of a subset of STAG2-regulated genes, some of which are likely responsible for PAXIP1/PAGR1-mediated tumor suppression (Fig. 5; and Fig. S5, a and b).

STAG2-mediated tumor suppression is also mediated by PAXIP1/PAGR1-independent gene expression programs (Fig. 5 and Fig. S3). Interestingly, in contrast to many genetic alterations that increase tumorigenesis while decreasing or changing differentiation, STAG2-deficient tumors express higher levels of the lung lineage defining transcription factor NKX2-1 as well as canonical genes related to ATII differentiation (Fig. 4, g, h, j, and k; and Figs. S2 and S3). Additionally, there is increased DNA loop formation in STAG2-deficient tumors at loci that overlap with genes involved in lung differentiation and tumorigenesis, such as genes related to ATII differentiation and the EGF pathway (Fig. 6 d and Fig. S4, e–g). Thus, it is interesting to note the potential stabilization of this cellular state by STAG2 deficiency, which otherwise might be anticipated to be selected against (Winslow et al., 2011). Genomic regions with increased accessibility in STAG2-deficient tumors are also enriched for NKX2-1 binding sites (Fig. 4 k). Whether NKX2-1 interacts directly or indirectly with STAG2-cohesin-PAXIP1/PAGR1 and whether NKX2-1-regulated oncogenic programs drive growth in this context will be important areas of future investigation (Weir et al., 2007).

While we show clear evidence for a genetic interaction between STAG2-cohesin and the PAXIP1/PAGR1 complex during lung tumor suppression, these data do not distinguish whether PAXIP1/PAGR1 functions upstream or downstream to STAG2-cohesin (Fig. 3 f). Our data and existing literature on PAXIP1/PAGR1 and STAG2-cohesin are most consistent with a model in which PAXIP1/PAGR1 is initially recruited to genome loci (perhaps through direct interactions with lineage-specific transcription factors), followed by STAG2-cohesin localization at those regions, and the formation of genomic contacts that ultimately impact gene expression to suppress lung tumor growth (Mayayo-Peralta et al., 2023; Kojic et al., 2018). A direct interaction between PAXIP1/PAGR1 and STAG2-cohesin is only weakly supported by existing mass spectrometry data on STAG2-, cohesin-, and PAXIP1-interacting proteins (Kim et al., 2019b; Hara et al., 2014). PAXIP1/PAGR1 and STAG2-cohesin could interact

weakly or transiently, through an intermediate protein, or be functionally linked through a less direct mechanism, such as PAXIP1/PAGR1-mediated recruitment of histone modifying enzymes followed by preferential localization of STAG2-cohesin at these regions (Grimley and Dressler, 2018; Yan et al., 2018).

Interestingly, while our Hi-C data uncovered unique loops that overlap with genes impacted by STAG2 inactivation, these unique loops did not overlap with genes impacted by both STAG2 inactivation and PAXIP1 inactivation (Fig. S4 n; and Tables S7 and S11). Further biochemical, genetic, and genomic experiments including CTCF ChIA-PET and CUT&RUN should further clarify the interaction between STAG2-cohesin and the PAXIP1/PAGR1 complex as well as their genomic localization during tumor suppression.

Interestingly, the tumor-suppressive STAG2-PAXIP1/PAGR1 axis may transcend cancer type. In human cell lines, *STAG2*, *PAXIP1*, and *PAGR1* have highly correlated effects in bladder cancer, Ewing's sarcoma, AML, and lung adenocarcinoma, all of which also have frequent mutations in *STAG2* (Fig. S1, f-i) (Gao et al., 2013; Cerami et al., 2012; de Bruijn et al., 2023). In particular, the correlation of *STAG2* with *PAXIP1* and *PAGR1* in AML was strikingly high ($r = 0.86$ and $r = 0.90$, respectively) (Fig. S1, f and g). In AML, *STAG2* mutations and heterozygous mutations in *SMC3* are well-established regulators of the preleukemic state (Viny et al., 2019). Interestingly, cancer sequencing studies have also identified mutations in *PAXIP1* and *PAGR1* and the down-regulation of *PAXIP1* expression in AML (Fig. S5, c-e) (Gao et al., 2013; Cerami et al., 2012; de Bruijn et al., 2023; AACR Project GENIE Consortium, 2017). Thus, the PAXIP1/PAGR1 complex may represent an important unexplored driver in AML.

Collectively, our data not only highlight a STAG2-PAXIP1/PAGR1 tumor-suppressive axis that may transcend cancer types but also establish distinct PAXIP1-independent and PAXIP1-dependent STAG2-mediated tumor-suppressive functions. Our findings underscore the important roles of STAG2-cohesin in suppressing lung tumorigenesis.

Materials and methods

Design, generation, barcoding, and production of lentiviral and adenoviral vectors

sgRNA sequences targeting each putative TSG in the Tuba-seq^{Ultra} pools were chosen utilizing a combination of CRISPICK (<https://portals.broadinstitute.org/gppx/crispick/public>) and the Brie library (DeWeirdt et al., 2022). First, the sgRNAs from the Brie library were filtered based on (1) their predicted on-target efficacy score (score $> 0.2^2$) from CRISPICK and (2) off-target risk. To evaluate off-target risk, the sgRNA sequence + NGG were aligned to mouse reference genome assembly (GRCm39). sgRNAs that have a perfect match or a tolerable mismatch (a single mismatch occurring at a position located >10 bases away from the PAM sequence) in an off-target location were regarded as high off-target risk and were excluded. For the Brie sgRNAs that successfully passed this filter, a maximum of two sgRNAs were selected based on their on-target efficacy score. To ensure the inclusion of three sgRNAs per gene, additional sgRNAs were designed using CRISPICK, adhering to the aforementioned two

criteria. This approach resulted in a final set of sgRNAs, comprising both tested sgRNAs from the Brie library and novel sgRNAs designed based on one of the most reliable sgRNA-predicting algorithms available (Doench et al., 2016). All sgRNA sequences used are shown in Tables S1, S2, and S4.

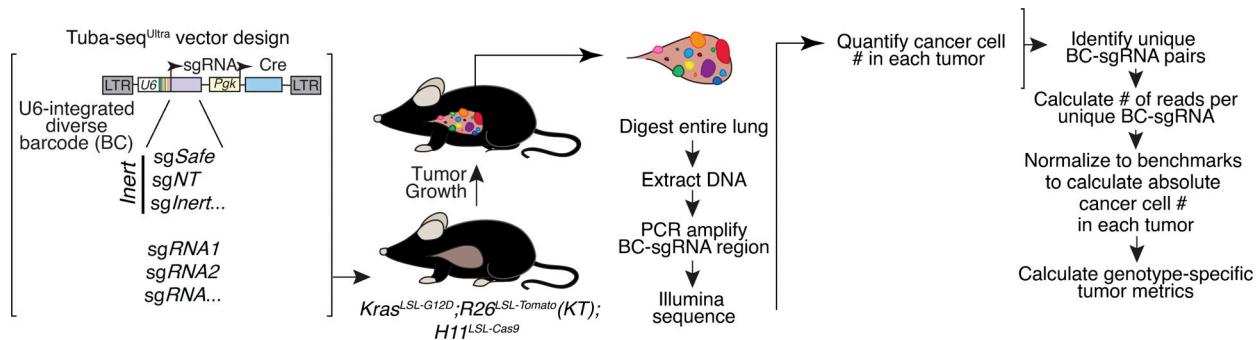
To generate vectors for Tuba-seq^{Ultra}, BC-containing oligo pools were cloned at the 3' end of the bU6 promoter in the Tuba-seq^{Ultra} backbone vector by Gibson assembly, and sgRNA libraries were added using Golden Gate assembly (Tang et al., 2024, Preprint). Prior to adding the barcoded oligo pools, the last 20 bp of the bU6 promoter and 137 bp filler from the Tuba-seq^{Ultra} backbone vector were removed, followed by DpnI digestion to remove the remaining methylated parental vector. The single-stranded barcoded oligos were cloned into the linearized and DpnI-digested Tuba-seq^{Ultra} backbone vector with Gibson assembly to diversify the sequence of the last 20 bp of the bU6 promoter. PCR-amplified sgRNA pools (TWIST Biosciences) were cloned into the BsmBI-v2-digested barcoded Tuba-seq^{Ultra} vector pool using Golden Gate assembly. The products were transformed into NEB 10- β electrocompetent bacteria and several million colonies were pooled, followed by plasmid extractions (Qiagen Plasmid Plus Midiprep kit) and sequencing (Novogene Corporation, Inc.) to determine the BC and sgRNA representation (Fig. S2 a). To generate Lenti-sgRNA/Cre vectors encoding individual sgRNAs, individual sgRNAs were cloned downstream of the U6 promoter in a pLL3.3 backbone containing the PGK promoter and Cre using site-directed mutagenesis (Rogers et al., 2017).

Lentiviral vectors were produced using polyethylenimine-based transfection of 293T cells with delta8.2 and VSV-G packaging plasmids in 150-mm cell culture plates. Sodium butyrate (B5887; Sigma-Aldrich) was added 8 h after transfection to achieve a final concentration of 20 mM. The medium was refreshed 24 h after transfection. 20 ml of virus-containing supernatant was collected 36 and 48 h after transfection. The two collections were then pooled and concentrated by ultracentrifugation (25,000 rpm for 1.5 h), resuspended overnight in 100 μ l PBS, and frozen at -80°C . Adeno-Spc-Cre was purchased from the University of Iowa Core Web.

We generated a pool of barcoded lentiviral vectors encoding Cre along with sgRNAs targeting a panel of genes. Vectors with sgRNAs targeting essential genes (Table S1, S2, and S4) and "inert" sgRNAs were also generated. In this lentiviral system, a diverse BC is integrated into the bovine U6 (bU6) promoter directly 5' of the sgRNA. As a result, each clonal tumor is uniquely identified by a BC-sgRNA element, and the sgRNA sequence indicates the specific gene targeted in each tumor (Scheme 1).

Mice and tumor initiation

The use of mice for the current study has been approved by the Institutional Animal Care and Use Committee at Stanford University, protocol number 26696. *Kras*^{LSL-G12D/+} (RRID:IMSR_JAX:008179), *R26*^{LSL-tdTomato} (RRID:IMSR_JAX:007909), *H11*^{LSL-Cas9} (RRID:IMSR_JAX:027632), *Stag2*^{fl/fl} (RRID:IMSR_JAX:030902), and *Smc3*^{fl/fl} (RRID:IMSR_JAX:030559) mice have been previously described (Cai et al., 2021; Jackson et al., 2001; DuPage



Scheme 1. Schematic of Tuba-seq^{Ultra} vector design and tumor initiation. U6-integrated diverse BC ($U6^{BC}$) contains a 21-nucleotide region at the 3' end of the bovine U6 promoter immediately downstream of the TATA box. The sgRNA-Pool/Cre consists of barcoded lenti-sgRNA/Cre vectors that contain three sgRNAs targeting each gene of interest, safe cutting “inert” sgRNAs, and sgRNAs targeting an essential gene. $KT;H11^{LSL-Cas9}$ are transduced with the lenti- $U6^{BC}$ -sgRNA-Pool/Cre. 14–15 wk after tumor initiation, we extract DNA from bulk tumor-bearing lungs and used Tuba-seq^{Ultra} to quantify the impact of targeting each regulator on tumor growth, and tumor initiation for each tumor of each genotype.

Downloaded from http://jupress.org/jem/article-pdf/2221/1/e20240765/1936305/jem_20240765.pdf by guest on 09 February 2026

et al., 2009). All mice were on a C57BL/6 background. Lung tumors were initiated by intratracheal delivery of Lenti-sgRNA/Cre vectors or Adeno-Spc-Cre vectors (Cai et al., 2021; Doench et al., 2016). Adenoviral vectors were $CaCl_2$ -precipitated prior to delivery.

