

1    **A STAG2-PAXIP1/PAGR1 axis suppresses lung tumorigenesis**

2

3    Emily L. Ashkin<sup>1</sup>, Yuning J. Tang<sup>2\*</sup>, Haiqing Xu<sup>3\*</sup>, King L. Hung<sup>1,5</sup>, Julia Belk<sup>5,6</sup>, Hongchen Cai<sup>2</sup>,  
4    Steven Lopez<sup>2</sup>, Deniz Nesli Dolcen<sup>2</sup>, Jess D. Hebert<sup>2</sup>, Rui Li<sup>4,5</sup>, Paloma A. Ruiz<sup>2</sup>, Tula Keal<sup>4</sup>,  
5    Laura Andrejka<sup>2</sup>, Howard Y. Chang<sup>5,6</sup>, Dmitri A. Petrov<sup>1,3,7</sup>, Jesse R. Dixon<sup>4</sup>, Zhichao Xu<sup>4</sup>, and  
6    Monte M. Winslow<sup>1,2,8,#</sup>

7

8    1. Cancer Biology Program, Stanford University School of Medicine, Stanford, CA, USA 94305

9    2. Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA 94305

10    3. Department of Biology, Stanford University, Stanford, CA, USA 94305

11    4. Gene Expression Laboratory; Salk Institute for Biological Studies, La Jolla, CA, USA 92037

12    5. Center for Personal Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA 94305

13    6. Howard Hughes Medical Institute, Stanford University School of Medicine, Stanford, CA, USA 94305

14    7. Chan Zuckerberg Biohub, San Francisco, CA, USA 94158

15    8. Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA 94305

16    \*These authors contributed equally

17    #Corresponding Author: Monte M. Winslow: [mwinslow@stanford.edu](mailto:mwinslow@stanford.edu)

18

19

20

21

22

23

24

25

26

27

28 **SUMMARY**

29 STAG2 is a frequently mutated cohesin subunit across several cancers and one of the most  
30 important functional suppressors of lung adenocarcinoma. Our findings underscore important roles  
31 of STAG2 in suppressing lung tumorigenesis and highlight a STAG2-PAXIP1/PAGR1 tumor-  
32 suppressive program that may transcend cancer type.

33

34

35 **ABSTRACT**

36

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

50

51 **INTRODUCTION**

52       The cohesin complex is a large multi-subunit complex that plays important roles in  
53 regulating genome conformation, gene expression and determining cell fate (1). There are two  
54 classes of cohesin complexes, distinguished by whether they contain the paralogs STAG1 or  
55 STAG2 (2). STAG1- and STAG2-cohesin are thought to control different gene expression  
56 programs (3–5). STAG1-cohesin colocalizes with CTCF to form topologically associating domain  
57 boundaries that drive long-range chromatin interactions for gene transcription (5,6). Conversely,  
58 STAG2-cohesin has been shown to drive mid-range chromatin interactions (5,7); however, it  
59 remains unknown whether additional proteins contribute specifically to STAG2-cohesin regulated  
60 gene expression (8,9).

61       Notably, *STAG2* is the most frequently mutated cohesin complex member across many  
62 cancer types, including bladder cancer, Ewing's sarcoma, acute myeloid leukemia (AML), and  
63 lung adenocarcinoma (9–18). Using *in vivo* CRISPR/Cas9 screens within genetically engineered  
64 mouse models of lung cancer, we and others found *STAG2* to be an important tumor suppressor  
65 in lung adenocarcinoma (19,20). *STAG2* mutations occur in ~4% of human lung adenocarcinomas,  
66 and *STAG2* protein expression is low or absent in ~20% of these tumors (19). In lung cancer,  
67 *STAG2*-mediated tumor suppression is not through an impact on chromosomal instability or  
68 hindering MAPK signaling, as previously suggested for other cancer types (10,19,21–25). Thus,  
69 the mechanisms through which *STAG2* inactivation drives lung tumor growth remain largely  
70 unknown and crucial for understanding cellular processes that contribute to tumor suppression.

71       Here, using CRISPR/Cas9-mediated somatic genome editing within genetically engineered  
72 mouse models of lung adenocarcinoma, we investigate the tumor-suppressive capacity of *STAG2*,  
73 every other component of the cohesin complex, and potential co-operators of *STAG2*-cohesin in

74 lung cancer. We identify PAXIP1 and PAGR1, which form an obligatory heterodimer that has  
75 been suggested to be involved in DNA damage repair, cell cycle control, and transcription  
76 regulation (26–29) as potent lung tumor suppressors, and provide multiple complementary lines  
77 of evidence for a STAG2-PAXIP1/PAGR1 tumor-suppressive axis.