In our initial screen for cohesin complex members, which was part of a larger screen of regulators of lung tumorigenesis (Tang et al., 2024, Preprint), cohorts of KT ($n = 14$) and $KT;H11^{LSL-Cas9}$ ($n = 68$) were transduced with 3×10^5 and 1×10^5 infectious units (IFU) per mouse, respectively (Fig. 1a). To generate lung tumors that were heterozygous mutant for cohesin, tumors were initiated in KT and $KT;Smc3^{flx/wt}$ mice ($n = 13$ –15 mice/group) with 1×10^9 IFU Adeno-Spc-Cre (Fig. 1f). For our screen that included $sgStag2$, $sgPaxip1$, and $sgPagr1$, which was part of a larger screen of regulators of lung tumorigenesis, tumor cohorts of KT ($n = 5$) and $KT;H11^{LSL-Cas9}$ ($n = 31$) (Tang et al., 2024, Preprint) were transduced with 3×10^5 IFU and 1×10^5 IFU, respectively, of the Lenti- $U6^{BC}$ -sgRNA/Cre pool. To generate tumors with inactivation of PAXIP1 for RNA-seq and ATAC-seq, cohorts of KT ($n = 13$) and $KT;H11^{LSL-Cas9}$ ($n = 32$) mice were transduced with Lenti-sgPaxip1/Cre and Lenti-sgInert/Cre at 1.5×10^5 IFU and 3×10^5 IFU per mouse, respectively (Fig. 2d). For our Tuba-seq^{Ultra} screen of STAG2-mediated tumor suppression, cohorts of KT ($n = 12$), $KT;H11^{LSL-Cas9}$ ($n = 33$), and $KT;H11^{LSL-Cas9};Stag2^{flx}$ ($n = 35$) mice were transduced with 3×10^5 IFU, 1×10^5 IFU, and 5×10^4 IFU of Lenti- $U6^{BC}$ -sgRNA/Cre per mouse, respectively (Fig. 3, a and b). To generate lung tumors for RNA-seq and ATAC-seq, tumors were initiated in KT ($n = 20$) with 5×10^9 and 1×10^9 IFU per mouse for collection at the 8- and 16-wk time points, respectively, and in $KT;Stag2^{flx}$ ($n = 20$) with 2.5×10^8 and 2.5×10^7 IFU per mouse for the 8- and 16-wk time points, respectively (Fig. 4a).

Tuba-seq^{Ultra} library generation

Genomic DNA was isolated from bulk tumor-bearing lung tissue from each mouse as previously described (Cai et al., 2021; Tang et al., 2024, Preprint; Rogers et al., 2018; Yousefi et al., 2022). Following homogenization and overnight proteinase K digestion, genomic DNA was extracted from the lung lysates using

standard phenol-chloroform methods. Subsequently, Q5 High-Fidelity 2x Master Mix (M0494X; New England Biolabs) was used to amplify the $U6$ -BC-sgRNA region from $32 \mu\text{g}$ of genomic DNA in a total reaction volume of $800 \mu\text{l}$ per sample. The unique dual-indexed primers used were as follows: forward: 5'-AAT GATACGGCACCACCGAGATCTACAC-3', eight nucleotides for i5 index: 5'-ACACTCTTCCTACACGACGCTCTCCGATCT-3', six to nine random nucleotides for increased diversity: 5'-GCG CACGCTTGCCGCGCTG-3'; and reverse: 5'-CAAGCAGAAGAC GGCATACGAGAT-3, six nucleotides for i7 index: 5'-GTGACT GGAGTTCAGACGTGTGCTCTCCGATCT-3', nine to six random nucleotides for increased diversity: 5'-CAGGTTCTTGCGAAC CTCAT-3'. The PCR products were purified with Agencourt AMPure XP beads (A63881; Beckman Coulter) using a double-size selection protocol. The concentration and quality of the purified libraries were determined using the Agilent High Sensitivity DNA kit (5067-4626; Agilent Technologies) on the Agilent 2100 Bioanalyzer (G2939BA; Agilent Technologies). The libraries were pooled based on lung weight, cleaned up using AMPure XP beads, and sequenced (read length, 2×150 bp) on the Illumina HiSeq 2500 or NextSeq 500 platform (Novogene Corporation, Inc.).

Processing of paired-end reads to identify the U6-BC and sgRNA

Paired ends were first merged using AdapterRemoval (Schubert et al., 2016), and merged reads were parsed using regular expressions to identify the sgRNA sequence and clonal BC. When identifying sgRNA sequences, we strictly required a perfect match with the designed sequences. The 14-nucleotide random BC sequence possesses a high theoretical diversity of $\sim 4^{14}$ ($>10^8$). In each mouse, there are typically fewer than 100 unique tumors associated with each sgRNA. This indicates that the probability of two genuine unique clonal BCs being within a hamming distance of each other is extremely low. Consequently, when we encounter low-frequency clonal BCs within a 1-hamming distance of high-frequency clonal BCs, we attribute them to sequencing or PCR errors. These low-frequency BCs were merged with BCs of higher frequencies.

After applying these filtering steps, we converted the read counts associated with each BC-sgRNA into absolute neoplastic cell numbers. This conversion was accomplished by normalizing the read counts to the number of reads in the “spike-in” cell lines added to each sample prior to lung lysis and DNA extraction. The median sequencing depth across all experiments was ~1 read per 50 cells. To perform statistical comparisons of tumor genotypes, we imposed a minimum tumor size cutoff of 300 cells.

Summary statistics for overall growth rate

To quantify the impact of each gene on tumor growth, we employed a normalization process that involved calculating statistics for tumors produced by a given sgRNA X (sgX tumors) and comparing them to the corresponding statistics of tumors generated by control sgRNAs (sgInert tumors). Two key statistical measures were employed to characterize these distributions: the size of tumors at defined percentiles of the distribution (specifically the 50th, 70th, 80th, 90th, and 95th percentile tumor sizes) and the LN mean size. The percentile sizes are nonparametric summary statistics of the tumor size distribution. By focusing on percentiles corresponding to the upper tail of the distribution, we specifically examined the growth of larger tumors, thereby mitigating potential confounding factors arising from variations in cutting efficiency among guides. The LN mean is the maximum-likelihood estimate of mean tumor size assuming a log-normal distribution. We normalized these two statistics calculated on tumors of each genotype to the corresponding sgInert statistic. The resulting ratios reflect the growth advantage or disadvantage associated with each tumor genotype relative to the growth of sgInert tumors.

For example, the relative i^{th} percentile size for tumors of genotype X was calculated as:

$$\text{Relative tumor size at } i^{\text{th}} \text{ percentile for sgX tumors} = \frac{\text{tumor size at } i^{\text{th}} \text{ percentile for sgX tumors}}{\text{tumor size at } i^{\text{th}} \text{ percentile for sgInert tumors}}$$

Likewise, the relative LN mean size for tumors of genotype X was calculated as:

$$\text{Relative LN mean size for sgX tumors} = \frac{\text{LN mean of tumor size for sgX tumors}}{\text{LN mean of tumor size for sgInert tumors}}$$

Summary statistics for relative tumor number

In addition to the tumor size metrics described above, we characterized the effects of gene inactivation on tumorigenesis in terms of the number of tumors (“tumor number”) associated with each genotype. Unlike the aforementioned metrics of tumor size, tumor number and burden are linearly affected by lentiviral titer and are thus sensitive to underlying differences in the representation of each Lenti-sgRNA/Cre vector in the viral pool. Critically, each Tuba-seq^{Ultra} experiment included a cohort of *KT* control mice. *KT* mice lack expression of Cas9, rendering all Lenti-sgRNA/Cre vectors functionally equivalent to non-targeting sgRNA vectors in these mice. Therefore, the observed tumor number and burden associated with each sgRNA reflect their respective representation within the viral

pool. This control allows us to accurately assess the specific effects of gene inactivation on tumor formation and growth, accounting for any biases introduced by the viral vector composition.

To assess the extent to which a given gene (X) affects tumor number, we therefore first normalized the number of sgX tumors in *KT;H11^{LSL-Cas9}* mice by the number of sgX tumors in the *KT* mice:

$$\text{Tumor number for sgX tumors} = \frac{\text{Number of sgX tumors in } \text{KT;H11}^{\text{LSL-Cas9}} \text{ mice}}{\text{Number of sgX tumors in } \text{KT mice}}$$

As with the tumor size metrics, we then calculated a relative tumor number by normalizing this statistic to the corresponding statistic calculated using sgInert tumors:

$$\text{Relative tumor number for sgX tumors} = \frac{\text{Tumor number for sgX tumors}}{\text{Tumor number for sgInert tumors}}$$

Genes that influence relative tumor number modify the probability of tumor initiation and/or the very early stages of oncogene-driven epithelial expansion, which prior work suggests are imperfectly correlated with tumor growth at later stages (Chuang et al., 2017). Relative tumor number thus captures an additional and potentially important aspect of TSG function.

Calculation of confidence intervals and P values for tumor growth and number metrics

Confidence intervals and P values were calculated using bootstrap resampling for each sample statistic. To account for both mouse-to-mouse and within-mouse variability, we adopted a two-step, nested bootstrap approach where we first resampled mice and then resampled tumors within each mouse to generate resampled data. 10,000 times of bootstrapping was performed to calculate 10,000 values of each statistic. 95% confidence intervals were calculated using the 2.5th and 97.5th percentiles of the bootstrapped statistics. Because we calculate metrics of tumor growth that are normalized to the same metrics in sgInert tumors, under the null model where genotype does not affect tumor growth, the test statistic is equal to 1. Two-sided P values were thus calculated as follows:

$$p = 2 \cdot \min[\Pr(T > 1), \Pr(T < 1)]$$

where T is the test statistic and $\Pr(T > 1)$ and $\Pr(T < 1)$ were calculated empirically as the proportion of bootstrapped statistics that were more extreme than the baseline of 1. To account for multiple hypothesis testing, P values were FDR-adjusted using the Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995). Summarized statistics of all Tuba-seq^{Ultra} experiments in this study can be found in Tables S1, S2, S3, and S4.

Analysis of DepMap data

Cancer cell line dependency data (DepMap Public 22Q2) and mutation data (Cancer Cell Line Encyclopedia) were acquired from the Broad Institute DepMap Portal (<https://depmap.org/>)

portal). First, cell lines without non-silent mutations in *STAG2* were selected. Then the remaining cell lines were further categorized into five overlapping sets: (1) cell lines with *KRAS* mutations (with G12 or G13 substitutions), (2) cell lines derived from adenocarcinoma or non-small cell lung cancers (NSCLC), (3) cell lines obtained from bladder cancer, (4) cell lines derived from Ewing's sarcoma, and (5) cell lines derived from AML. Pearson's and Spearman's correlation coefficients were calculated to measure the correlation between *STAG2* and other genes using gene effect scores across the aforementioned cell line sets.

Tumor dissociation and cancer cell sorting

Tumors were dissociated using collagenase IV, dispase, and trypsin at 37°C for 30 min as previously described (Chuang et al., 2017). Cells were stained with DAPI and antibodies against CD45 (30-F11), CD31 (390), F4/80 (BM8), and Ter119 (all from BioLegend) to exclude hematopoietic and endothelial cells. FACSaria sorters (BD Biosciences) were used for cell sorting.

Western blotting on sorted cancer cells

Cells were lysed in RIPA buffer (50 mM Tris-HCl [pH 7.4], 150 mM NaCl, 1% Nonidet P-40, and 0.1% SDS) and incubated at 4°C with continuous rotation for 30 min, followed by centrifugation at 12,000 relative centrifugal force in units of *g* for 10 min. The supernatant was collected and the protein concentration was determined by BCA assay (23250; Thermo Fisher Scientific). Protein extracts (10–50 µg) were separated on 4–12% SDS-PAGE and transferred onto polyvinylidene fluoride membranes. The membranes were blocked with 5% non-fat milk in TBS with 0.1% Tween 20 (TBST) at room temperature for 1 h, cut according to the molecular weight of the target protein (with at least two flanking protein markers), followed by incubation with primary antibodies diluted in TBST (1:1,000) at 4°C overnight. After three 10-min washes with TBST, the membranes were incubated with the appropriate secondary antibody conjugated to HRP diluted in TBST (1:10,000) at room temperature for 1 h. After three 10-min washes with TBST, protein expression was quantified with enhanced chemiluminescence reagents (PI80196; Thermo Fisher Scientific).

Antibodies used in this study: HSP90 (610418; BD Biosciences), GAPDH (5174S; Cell Signaling), PAXIP1 (ABE1877; Sigma-Aldrich), STAG2 (SC-81852; Santa Cruz Biotech), goat anti-rabbit IgG antibody, HRP-conjugate (12-348; Sigma-Aldrich), and goat anti-mouse IgG antibody, HRP-conjugate (62-6520; Thermo Fisher Scientific).

RNA-seq on sorted cancer cells

Total RNA was prepared from FACS-sorted neoplastic cells ranging from 5×10^4 to 4×10^5 cells. RNA quality was assessed using the RNA 6000 Pico Assay Kit on the Agilent 2100 Bioanalyzer (Agilent). RNA used for RNA-seq had a mean RIN of 6.6. RNA-seq libraries were generated by Novogene, Inc. Total RNA (from 2.5 to 10 ng/sample) was used for cDNA synthesis using the Ovation RNA-seq System (NuGEN Technologies, Inc.). Two micrograms of NuGEN-amplified double-stranded cDNA was sheared using a Covaris sonicator to an average length of 490 bp and subjected to library preparation using the Takara

V4 Ultra-low input mRNA non-directional library preparation according to the manufacturer's protocol and sequenced on a NovaSeq PE150 (~6G raw data per sample).

Analysis of RNA-seq data sets

Paired-end RNA-seq reads were aligned to the mm10 mouse genome using STAR under standard input parameters. Two *KT; Stag2^{flx}* samples were excluded from subsequent analysis based on two parameters: (1) >20% the average number of nucleotide counts in Exon 8 of *STAG2*, and (2) >40% mRNA expression of *Stag2* from RNA-seq results (TPM). The differentially expressed genes between different tumor genotypes were called by DESeq2 using the HTSeq-derived counts as input (Love et al., 2014). Unsupervised hierarchical clustering and heatmap visualization of DEGs were performed using the "pheatmap" package in R (v.1.0.12; <https://github.com/raivokolde/pheatmap>).

The DESeq2-calculated fold changes were used to generate ranked gene lists for input into GSEA (v.4.0.3). GSEA was performed using normalized RNA-seq counts against gene signatures from the MSigDB database (Liberzon et al., 2015). Default parameters were used with the following exception: max size = 20,000. Plots were made using the Rtoolbox package in R (<https://github.com/PeeperLab/Rtoolbox>). clusterProfiler (v.4.3.1.900) was used to perform GO analyses, GO term enrichment, and functional enrichment analysis (Yu et al., 2012). GO terms were then categorized using EMBL's European Bioinformatics Institute (EMBL-EBI) goslim_mouse annotations and statistics analysis by "all" relationships ("occurs_in," "has_input," etc.).

ssGSEA analysis of previously published data set

KT and *KT; p53^{flx;flx}* (*KPT*) bulk RNA-seq samples were analyzed at various stages of tumor progression (55 samples) (Chuang et al., 2017). Single-sample GSEA was performed on all samples using the GSVA package v1.46.0 in R (Foroutan et al., 2018), and linear regression and unpaired *t* tests were performed in GraphPad Prism.

ATAC-seq library preparation and data analysis

After cells were sorted and counted, 5×10^4 to 1×10^5 cells were resuspended in 1 ml of cold ATAC-seq resuspension buffer (RSB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM MgCl₂ in water) (Buenrostro et al., 2015). Cells were centrifuged $500 \times g$ for 5 min at 4°C. After centrifugation, 900 µl of supernatant was aspirated. The remaining 100 µl of supernatant was carefully aspirated with a P20 pipette tip to avoid disturbing the cell pellet. Cell pellets were then resuspended in 50 µl of ATAC-seq RSB containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin and mixed by pipetting up and down five times. This reaction was incubated on ice for 3 min, and after lysis, 1 ml of ATAC-seq RSB containing 0.1% Tween-20 was added, and tubes were inverted six times to mix. Nuclei were then centrifuged for 10 min at $500 \times g$ at 4°C. The supernatant was again removed using the two-pipette steps, as described before, and nuclei were resuspended in 50 µl of transposition mix (25 µl 2× TD buffer, 2.5 µl transposase [100 nM final], 16.5 µl PBS, 0.5 µl 10% Tween-20, and 5 µl water) by pipetting up and down six times.

Transposition reactions were incubated at 37°C for 30 min at 1,000 rpm. Reactions were purified using Qiagen MinElute Reaction Cleanup Kit. ATAC-seq library preparation was performed as described using the primers listed in Table S10 and sequenced on the NovaSeq 6000 platform (Illumina) with 2 × 75 bp reads. Adapter-trimmed reads were aligned to the mm10 genome using Bowtie2 (2.1.0). Aligned reads were filtered for quality using SAMtools (v.1.9), duplicate fragments were removed using Picard (v.2.21.9-SNAPSHOT), and peaks were called using MACS2 (v.2.1.0.20150731) with a q-value cut-off of 0.01 and with a no-shift model. Peaks from replicates were merged, and read counts were obtained using bedtools (v.2.17.0) and normalized using DESeq2 (v.1.26.0). For the *KT* versus *KT;Stag2^{flx}* samples, all libraries had a TSS enrichment >10. For the *KT;H11^{LSL-Cas9}* sgInert and sgPaxip1 samples, all libraries had a TSS enrichment >8.

In situ Hi-C library construction on sorted cancer cells

In situ Hi-C libraries were prepared from FACS-sorted neoplastic cells with low input ranging from 2.5×10^5 to 3.5×10^5 cells. In detail, cells were first pelleted at $300 \times g$ and re-suspended in 1 ml Room Temp PBS containing 3% BSA and then fixed as previously described (Rao et al., 2014), only with a higher centrifuge speed at $2,500 \times g$ for pelleting the fixed cells. Fixed cell pellets were processed through the standard Arima Hi-C kit (Catalog #A510008) until the step of Covaris DNA sonication to an average size of 400 bp. Sheared DNA was purified and size-selected by a two-step AMPure XP bead (#A63882; Beckman Coulter) cleanup at 0.6× and 1.0×, and then inputted into the 2S Plus DNA Library Kit from Integrated DNA Technologies (Catalog #10009878) for low-input library preparation until the step before library PCR amplification. Libraries underwent a biotin pulldown using streptavidin to enrich chromatin interactions before an initial five-cycle amplification was performed and a quantitative PCR quantification determining the number of additional cycles required for each library to reach a final concentration around 20 nM in 20 μ l. Amplified libraries were then examined by Agilent 4200 TapeStation for confirmation of size distribution before submission for sequencing. Libraries were first sequenced using the Illumina MiniSeq PE50 for quality control (analysis described in the next section) and then deeply sequenced using the Illumina NovaSeq PE50 aiming at $>3 \times 10^8$ read pairs for each library that passed the MiniSeq QC.

Analysis of in situ Hi-C data

Hi-C data was aligned to the mm10 reference genome using MWA-MEM (Li, 2013, Preprint). Reads were filtered (MAPQ ≥ 30) and paired using the previously described pipeline (Dixon et al., 2018). PCR duplicate reads were detected and removed by Picard (<https://broadinstitute.github.io/picard/>). For MiniSeq quality control analysis, we sorted unique read pairs into three categories: (1) cis-reads that aligned to the same chromosome with a distance over 1 kb, (2) cis-reads that aligned to the same chromosome with a distance equal or <1 kb, and (3) trans-reads that aligned to two different chromosomes. Hi-C libraries with a ratio of cis-reads with distances over 1 kb

>50% and trans-reads ratio <25% proceeded with deep sequencing using the NovaSeq.

For deeply sequenced Hi-C data, .hic files were generated and normalized (-w 5,000) using JuiceBox (Durand et al., 2016) after alignment and duplication removal described above. Hi-C data from replicates within the same treatment were merged before visualization and loop-calling using JuiceBox. For loop-calling, genome-wide KR normalization and three resolutions were used (-k GW_KR -r 5,000; 10,000; and 25,000). Called loop lists from wildtype and *Stag2*-deficient neoplastic cells were compared with each other and sorted into indicated categories: (1) common loops that both loop anchors have overlap between *KT* and *KT;Stag2^{flx}* samples, (2) unique loops that neither loop anchors have overlap between *KT* and *KT;Stag2^{flx}* samples, (3) loops that are unique to *KT* that have one loop anchor that overlaps with a loop unique to *KT;Stag2^{flx}* samples, (4) loops that are unique to *KT* that have one loop anchor that overlaps with common loops between *KT* and *KT;Stag2^{flx}* samples, (5) loops that are unique to *KT;Stag2^{flx}* that have one loop anchor that overlaps with common loops between *KT* and *KT;Stag2^{flx}* samples, (6) loops that are unique to *KT;Stag2^{flx}* that have one loop anchor that overlaps with a loop unique to *KT* samples (Fig. 5 e). For loop intensity analysis, loops from both *KT* and *KT;Stag2^{flx}* samples were concatenated and removed for redundancy before using the juicer dump function to output the ratio of observed/expected for every loop in each separate sample (Table S11).

Histology and immunohistochemistry (IHC)

Lung lobes were preserved in 4% formalin and embedded in paraffin. Hematoxylin and eosin stains were performed by Stanford Pathology/Histology Service Center (Stanford, CA, USA) or Histo-Tec Laboratories (Hayward, CA, USA). Total tumor burden (tumor area/total area × 100%) and individual tumor sizes were determined using ImageJ. IHC was conducted on 4- μ m sections. Antigen retrieval was performed with 10 mM citrate buffer in a pressurized decloaking chamber. The slides were washed with 1× PBST and prepared with VECTASTAIN ABC-HRP Kit (PK-4000; Vector Laboratories). The following primary antibody with its dilution was used: NKX2-1 (1:250, ab7013; Abcam). Sections were developed with DAB (SK-4100; Vector Laboratories) and counterstained with hematoxylin. To assess NKX2-1 expression in tumors while taking into account potential differences in staining between samples and across sections, we compared NKX2-1 staining intensity between tumors and adjacent normal tissue. Tumors were binned as having (1) higher expression relative to adjacent normal tissue, (2) equal expression relative to adjacent normal tissue, or (3) lower expression relative to adjacent normal tissue.

Online supplemental material

Fig. S1 shows that the effect of *STAG2* inactivation is highly correlated with that of *PAXIP1* and *PAGRI* inactivation in human cancer cell lines and genetic interactions with *STAG2* are robust to different measures of tumorigenesis and tumor growth. Fig. S2 shows validation of *STAG2* inactivation, clustering of *KT* and *KT;Stag2^{flx}* samples, and gene expression of surfactant genes regulated by *STAG2*, TSS enrichment, and TF enrichment for *KT*

and *KT;Stag2^{flx}* samples. **Fig. S3** shows DNA loops and DNA anchors that correlate with gene expression but not chromatin accessibility when comparing *KT* and *KT;Stag2^{flx}* samples. **Fig. S4** shows that PAXIP1 and STAG2-cohesin regulation of gene expression is conserved in downregulation but not upregulation of genes, *Paxip1* clustering, TSS enrichment, and TF enrichment for *KT;H11^{LSL-Cas9} sgInert* and *KT;H11^{LSL-Cas9}*. **Fig. S5** shows the summary of the in vivo genetic experiments as specific tumor suppressive function of STAG2-cohesin and PAXIP1/PAGR1, as STAG2-cohesin and PAXIP1/PAGR1 mediated tumor suppression, and mutations in *PAXIP1* and *PAGR1* and downregulation of *PAXIP1* mRNA expression in a subset of AMLs. Table S1 shows tumor size and number metrics from the cohesin complex member Tuba-seq^{Ultra} analysis. Table S2 shows tumor size and number metrics from *Paxip1*, *Pagr1*, *Cdca5*, and various TSGs and essential genes Tuba-seq^{Ultra} analysis. Table S3 shows genes selected for Tuba-seq^{Ultra} analysis of putative genetic interactors of STAG2. Table S4 shows tumor size and number metrics from Tuba-seq^{Ultra} analysis of putative genetic interactors of STAG2. Table S5 shows metrics and RLE-normalized (DESeq2) counts for *KT;Stag2^{flx}* versus *KT* 8 and 16 wk after tumor initiation RNA-seq experiment. Table S6 shows metrics and RLE-normalized (DESeq2) counts for *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* RNA-seq experiment. Table S7 shows metrics for differentially expressed genes altered by both STAG2 and PAXIP1 with $\log_2\text{FoldChange} > (\text{abs}) 1$. Table S8 shows metrics and RLE-normalized (DESeq2) counts for *KT;Stag2^{flx}* versus *KT* 16 wk after tumor initiation ATAC-seq experiment. Table S9 shows metrics and RLE-normalized (DESeq2) counts for *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* 16 wk after tumor initiation ATAC-seq experiment. Table S10 shows a list of forward and reverse primers used to generate ATACseq libraries for *KT;Stag2^{flx}* versus *KT* and *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* ATACseq experiments. Table S11 shows loci for all loops and normalized metrics for *KT;Stag2^{flx}* versus *KT* 16 wk after tumor initiation Hi-C experiment.

Data availability

All data generated or analyzed during this study are included in this published article (and its supplementary information files). RNA-seq, ATAC-seq, and Hi-C metadata and raw sequencing files can also be found in the Gene Expression Omnibus (accession IDs: GSE274204, GSE274205, and GSE277290).

Acknowledgments

We thank the Stanford Shared FACS Facility for flow cytometry and cell sorting services, the Stanford Veterinary Animal Care Staff for expert animal care, Human Pathology/Histology Service Center, and the Stanford Protein and Nucleic Acids Facility. We would like to acknowledge the American Association for Cancer Research and its financial and material support in the development of the AACR Project GENIE registry, as well as members of the consortium for their commitment to data sharing. Interpretations are the responsibility of the study authors. We thank D. Solomon, D. Hargreaves, J. Sage, L. Attardi, P. Khavari, J. Carette, P. Du, B. Chick, N. Avina Ochoa, and

members of the Winslow and Crabtree laboratories for helpful comments and experimental support. We thank A. Melnick for providing *Smc3^{flx}* mice.

E.L. Ashkin was supported by the Howard Hughes Medical Institute (HHMI) Gilliam Fellowship for Advanced Study (GT14928) and a National Cancer Institute Predoctoral to Postdoctoral Fellow Transition Award (F99CA284289). Y.J. Tang was supported by the Canadian Institute of Health Research (CIHR) postdoctoral fellowship (MFE-176568). K.L. Hung was supported by a Stanford Graduate Fellowship and a National Cancer Institute Predoctoral to Postdoctoral Fellow Transition Award (F99CA274692). J.A. Belk was supported by the HHMI Hanna H. Gray Fellowship Award. H. Cai was supported by a Tobacco-Related Disease Research Program (TRDRP) fellowship (28FT-0019). D.N. Dolcen was supported by the National Cancer Institute Predoctoral to Postdoctoral Fellow Transition Award (K01-CA245784). J.D. Hebert was supported by an American Cancer Society postdoctoral fellowship (PF-21-112-01-MM) and a TRDRP fellowship (T31FT1619). P.A. Ruiz was supported by the National Science Foundation Graduate Research Fellowship Program (DGE-2146755) and the Stanford Graduate Fellowship. This work was supported by National Institutes of Health (NIH) R01-CA231253 (to M.M. Winslow) and R01-CA234349 (to M.M. Winslow and D.A. Petrov) and in part by the Stanford Cancer Institute support grant (NIH P30-CA124435).

Author contributions: E.L. Ashkin: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing, Y.J. Tang: Data curation, Methodology, H. Xu: Data curation, Formal analysis, Visualization, Writing - original draft, Writing - review & editing, K.L. Hung: Conceptualization, Formal analysis, Methodology, Resources, Writing - review & editing, J.A. Belk: Data curation, Formal analysis, Investigation, Software, Writing - review & editing, H. Cai: Investigation, Resources, Writing - review & editing, S.S. Lopez: Data curation, Writing - review & editing, D.N. Dolcen: Conceptualization, J.D. Hebert: Conceptualization, Writing - review & editing, R. Li: Investigation, P.A. Ruiz: Investigation, Visualization, Writing - review & editing, T. Keal: Formal analysis, Investigation, L. Andrejka: Investigation, H.Y. Chang: Resources, D.A. Petrov: Investigation, J.R. Dixon: Methodology, Supervision, Writing - review & editing, Z. Xu: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing - original draft, Writing - review & editing, M.M. Winslow: Conceptualization, Funding acquisition, Supervision, Visualization, Writing - original draft.

Disclosures: H.Y. Chang is a co-founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, and Orbital Therapeutics, and an advisor of 10x Genomics, Arsenal Bio, Chroma Medicine, Exai Bio, and Spring Discovery. M.M. Winslow and D.A. Petrov reported grants from the NIH (R01-CA231253, R01-CA234349, and P30-CA124435) during the conduct of the study and are founders and equity holders in Guide Oncology Inc. No other disclosures were reported.

Submitted: 30 April 2024
 Revised: 16 September 2024
 Accepted: 30 October 2024

References

AACR Project GENIE Consortium. 2017. AACR Project GENIE: Powering precision medicine through an international consortium. *Cancer Discov.* 7:818–831. <https://doi.org/10.1158/2159-8290.CD-17-0151>

Adane, B., G. Alexe, B.K.A. Seong, D. Lu, E.E. Hwang, D. Hnisz, C.A. Lareau, L. Ross, S. Lin, F.S. Dela Cruz, et al. 2021. STAG2 loss rewires oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer Cell.* 39:827–844.e10. <https://doi.org/10.1016/j.ccr.2021.05.007>

Alonso-Gil, D., and A. Losada. 2023. NIPBL and cohesin: New take on a classic tale. *Trends Cell Biol.* 33:860–871. <https://doi.org/10.1016/j.tcb.2023.03.006>

Aquila, L., J. Ohm, and A. Woloszynska-Read. 2018. The role of STAG2 in bladder cancer. *Pharmacol. Res.* 131:143–149. <https://doi.org/10.1016/j.phrs.2018.02.025>

Arruda, N.L., Z.M. Carico, M. Justice, Y.F. Liu, J. Zhou, H.C. Stefan, J.M. Dowen, J.T. Lee, and E.F. Joyce. 2020. Distinct and overlapping roles of STAG1 and STAG2 in cohesin localization and gene expression in embryonic stem cells. *Epigenetics Chromatin.* 13:32. <https://doi.org/10.1186/s13072-020-00353-9>

Bailey, M.L., N.J. O’Neil, D.M. van Pel, D.A. Solomon, T. Waldman, and P. Hieter. 2014. Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition. *Mol. Cancer Ther.* 13:724–732. <https://doi.org/10.1158/1535-7163.MCT-13-0749>

Bailey, M.L., D. Tieu, A. Habsid, A.H.Y. Tong, K. Chan, J. Moffat, and P. Hieter. 2021. Paralogous synthetic lethality underlies genetic dependencies of the cancer-mutated gene STAG2. *Life Sci. Alliance.* 4:e202101083. <https://doi.org/10.26508/lsci.202101083>

Ball, A.R., Y.-Y. Chen, and K. Yokomori. 2014. Mechanisms of cohesin-mediated gene regulation and lessons learned from cohesinopathies. *Biochim. Biophys. Acta.* 1893:191–202. <https://doi.org/10.1016/j.bbagen.2013.11.002>

Benjamini, Y., and Y. Hochberg. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Stat. Soc. B.* 57:289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>

Blair, L.M., J.M. Juan, L. Sebastian, V.B. Tran, W. Nie, G.D. Wall, M. Gerceker, I.K. Lai, E.A. Apilado, G. Grenot, et al. 2023. Oncogenic context shapes the fitness landscape of tumor suppression. *Nat. Commun.* 14:6422. <https://doi.org/10.1038/s41467-023-42156-y>

Buenrostro, J.D., B. Wu, H.Y. Chang, and W.J. Greenleaf. 2015. ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109:21.29.1–21.29.9. <https://doi.org/10.1002/0471142727.mbo2129s109>

Cai, H., S.K. Chew, C. Li, M.K. Tsai, L. Andrejka, C.W. Murray, N.W. Hughes, E.G. Shuldiner, E.L. Ashkin, R. Tang, et al. 2021. A functional taxonomy of tumor suppression in oncogenic KRAS-driven lung cancer. *Cancer Discov.* 11:1754–1773. <https://doi.org/10.1158/2159-8290.CD-20-1325>

Callen, E., R.B. Faryabi, M. Luckey, B. Hao, J.A. Daniel, W. Yang, H.W. Sun, G. Dressler, W. Peng, H. Chi, et al. 2012. The DNA damage- and transcription-associated protein paxip1 controls thymocyte development and emigration. *Immunity.* 37:971–985. <https://doi.org/10.1016/j.jimmuni.2012.10.007>

Canudas, S., and S. Smith. 2009. Differential regulation of telomere and centromere cohesion by the Scc3 homologues SA1 and SA2, respectively, in human cells. *J. Cell Biol.* 187:165–173. <https://doi.org/10.1083/jcb.200903096>

Cerami, E., J. Gao, U. Dogrusoz, B.E. Gross, S.O. Sumer, B.A. Aksoy, A. Jacobsen, C.J. Byrne, M.L. Heuer, E. Larsson, et al. 2012. The cBio cancer genomics portal: An open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* 2:401–404. <https://doi.org/10.1158/2159-8290.CD-12-0095>

Chuang, C.-H., P.G. Greenside, Z.N. Rogers, J.J. Brady, D. Yang, R.K. Ma, D.R. Caswell, S.H. Chiou, A.F. Winters, B.M. Grüner, et al. 2017. Molecular definition of a metastatic lung cancer state reveals a targetable CD109-Janus kinase-Stat axis. *Nat. Med.* 23:291–300. <https://doi.org/10.1038/nm.4285>

Crompton, B.D., C. Stewart, A. Taylor-Weiner, G. Alexe, K.C. Kurek, M.L. Calicchio, A. Kiezun, S.L. Carter, S.A. Shukla, S.S. Mehta, et al. 2014. The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* 4:1326–1341. <https://doi.org/10.1158/2159-8290.CD-13-1037>

Cuadrado, A., D. Giménez-Llorente, A. Kojic, M. Rodríguez-Corsino, Y. Cuartero, G. Martín-Serrano, G. Gómez-López, M.A. Martí-Renom, and A. Losada. 2019. Specific contributions of cohesin-SAI and cohesin-SA2 to TADs and Polycomb domains in embryonic stem cells. *Cell Rep.* 27:3500–3510.e4. <https://doi.org/10.1016/j.celrep.2019.05.078>

de Bruijn, I., R. Kundra, B. Mastrogiacomo, T.N. Tran, L. Sikina, T. Mazor, X. Li, A. Ochoa, G. Zhao, B. Lai, et al. 2023. Analysis and visualization of longitudinal genomic and clinical data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. *Cancer Res.* 83:3861–3867. <https://doi.org/10.1158/0008-5472.CAN-23-0816>

De Koninck, M., E. Lapi, C. Badia-Careaga, I. Cossío, D. Giménez-Llorente, M. Rodríguez-Corsino, E. Andrada, A. Hidalgo, M. Manzanares, F.X. Real, and A. Losada. 2020. Essential roles of cohesin STAG2 in mouse embryonic development and adult tissue homeostasis. *Cell Rep.* 32:108014. <https://doi.org/10.1016/j.celrep.2020.108014>

DeWeirdt, P.C., A.V. McGee, F. Zheng, I. Nwolah, M. Hegde, and J.G. Doench. 2022. Accounting for small variations in the tracrRNA sequence improves sgRNA activity predictions for CRISPR screening. *Nat. Commun.* 13:5255. <https://doi.org/10.1038/s41467-022-33024-2>

Ding, S., J. Diep, N. Feng, L. Ren, B. Li, Y.S. Ooi, X. Wang, K.F. Bruliois, L.L. Yasukawa, X. Li, et al. 2018. STAG2 deficiency induces interferon responses via cGAS-STING pathway and restricts virus infection. *Nat. Commun.* 9:1485. <https://doi.org/10.1038/s41467-018-03782-z>

Dixon, J.R., J. Xu, V. Dileep, Y. Zhan, F. Song, V.T. Le, G.G. Yardimci, A. Chakraborty, D.V. Bann, Y. Wang, et al. 2018. Integrative detection and analysis of structural variation in cancer genomes. *Nat. Genet.* 50:1388–1398. <https://doi.org/10.1038/s41588-018-0195-8>

Doench, J.G., N. Fusi, M. Sullender, M. Hegde, E.W. Vainberg, K.F. Donovan, I. Smith, Z. Tothova, C. Wilen, R. Orchard, et al. 2016. Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* 34:184–191. <https://doi.org/10.1038/nbt.3437>

DuPage, M., A.L. Dooley, and T. Jacks. 2009. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protoc.* 4:1064–1072. <https://doi.org/10.1038/nprot.2009.95>

Durand, N.C., J.T. Robinson, M.S. Shamim, I. Machol, J.P. Mesirov, E.S. Lander, and E.L. Aiden. 2016. Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom. *Cell Syst.* 3:99–101. <https://doi.org/10.1016/j.cels.2015.07.012>

Evers, L., P.A. Perez-Mancera, E. Lenkiewicz, N. Tang, D. Aust, T. Knösel, P. Rümmel, T. Holley, M. Kassner, M. Aziz, et al. 2014. STAG2 is a clinically relevant tumor suppressor in pancreatic ductal adenocarcinoma. *Genome Med.* 6:9. <https://doi.org/10.1186/gm526>

Foroutan, M., D.D. Bhuva, R. Lyu, K. Horan, J. Cursons, and M.J. Davis. 2018. Single sample scoring of molecular phenotypes. *BMC Bioinformatics.* 19:404. <https://doi.org/10.1186/s12859-018-2435-4>

Gao, J., B.A. Aksoy, U. Dogrusoz, G. Dresdner, B. Gross, S.O. Sumer, Y. Sun, A. Jacobsen, R. Sinha, E. Larsson, et al. 2013. Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci. Signal.* 6:pl1. <https://doi.org/10.1126/scisignal.2004088>

Gong, Z., Y.-W. Cho, J.-E. Kim, K. Ge, and J. Chen. 2009. Accumulation of Pax2 transactivation domain interaction protein (PTIP) at sites of DNA breaks via RNF8-dependent pathway is required for cell survival after DNA damage. *J. Biol. Chem.* 284:7284–7293. <https://doi.org/10.1074/jbc.M809158200>

Grimley, E., and G.R. Dressler. 2018. Are Pax proteins potential therapeutic targets in kidney disease and cancer? *Kidney Int.* 94:259–267. <https://doi.org/10.1016/j.kint.2018.01.025>

Hara, K., G. Zheng, Q. Qu, H. Liu, Z. Ouyang, Z. Chen, D.R. Tomchick, and H. Yu. 2014. Structure of cohesin subcomplex pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat. Struct. Mol. Biol.* 21:864–870. <https://doi.org/10.1038/nsmb.2880>

He, H., S.M. Bell, A.K. Davis, S. Zhao, A. Sridharan, C.-L. Na, E. Gao, A. Mathews, Q. Zhong, L. Zhang, et al. 2023. PRDM3/16 regulate chromatin accessibility required for NFKX2-1 mediated alveolar epithelial differentiation and function. *bioRxiv.* <https://doi.org/10.1101/2023.12.20.570481> (Preprint posted December 20, 2023).

Hwang, D.H., L.M. Sholl, V. Rojas-Rudilla, D.L. Hall, P. Shivdasani, E.P. Garcia, L.E. MacConaill, M. Vivero, J.L. Hornick, F.C. Kuo, et al. 2016. KRAS and NFKX2-1 mutations in invasive mucinous adenocarcinoma of the lung. *J. Thorac. Oncol.* 11:496–503. <https://doi.org/10.1016/j.jtho.2016.01.010>

Ingram, K., S.C. Samson, R. Zewdu, R.G. Zitnay, E.L. Snyder, and M.C. Mendoza. 2022. NFKX2-1 controls lung cancer progression by inducing DUSP6 to dampen ERK activity. *Oncogene.* 41:293–300. <https://doi.org/10.1038/s41388-021-02076-x>

Jackson, E.L., N. Willis, K. Mercer, R.T. Bronson, D. Crowley, R. Montoya, T. Jacks, and D.A. Tuvesson. 2001. Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. *Genes Dev.* 15:3243–3248. <https://doi.org/10.1101/gad.94301>

Kim, E., M. Dede, W.F. Lenoir, G. Wang, S. Srivivasan, M. Colic, and T. Hart. 2019a. A network of human functional gene interactions from knockout fitness screens in cancer cells. *Life Sci. Alliance.* 2:e201800278. <https://doi.org/10.26508/lسا.201800278>

Kim, J.-S., X. He, J. Liu, Z. Duan, T. Kim, J. Gerard, B. Kim, M.M. Pillai, W.S. Lane, W.S. Noble, et al. 2019b. Systematic proteomics of endogenous human cohesin reveals an interaction with diverse splicing factors and RNA-binding proteins required for mitotic progression. *J. Biol. Chem.* 294:8760–8772. <https://doi.org/10.1074/jbc.RA119.007832>

Kim, M.S., S.S. Kim, E.M. Je, N.J. Yoo, and S.H. Lee. 2012. Mutational and expressional analyses of STAG2 gene in solid cancers. *Neoplasma.* 59: 524–529. https://doi.org/10.4149/neo_2012_067

Kojic, A., A. Cuadrado, M. De Koninck, D. Giménez-Llorente, M. Rodríguez-Corsino, G. Gómez-López, F. Le Dily, M.A. Martí-Renom, and A. Losada. 2018. Distinct roles of cohesin-SAI and cohesin-SA2 in 3D chromosome organization. *Nat. Struct. Mol. Biol.* 25:496–504. <https://doi.org/10.1038/s41594-018-0070-4>

Kong, X., A.R. Ball Jr., H.X. Pham, W. Zeng, H.-Y. Chen, J.A. Schmiesing, J.S. Kim, M. Berns, and K. Yokomori. 2014. Distinct functions of human cohesin-SAI and cohesin-SA2 in double-strand break repair. *Mol. Cell. Biol.* 34:685–698. <https://doi.org/10.1128/MCB.01503-13>

Lechner, M.S., I. Levitan, and G.R. Dressler. 2000. PTIP, a novel BRCT domain-containing protein interacts with Pax2 and is associated with active chromatin. *Nucleic Acids Res.* 28:2741–2751. <https://doi.org/10.1093/nar/28.14.2741>

Li, H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv.* <https://doi.org/10.48550/arXiv.1303.3997> (Preprint posted March 16, 2013).

Liberzon, A., C. Birger, H. Thorvaldsdóttir, M. Ghandi, J.P. Mesirov, and P. Tamayo. 2015. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 1:417–425. <https://doi.org/10.1016/j.cels.2015.12.004>

Liu, B., and Z. Li. 2022. PTIP-associated protein 1: More than a component of the MLL3/4 complex. *Front. Genet.* 13:889109. <https://doi.org/10.3389/fgene.2022.889109>

Love, M.I., W. Huber, and S. Anders. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15: 550. <https://doi.org/10.1186/s13059-014-0550-8>

Marjanovic, N.D., M. Hofree, J.E. Chan, D. Canner, K. Wu, M. Trakala, G.G. Hartmann, O.C. Smith, J.Y. Kim, K.V. Evans, et al. 2020. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell.* 38: 229–246.e13. <https://doi.org/10.1016/j.ccel.2020.06.012>

Mayayo-Peralta, I., S. Gregoricchio, K. Schuurman, S. Yavuz, A. Zaalberg, A. Kojic, N. Abbott, B. Geverts, S. Beertthuijzen, J. Siefert, et al. 2023. PAXIP1 and STAG2 converge to maintain 3D genome architecture and facilitate promoter/enhancer contacts to enable stress hormone-dependent transcription. *Nucleic Acids Res.* 51:9576–9593. <https://doi.org/10.1093/nar/gkad267>

Mondal, G., M. Stevers, B. Goode, A. Ashworth, and D.A. Solomon. 2019. A requirement for STAG2 in replication fork progression creates a targetable synthetic lethality in cohesin-mutant cancers. *Nat. Commun.* 10: 1686. <https://doi.org/10.1038/s41467-019-09659-z>

Pan, J., R.M. Meyers, B.C. Michel, N. Mashtalir, A.E. Sizemore, J.N. Wells, S.H. Cassel, F. Vazquez, B.A. Weir, W.C. Hahn, et al. 2018. Interrogation of mammalian protein complex structure, function, and membership using genome-scale fitness screens. *Cell Syst.* 6:555–568.e7. <https://doi.org/10.1016/j.cels.2018.04.011>

Rao, S.S.P., S.-C. Huang, B. Glenn St Hilaire, J.M. Engreitz, E.M. Perez, K.R. Kieffer-Kwon, A.L. Sanborn, S.E. Johnstone, G.D. Bascom, I.D. Bochkov, et al. 2017. Cohesin loss eliminates all loop domains. *Cell.* 171:305–320.e24. <https://doi.org/10.1016/j.cell.2017.09.026>

Rao, S.S.P., M.H. Huntley, N.C. Durand, E.K. Stamenova, I.D. Bochkov, J.T. Robinson, A.L. Sanborn, I. Machol, A.D. Omer, E.S. Lander, and E.L. Aiden. 2014. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell.* 159:1665–1680. <https://doi.org/10.1016/j.cell.2014.11.021>

Ren, G., W. Jin, K. Cui, J. Rodriguez, G. Hu, Z. Zhang, D.R. Larson, and K. Zhao. 2017. CTCF-mediated enhancer-promoter interaction is a critical regulator of cell-to-cell variation of gene expression. *Mol. Cell.* 67:1049–1058.e6. <https://doi.org/10.1016/j.molcel.2017.08.026>

Richart, L., E. Lapi, V. Pancaldi, M. Cuenca-Ardura, E.C.-S. Pau, M. Madrid-Mencía, H. Neyret-Kahn, F. Radvanyi, J.A. Rodríguez, Y. Cuartero, et al. 2021. STAG2 loss-of-function affects short-range genomic contacts and modulates the basal-luminal transcriptional program of bladder cancer cells. *Nucleic Acids Res.* 49:11005–11021. <https://doi.org/10.1093/nar/gkab864>

Rivas, M.A., C. Durmaz, A. Kloetgen, C.R. Chin, Z. Chen, B. Bhinder, A. Koren, A.D. Viny, C.D. Scharer, J.M. Boss, et al. 2021a. Cohesin core complex gene dosage contributes to germinal center derived lymphoma phenotypes and outcomes. *Front. Immunol.* 12:688493. <https://doi.org/10.3389/fimmu.2021.688493>

Rivas, M.A., C. Meydan, C.R. Chin, M.F. Challman, D. Kim, B. Bhinder, A. Kloetgen, A.D. Viny, M.R. Teater, D.R. McNally, et al. 2021b. Smc3 dosage regulates B cell transit through germinal centers and restricts their malignant transformation. *Nat. Immunol.* 22:240–253. <https://doi.org/10.1038/s41590-020-00827-8>

Rogers, Z.N., C.D. McFarland, I.P. Winters, J.A. Seoane, J.J. Brady, S. Yoon, C. Curtis, D.A. Petrov, and M.M. Winslow. 2018. Mapping the *in vivo* fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat. Genet.* 50:483–486. <https://doi.org/10.1038/s41588-018-0083-2>

Rogers, Z.N., C.D. McFarland, I.P. Winters, S. Naranjo, C.H. Chuang, D. Petrov, and M.M. Winslow. 2017. A quantitative and multiplexed approach to uncover the fitness landscape of tumor suppression *in vivo*. *Nat. Methods.* 14:737–742. <https://doi.org/10.1038/nmeth.4297>

Schubert, M., S. Lindgreen, and L. Orlando. 2016. AdapterRemoval v2: Rapid adapter trimming, identification, and read merging. *BMC Res. Notes.* 9: 88. <https://doi.org/10.1186/s13104-016-1900-2>

Serresi, M., G. Gargiulo, N. Proost, B. Siteur, M. Cesaroni, M. Koppens, H. Xie, K.D. Sutherland, D. Hulsman, E. Citterio, et al. 2016. Polycomb Repressive Complex 2 is a barrier to KRAS-driven inflammation and epithelial-mesenchymal transition in non-small-cell lung cancer. *Cancer Cell.* 29:17–31. <https://doi.org/10.1016/j.ccr.2015.12.006>

Serresi, M., B. Siteur, D. Hulsman, C. Company, M.J. Schmitt, C. Lieftink, B. Morris, M. Cesaroni, N. Proost, R.L. Beijersbergen, et al. 2018. Ezh2 inhibition in Kras-driven lung cancer amplifies inflammation and associated vulnerabilities. *J. Exp. Med.* 215:3115–3135. <https://doi.org/10.1084/jem.20180801>

Shen, C.-H., S.H. Kim, S. Trousil, D.T. Frederick, A. Piris, P. Yuan, L. Cai, L. Gu, M. Li, J.H. Lee, et al. 2016. Loss of cohesin complex components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. *Nat. Med.* 22:1056–1061. <https://doi.org/10.1038/nm.4155>

Solomon, D.A., J.-S. Kim, J. Bondaruk, S.F. Shariat, Z.-F. Wang, A.G. Elkahloun, T. Ozawa, J. Gerard, D. Zhuang, S. Zhang, et al. 2013. Frequent truncating mutations of STAG2 in bladder cancer. *Nat. Genet.* 45: 1428–1430. <https://doi.org/10.1038/ng.2800>

Solomon, D.A., T. Kim, L.A. Diaz-Martinez, J. Fair, A.G. Elkahloun, B.T. Harris, J.A. Toretsky, S.A. Rosenberg, N. Shukla, M. Ladanyi, et al. 2011. Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science.* 333:1039–1043. <https://doi.org/10.1126/science.1203619>

Tang, Y., H. Xu, N.W. Hughes, S.H. Kim, P. Ruiz, E.G. Shuldin, S.S. Lopez, J.D. Hebert, S. Karmakar, L. Andrejka, et al. 2024. Functional mapping of epigenetic regulators uncovers coordinated tumor suppression by HBO1 and MLL1 complexes. *bioRxiv.* <https://doi.org/10.1101/2024.08.19.607671> (Preprint posted August 20, 2024).

Taylor, C.F., F.M. Platt, C.D. Hurst, H.H. Thygesen, and M.A. Knowles. 2014. Frequent inactivating mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and inversely related to chromosomal copy number changes. *Hum. Mol. Genet.* 23:1964–1974. <https://doi.org/10.1093/hmg/ddt589>

Toba-Ichihashi, Y., T. Yamaoka, T. Ohmori, and M. Ohba. 2015. Up-regulation of Syndecan-4 contributes to TGF-β1-induced epithelial to mesenchymal transition in lung adenocarcinoma A549 cells. *Biochem. Biophys. Rep.* 5:1–7. <https://doi.org/10.1016/j.bbrep.2015.11.021>

van der Lelij, P., S. Lieb, J. Jude, G. Wutz, C.P. Santos, K. Falkenberg, A. Schlattl, J. Ban, R. Schwentner, T. Hoffmann, et al. 2017. Synthetic lethality between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. *eLife.* 6:e26980. <https://doi.org/10.7554/eLife.26980>

van der Lelij, P., J.A. Newman, S. Lieb, J. Jude, V. Katis, T. Hoffmann, M. Hinterndorfer, G. Bader, N. Kraut, M.A. Pearson, et al. 2020. STAG1 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. *Life Sci. Alliance.* 3:e202000725. <https://doi.org/10.26508/lسا.202000725>

van Schie, J.J.M., K. de Lint, T.M. Molenaar, M. Moronta Gines, J.A. Balk, M.A. Rooimans, K. Roohollahi, G.M. Pai, L. Borghuis, A.R. Ramadhin, et al. 2023. CRISPR screens in sister chromatid cohesion defective cells reveal

PAXIP1-PAGR1 as regulator of chromatin association of cohesin. *Nucleic Acids Res.* 51:9594–9609. <https://doi.org/10.1093/nar/gkad756>

Viny, A.D., R.L. Bowman, Y. Liu, V.P. Lavallée, S.E. Eisman, W. Xiao, B.H. Durham, A. Navitski, J. Park, S. Braunstein, et al. 2019. Cohesin members Stag1 and Stag2 display distinct roles in chromatin accessibility and topological control of HSC self-renewal and differentiation. *Cell Stem Cell.* 25:682–696.e8. <https://doi.org/10.1016/j.stem.2019.08.003>

Viny, A.D., C.J. Ott, B. Spitzer, M. Rivas, C. Meydan, E. Papalex, D. Yelin, K. Shank, J. Reyes, A. Chiu, et al. 2015. Dose-dependent role of the cohesin complex in normal and malignant hematopoiesis. *J. Exp. Med.* 212: 1819–1832. <https://doi.org/10.1084/jem.20151317>

Weir, B.A., M.S. Woo, G. Getz, S. Perner, L. Ding, R. Beroukhim, W.M. Lin, M.A. Province, A. Kraja, L.A. Johnson, et al. 2007. Characterizing the cancer genome in lung adenocarcinoma. *Nature.* 450:893–898. <https://doi.org/10.1038/nature06358>

Winslow, M.M., T.L. Dayton, R.G. Verhaak, C. Kim-Kiselak, E.L. Snyder, D.M. Feldser, D.D. Hubbard, M.J. DuPage, C.A. Whittaker, S. Hoersch, et al. 2011. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature.* 473:101–104. <https://doi.org/10.1038/nature09881>

Yamaguchi, T., Y. Hosono, K. Yanagisawa, and T. Takahashi. 2013. NKX2-1/TTF-1: An enigmatic oncogene that functions as a double-edged sword for cancer cell survival and progression. *Cancer Cell.* 23:718–723. <https://doi.org/10.1016/j.ccr.2013.04.002>

Yan, J., S.-A.A. Chen, A. Local, T. Liu, Y. Qiu, K.M. Dorighi, S. Preissl, C.M. Rivera, C. Wang, Z. Ye, et al. 2018. Histone H3 lysine 4 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res.* 28:204–220. <https://doi.org/10.1038/cr.2018.1>

Yang, L., M. Lin, W.J. Ruan, L.L. Dong, E.G. Chen, X.H. Wu, K.J. Ying, Y. Guo, Y. Zhang, and C. Zhang. 2012. Nkx2-1: A novel tumor biomarker of lung cancer. *J. Zhejiang Univ. Sci. B.* 13:855–866. <https://doi.org/10.1631/jzus.B1100382>

Yousefi, M., G. Boross, C. Weiss, C.W. Murray, J.D. Hebert, H. Cai, E.L. Ashkin, S. Karmakar, L. Andrejka, L. Chen, et al. 2022. Combinatorial inactivation of tumor suppressors efficiently initiates lung adenocarcinoma with therapeutic vulnerabilities. *Cancer Res.* 82:1589–1602. <https://doi.org/10.1158/0008-5472.CAN-22-0059>

Yu, G., L.-G. Wang, Y. Han, and Q.-Y. He. 2012. clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS.* 16: 284–287. <https://doi.org/10.1089/omi.2011.0118>

Zheng, R., and G.A. Blobel. 2010. GATA transcription factors and cancer. *Genes Cancer.* 1:1178–1188. <https://doi.org/10.1177/1947601911404223>

Supplemental material

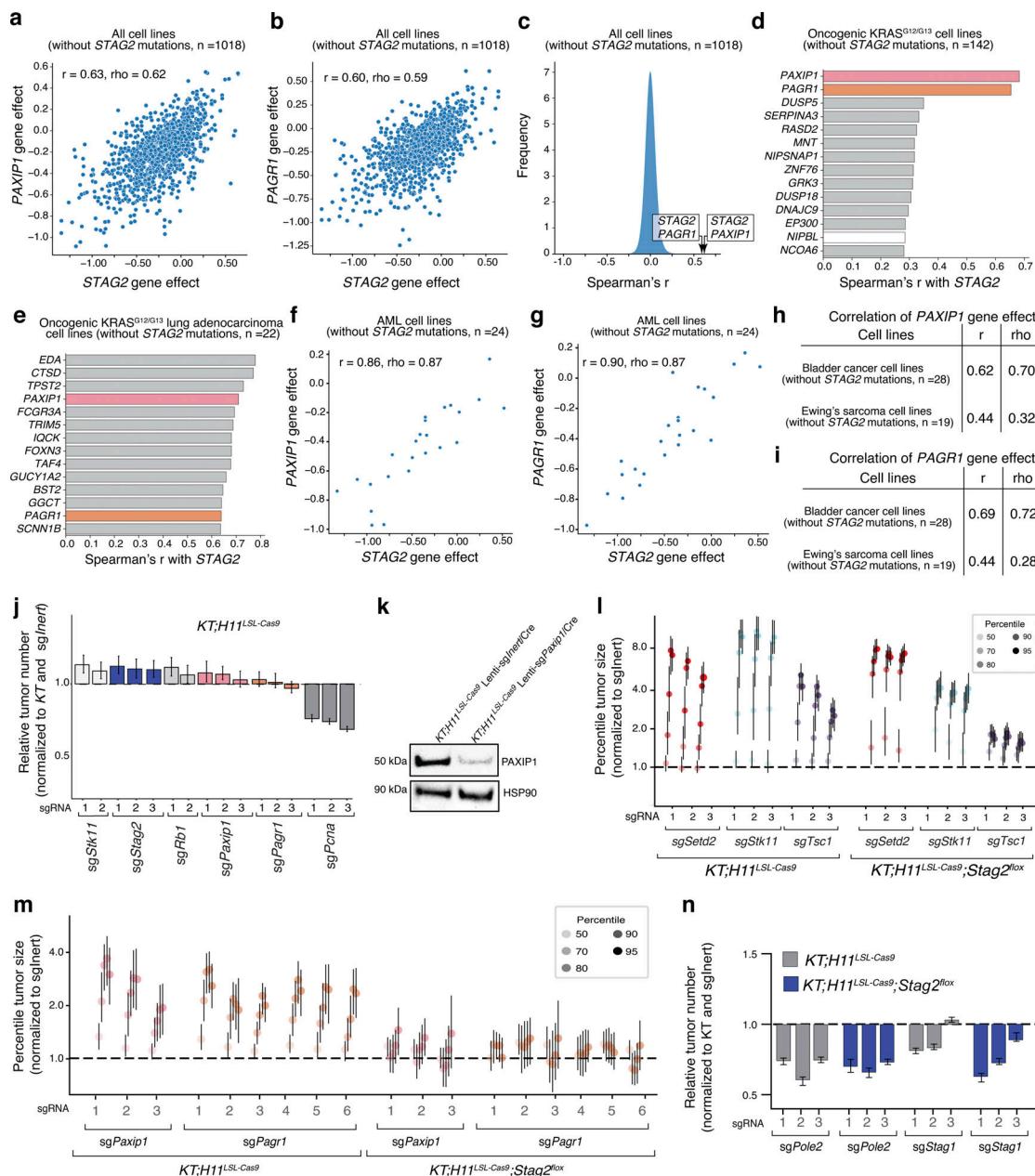


Figure S1. The effect of *STAG2* inactivation is highly correlated with that of *PAXIP1* and *PAGR1* inactivation in human cancer cell lines, and genetic interactions with *STAG2* are robust to different measures of tumorigenesis and tumor growth in oncogenic Kras lung cancer in vivo. (a and b) Gene knockout effects for *PAXIP1* and *STAG2* inactivation (a) and for *PAGR1* and *STAG2* inactivation (b). Each dot represents a cell line. Cell lines with *STAG2* mutation were excluded. Spearman's *r* and Pearson rho are indicated. (c) Bell curve showing the frequency of Spearman's correlations across all pairwise comparisons in DepMap. *STAG2-PAGR1* and *STAG2-PAXIP1* correlations are indicated. (d and e) Genes with the highest Spearman's correlation with the effect of *STAG2* inactivation from DepMap. Data from cell lines with oncogenic mutations at codons 12 or 13 of KRAS (d) and lung adenocarcinoma cell lines with oncogenic mutations at codons 12 or 13 of KRAS (e). Cell lines with *STAG2* mutation were excluded. *PAXIP1* and *PAGR1* are colored bars. Core cohesin complex genes are white bars. (f and g) Gene effects for *PAXIP1* and *STAG2* inactivation (f) and for *PAGR1* and *STAG2* inactivation (g) in AML. Each dot represents a cell line. Cell lines with *STAG2* mutations were excluded. Spearman's *r* and Pearson rho are indicated. (h and i) Tables indicating Spearman's *r* and Pearson rho for *PAXIP1* gene effect (h) and for *PAGR1* gene effect (i) in bladder cancer and Ewing's sarcoma. Cell lines with *STAG2* mutations were excluded. (j) Relative tumor number (normalized to *KT* and *sgInert*). Mean \pm 95% confidence intervals are shown. Dotted line indicates no effect. Raw values and significance of each effect are shown in Table S2 (j). (k) Western blot on sorted neoplastic cells from *KT;H11^{LSL-Cas9}* mice with tumors initiated with the indicated Lenti-sgRNA/Cre vectors. Data represents one replicate of three independent experiments. (l) Tumor sizes at the indicated percentiles for tumors with sgRNA targeting *Setd2*, *Stk11*, or *Tsc1* (normalized to *sgInert*) in *KT;H11^{LSL-Cas9}* and *KT;H11^{LSL-Cas9,Stag2^{flx}}* mice. Each gene was targeted with three sgRNAs. Error bars indicate 95% confidence intervals. Dotted line indicates no effect. (m) Tumor sizes at the indicated percentiles for tumors with sgRNA targeting *Paxip1* or *Pagr1* (normalized to *sgInert*) in *KT;H11^{LSL-Cas9}* and *KT;H11^{LSL-Cas9,Stag2^{flx}}* mice. Each gene was targeted with three and six sgRNAs, respectively. Error bars indicate 95% confidence intervals. Dotted line indicates no effect. (n) Comparison of relative tumor number for tumors with sgRNAs targeting for *Stag1*, *Stag2*, or *Pole2* in *KT;H11^{LSL-Cas9}* mice compared to *KT;H11^{LSL-Cas9,Stag2^{flx}}* mice. Mean \pm 95% confidence intervals are shown. Raw values and significance of each effect are shown in Table S4 (l-n). Source data are available for this figure: SourceData FS1.

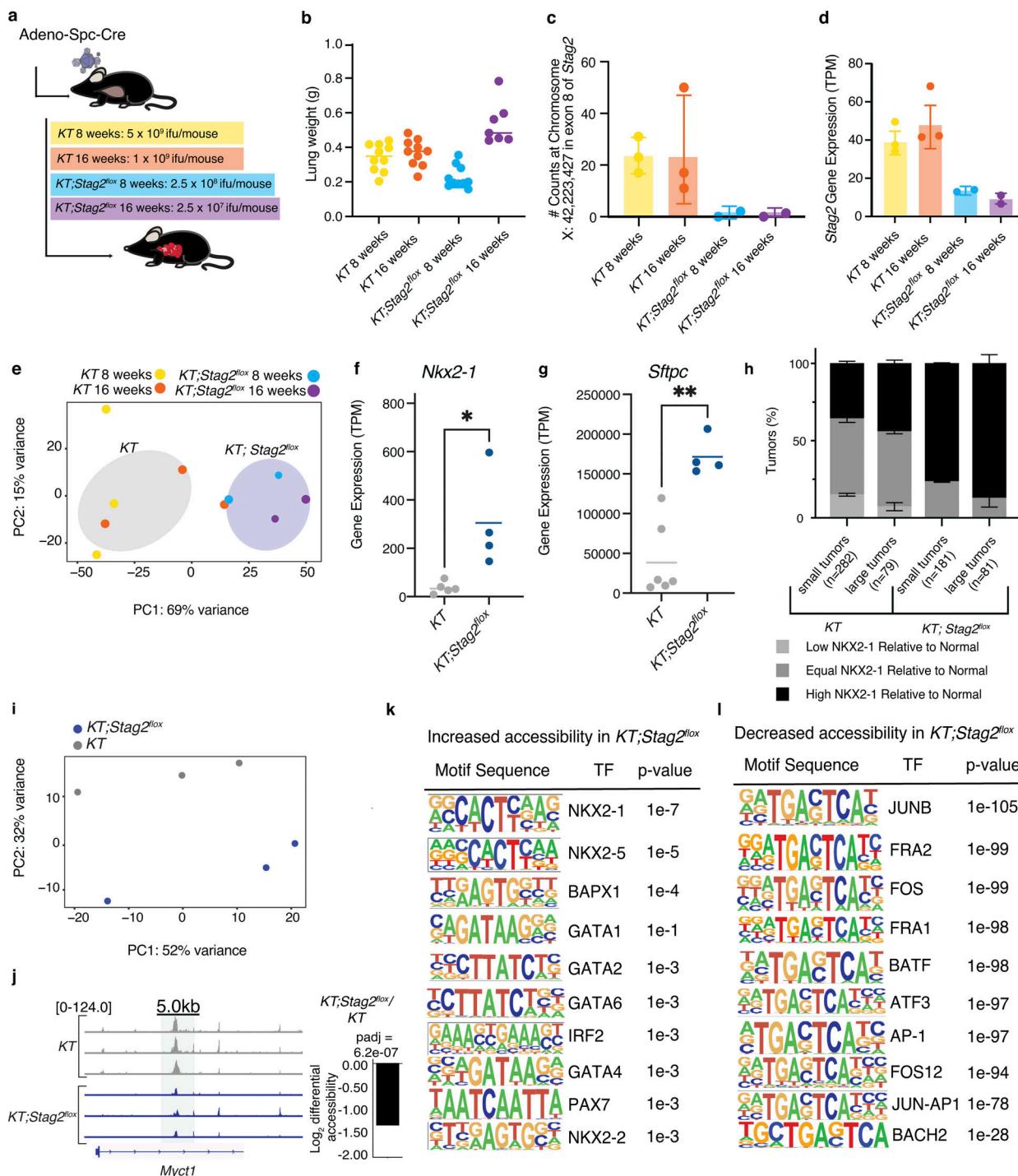


Figure S2. Validation of STAG2 inactivation, clustering of KT and *KT;Stag2^{flx}* samples, gene expression of surfactant genes, and differential accessibility regulated by STAG2. (a) Schematic of tumor initiation with Adeno-Spc-Cre with different viral titers in KT and *KT;Stag2^{flx}* mice to generate tumor samples to analyze at 8 and 16 wk after tumor initiation. (b) Lung weights of mice in each group. Each dot represents a mouse and the bar is the mean. (c) Number of DNA nucleotide counts at Chromosome X locus in Exon 8 of STAG2 in each group. Note that this region is within floxed exon. Mean \pm SEM is shown. (d) *Stag2* RNA expression in TPM via RNA-seq in each group. Bar is the standard error. Mean \pm SEM is shown. (e) PCA of *KT;Stag2^{flx}* and KT tumors processed with RNA-seq libraries at 8 and 16 wk clustered by genotype. (f and g) Gene expression (TPM) for *Nkx2-1* (f) and *Sftpc* (g) for KT and *KT;Stag2^{flx}* 8 and 16 wk. **P value <0.01 , *P value <0.1 by unpaired t test. Each dot represents a mouse, and the bar is the mean. (h) Quantification of Nkx2-1 expression in *KT;Stag2^{flx}* ($n = 2$ mice) and KT ($n = 2$ mice) tumors at 8 wk after initiation. Tumor Nkx2-1 expression quantified by comparison to Nkx2-1 expression in adjacent normal tissue. Tumor number quantified labeled in graph. Mean \pm SD is shown (P value <0.001 by ordinary one-way ANOVA). (i) PCA of KT and *KT;Stag2^{flx}* tumors processed with ATAC-seq libraries at 16 wk clustered by genotype. (j) ATAC-seq signal tracks for the *Myct1* gene locus in tumors from KT and *KT;Stag2^{flx}* mice. (k) Motif sequences and corresponding P values for transcription factors with motif enrichment in regions with increased accessibility in cancer cells from *KT;Stag2^{flx}* mice. (l) Motif sequences and corresponding P values for transcription factors with motif enrichment in regions with decreased accessibility in cancer cells from *KT;Stag2^{flx}* mice.

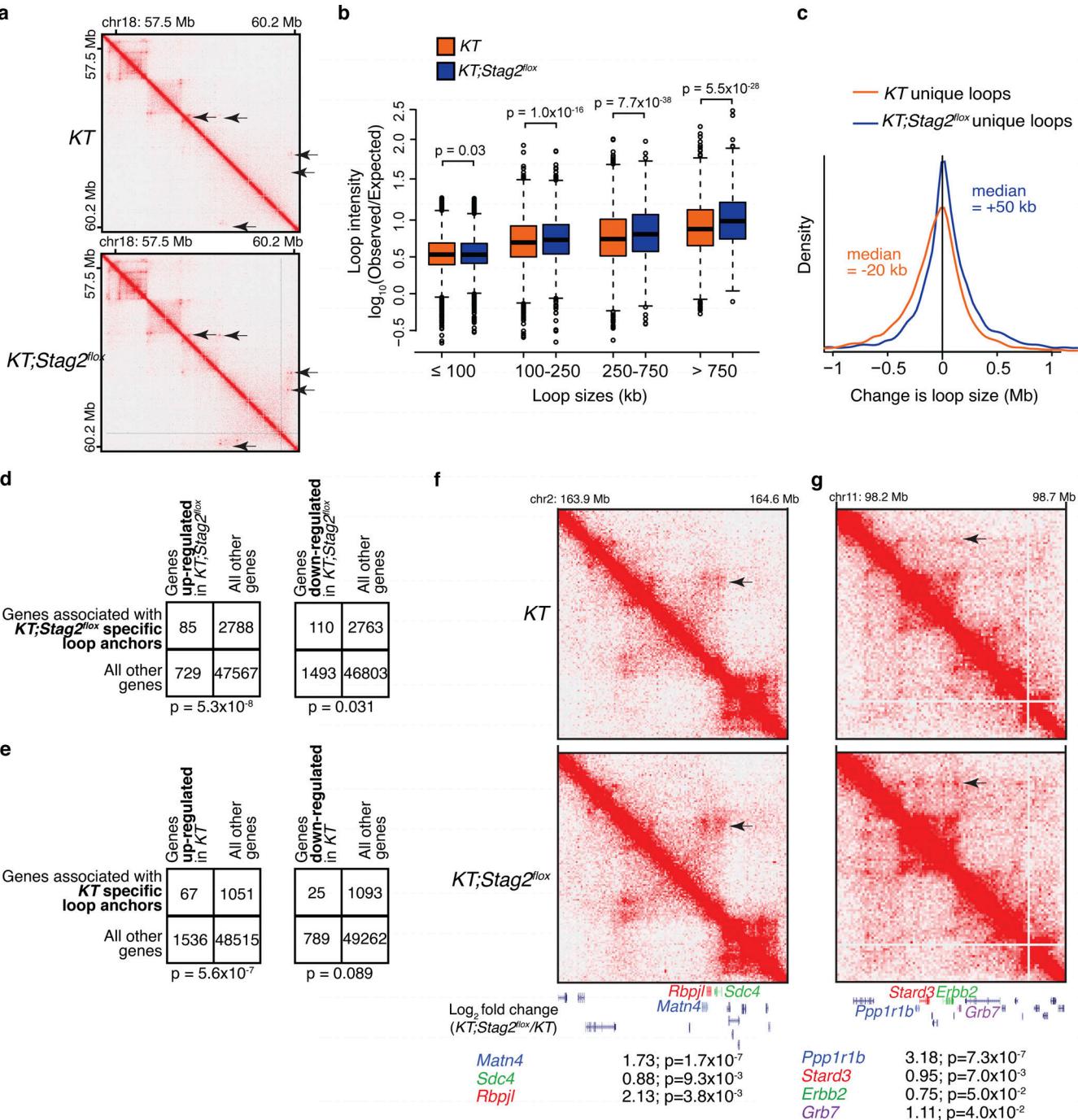


Figure S3. KT and KT;Stag2^{flx} samples contain both unique and common DNA loops and DNA anchors that correlate with gene expression but not chromatin accessibility. (a) Normalized contact matrices in KT (top) and KT;Stag2^{flx} (bottom) neoplastic cells. Contact matrices were visualized in Juicebox with the same color scale. Arrows indicate novel or stronger loop contacts in KT;Stag2^{flx}. (b) Comparison of loop intensity ($\log_{10}[\text{observed/expected}]$) of the indicated loop sizes between all loops in KT and KT;Stag2^{flx} samples. Boxes show median \pm interquartile range. Whiskers show standard error. P values (Wilcoxon rank test) are shown. (c) Change in loop size between the KT;Stag2^{flx} unique loops and the common loops or KT unique loops with which they share one anchor (blue). Change in loop size between KT unique loops and the common loops or KT;Stag2^{flx} unique loops with which they share one anchor (orange). (d) Number of genes associated with KT;Stag2^{flx} unique loops that are upregulated (left) or downregulated (right) in KT;Stag2^{flx} samples versus all other genes that are and are not associated with KT;Stag2^{flx} unique loops. P value (chi-square test) is shown. (e) Number of genes associated with KT unique loops that are upregulated (left) or down-regulated (right) in KT samples versus all other genes that are and are not associated with KT unique loops. P value (chi-square test) is shown. (f and g) Differences in chromatin looping and gene expression (from RNA-seq data, \log_2 fold change [$\text{KT;Stag2}^{\text{flx}}/\text{KT}$]; P value) of several differentially expressed genes proximal to unique anchor sites.

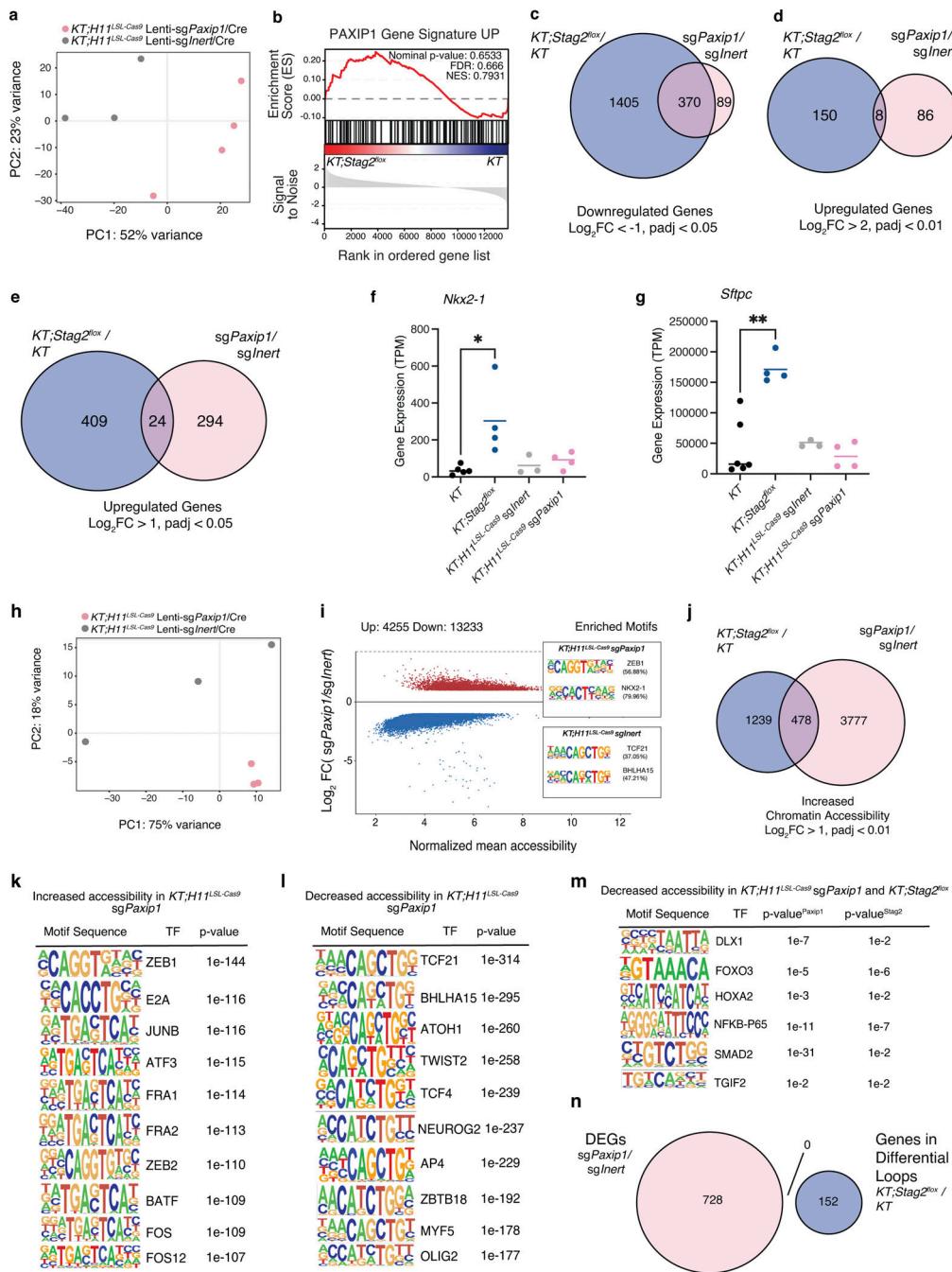


Figure S4. PAXIP1 and STAG2-cohesin regulation of gene expression and chromatin accessibility is conserved in downregulation but not upregulation of genes. (a) PCA of RNA-seq on *KT;H11^{LSL-Cas9} sgInert* and *KT;H11^{LSL-Cas9} sgPaxip1* tumors. (b) Upregulated PAXIP1 gene signature enrichment in rank-ordered gene list for *KT* versus *KT;Stag2^{flx}*. (c–e) Venn diagram of shared downregulated genes ($\log_2FC < -1$, $\text{Padj} < 0.05$) (c), shared upregulated genes ($\log_2FC > 2$, $\text{Padj} < 0.01$) (d), and shared upregulated genes ($\log_2FC > 1$, $\text{Padj} < 0.05$) (e) between *KT;Stag2^{flx}/KT* and *KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert*. (f and g) Gene expression (TPM) for *Nkx2-1* (f) and *Sftpc* (g) in neoplastic cells from tumors from *KT*, *KT;Stag2^{flx}*, *KT;H11^{LSL-Cas9} sgInert*, and *KT;H11^{LSL-Cas9} sgPaxip1* in mice. Each dot is an RNA-seq sample, and the bar is the mean. **P value <0.01, *P value <0.1 by unpaired t test (*KT* and *KT;Stag2^{flx}* data are same as Fig. S2). (h) PCA of *KT;H11^{LSL-Cas9} sgInert* and *KT;H11^{LSL-Cas9} sgPaxip1* tumors processed with ATAC-seq libraries. (i) Differential accessibility across 17,488 significant peaks in *KT;H11^{LSL-Cas9} sgInert* and *KT;H11^{LSL-Cas9} sgPaxip1* mice (three mice/group). The x-axis represents the \log_2 mean accessibility per peak and the y-axis represents the \log_2FC in accessibility. Colored dots are significant ($\log_2FC > 1$, $\text{FDR} < 0.05$). Red dots are increased chromatin accessibility in *KT;H11^{LSL-Cas9} sgPaxip1* accompanied by transcription factor hypergeometric motif enrichment in *KT;H11^{LSL-Cas9} sgPaxip1*, and blue dots are decreased chromatin accessibility in *KT;H11^{LSL-Cas9} sgPaxip1* accompanied by transcription factor hypergeometric motif enrichment in *KT;H11^{LSL-Cas9} sgInert*. (j) Venn diagram of shared regions of increased chromatin accessibility ($\log_2FC > 1$, $\text{Padj} < 0.01$) between *KT;Stag2^{flx}/KT* and *KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert*. (k and l) Motif sequences and corresponding P values for transcription factors with motif enrichment in regions with increased accessibility in *KT;H11^{LSL-Cas9} sgPaxip1* (k) and with decreased accessibility in *KT;H11^{LSL-Cas9} sgPaxip1* (l). (m) Motif sequences and corresponding P values for transcription factors with motif enrichment in regions with decreased accessibility in *KT;H11^{LSL-Cas9} sgPaxip1* and *KT;Stag2^{flx}*. (n) Venn diagram of differentially expressed genes regulated by *KT;H11^{LSL-Cas9} sgPaxip1/KT;H11^{LSL-Cas9} sgInert* ($|\log_2 FC| > 1$, $\text{Padj} < 0.05$; RNA-seq) versus genes found in differential loops regulated by *KT;Stag2^{flx}/KT* (HiC).

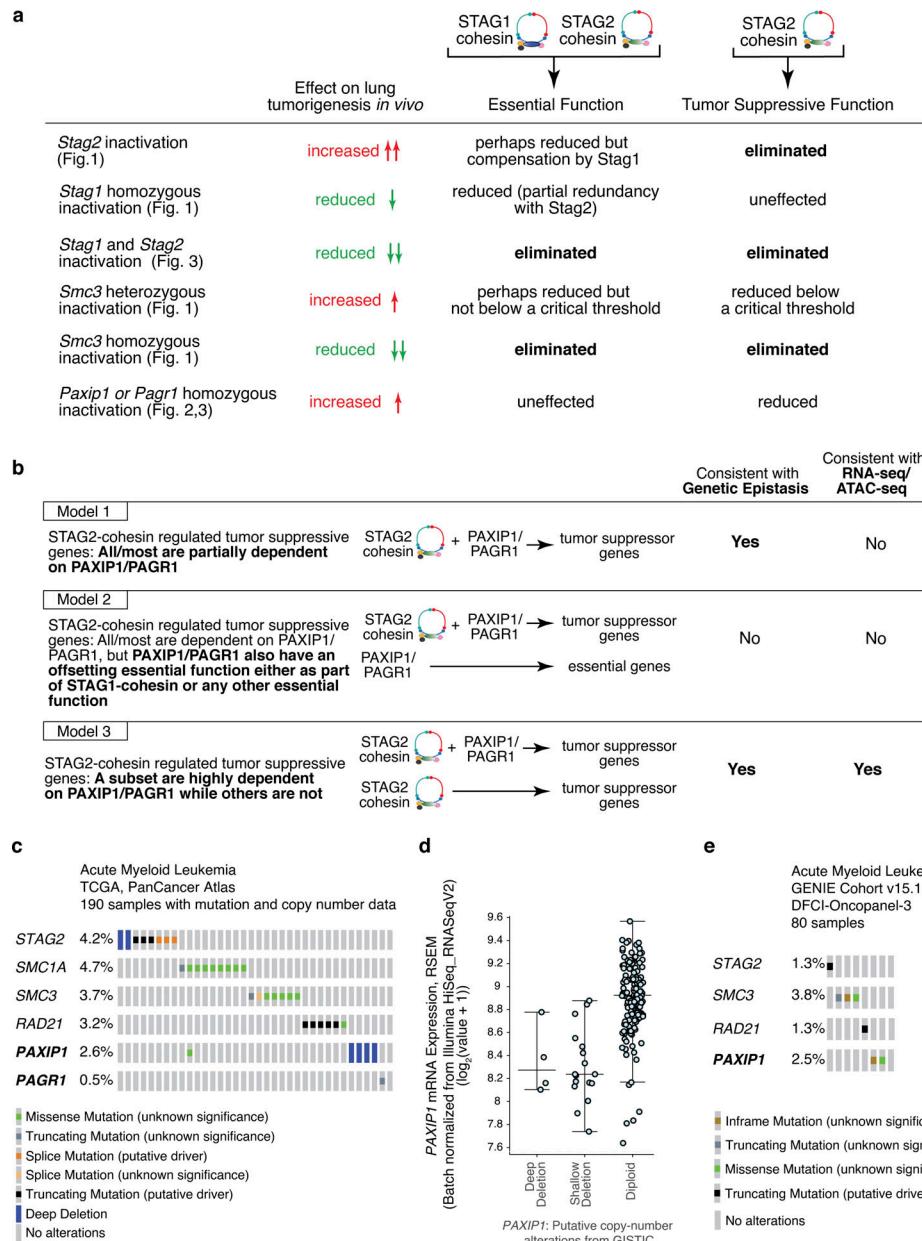


Figure S5. Summary of the *in vivo* genetic experiments as they related to the essential role of cohesin in general and the specific tumor suppressive function of STAG2-cohesin and PAXIP1/PAGR1, as well as a summary of the most parsimonious model of STAG2-cohesin and PAXIP1/PAGR1 mediated tumor suppression and mutations in PAXIP1 and PAGR1 and downregulation of PAXIP1 mRNA expression in a subset of AMLs. (a) Interpretations of results from Figs. 1, 2, and 3 describing the distinct phenotypes controlled by STAG2- and STAG1-cohesin, homozygous and heterozygous inactivation of cohesin components, and the role of PAXIP1 on lung tumorigenesis *in vivo*. The lung tumor suppressive effect of *Stag2* is well established in oncogenic KRAS-driven lung tumors (Cai et al., 2021; Blair et al., 2023), confirmed in the current study, and extended to lung cancer driven by other oncogenes (Blair et al., 2023). The compensation of STAG1 and STAG2 in the essential functions of cohesin, in general, is well established in cell lines (Arruda et al., 2020; van der Lelij et al., 2017; Canudas and Smith, 2009) and confirmed in lung cancer by our *in vivo* studies, including the genetic epistasis between *Stag1* and *Stag2*. While homozygous inactivation of each core and auxiliary cohesin component greatly reduced lung tumorigenesis, heterozygous inactivation of *Smc3* using a floxed allele increases tumorigenesis. This likely explains the mutations in cohesin components in human cancer and extends the importance of dysregulation of this complex to a much larger fraction of lung adenocarcinomas. Finally, while *Paxip1* or *Pagr1* inactivation increased lung tumorigenesis, inactivation of either gene did not reduce tumorigenesis of *Stag2*-deficient tumors, as would have been expected if *Paxip1/Pagr1* were also involved in the essential cohesin function. (b) Multiple models could have explained how STAG2-cohesin and the PAXIP1/PAGR1 complex cooperate to suppress lung tumorigenesis. However, our genetic epistasis data (Fig. 3) and molecular analyses (Figs. 4 and 5) are most consistent with Model 3 in which the major role of PAXIP1/PAGR1 is to work with STAG2-cohesin to regulate a subset of genes that are controlled by STAG2-cohesin. (c) Oncoprint of AMLs from TCGA accessed through cBioPortal. 190 samples with mutation and copy number data. Mutation type is indicated. (d) mRNA expression of *PAXIP1* in AMLs from TCGA accessed through cBioPortal. 165 samples with mutation, copy number, and gene expression data. Sample were split based on putative copy number of *PAXIP1*. Each dot is a sample. Note low expression in a subset of sample with likely unaltered DNA copy number (diploid samples). (e) Oncoprint of AMLs from DCFI-Oncopanel-3 samples from GENIE accessed through cBioPortal. 80 samples with mutation data. *SMC1A* and *PAGR1* were not profiled. Mutation type is indicated.

Provided online are Table S1, Table S2, Table S3, Table S4, Table S5, Table S6, Table S7, Table S8, Table S9, Table S10, and Table S11. Table S1 shows tumor size and number metrics from the cohesin complex member Tuba-seq^{Ultra} analysis. Table S2 shows tumor size and number metrics from *Paxip1*, *Pagr1*, *Cdca5*, and various TSGs and essential gene Tuba-seq^{Ultra} analysis. Table S3 shows genes selected for Tuba-seq^{Ultra} analysis of putative genetic interactors of STAG2. Table S4 shows tumor size and number metrics from Tuba-seq^{Ultra} analysis of putative genetic interactors of STAG2. Table S5 shows metrics and RLE-normalized (DESeq2) counts for *KT;Stag2^{flx}* versus *KT* 8 and 16 wk after tumor initiation RNA-seq experiment. Table S6 shows metrics and RLE-normalized (DESeq2) counts for *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* RNA-seq experiment. Table S7 shows metrics for differentially expressed genes altered by both STAG2 and PAXIP1 with \log_2 fold change $> (abs) 1$. Table S8 shows metrics and RLE-normalized (DESeq2) counts for *KT;Stag2^{flx}* versus *KT* 16 wk after tumor initiation ATAC-seq experiment. Table S9 shows metrics and RLE-normalized (DESeq2) counts for *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* 16 wk after tumor initiation ATAC-seq experiment. Table S10 shows a list of forward and reverse primers used to generate ATAC-seq libraries for *KT;Stag2^{flx}* versus *KT* and *KT;H11^{LSL-Cas9} sgPaxip1* versus *KT;H11^{LSL-Cas9} sgInert* ATAC-seq experiments. Table S11 shows loci for all loops and normalized metrics for *KT;Stag2^{flx}* versus *KT* 16 wk after tumor initiation Hi-C experiment.