78

## 79 RESULTS

80

### 81 Lung tumor suppression by STAG2 is unique among cohesin subunits

82 STAG2 is a tumor suppressor in oncogenic KRAS-driven lung cancer, and other cohesin  
83 subunits are frequently mutated in KRAS-driven lung cancer as well as in several other cancer  
84 types (9–14,19,30–33). To determine whether the inactivation of other cohesin subunits would  
85 also increase autochthonous lung tumor growth, we used tumor barcoding and somatic  
86 CRISPR/Cas9-mediated genome editing (19,34). We used a recently optimized version of tumor  
87 barcoding coupled with high-throughput barcode sequencing (Tuba-seq), in which lentiviral  
88 vectors have a diverse barcode (BC) integrated within the U6 promoter directly 5' of the sgRNA  
89 (Tuba-seq<sup>Ultra</sup>; U6 barcode Labeling with per-Tumor Resolution Analysis; **Fig. 1a-b and Methods**)  
90 (35). Tuba-seq<sup>Ultra</sup> enables the quantification of the size of each clonal tumor and the number of  
91 tumors with each sgRNA based on amplification of the BC-sgRNA region from bulk tumor-  
92 bearing lungs followed by high-throughput sequencing.

93 We generated a lentiviral pool that contained barcoded lentiviral-sgRNA/Cre Tuba-seq<sup>Ultra</sup>  
94 vectors targeting each cohesin gene, an essential gene, and Safe-cutting “inert” sgRNAs (Lenti-  
95 U6<sup>BC</sup>sgCohesin/Cre with 3 sgRNAs/gene; **Fig. 1b, Methods and Supplementary Table S1**). We  
96 initiated tumors with Lenti-U6<sup>BC</sup>sgCohesin/Cre in *Kras*<sup>LSL-G12D/+</sup>,*R26*<sup>LSL-Tomato</sup>,*H11*<sup>LSL-</sup>

97 *Cas9* (*KT;H1I<sup>LSL-Cas9</sup>*) mice and Cas9-negative *Kras<sup>LSL-G12D/+</sup>;R26<sup>LSL-Tomato</sup>* (*KT*) mice (**Methods**).  
98 After fifteen weeks of tumor growth, we extracted DNA from bulk tumor-bearing lungs, PCR  
99 amplified the BC-sgRNA region, and high-throughput sequenced the amplicon. By tallying the  
100 number of reads from each BC-sgRNA and normalizing to “spike-in” control cells with a known  
101 BC-sgRNA that were added at a defined cell number, we quantified the number of neoplastic cells  
102 in each clonal tumor with each sgRNA (**Methods**).

103 We calculated tumor number (the number of BCs associated with each sgRNA normalized  
104 to the expected number from Cas9-negative *KT* mice) and several metrics of tumor size (log-  
105 normal mean tumor size and tumor sizes at defined percentiles within the tumor size distribution;  
106 **Methods**). Consistent with previous observations, *Stag2* inactivation greatly increased tumor size  
107 and modestly increased tumor number (**Fig. 1c-d**) (19,20). Inactivation of most of the other cohesin  
108 complex components greatly reduced tumor number and tumor size, although not to the same  
109 extent as inactivation of the known essential gene *Pcna* (**Fig. 1c-e**). All three sgRNAs targeting  
110 *Stag1* only modestly reduced tumor number and tumor size. Collectively, these data are consistent  
111 with cohesin in general having an essential function and the well-described ability of STAG2 can  
112 compensate for STAG1 in this essential function (3,6,21,36) (**Fig. 1c-e**). Importantly, these results  
113 provided no indication that STAG1 is a tumor suppressor and are consistent with distinct roles of  
114 STAG1- and STAG2-cohesin in lung cancer.

115

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

117 While homozygous inactivation of cohesin subunits is detrimental to lung tumorigenesis,  
118 mutations in these genes are common in lung cancer, and heterozygous inactivation of core cohesin  
119 genes increase tumorigenesis in other contexts (37,38). Thus, we hypothesized that heterozygous

120 inactivation of core cohesin subunits could increase lung tumor growth. We chose to focus on the  
121 core cohesion component Smc3 due to the availability of mice with an *Smc3* floxed allele (39). To  
122 generate tumors with inactivation on one allele of *Smc3*, we initiated lung tumors in *KT;Smc3<sup>flx/+</sup>*  
123 and *KT* control mice. After fourteen weeks of tumor growth, *KT;Smc3<sup>flx/+</sup>* mice had significantly  
124 greater tumor burden and higher tumor number compared to *KT* mice as assessed by total lung  
125 weight, fluorescent imaging, and histology (**Fig. 1f-i**). These genetic data support a model in which  
126 the net effect of a reduction in cohesin is increased lung tumor growth potentially driven by trade-  
127 offs between a reduction in STAG2 cohesin-mediated tumor suppression and the essential function  
128 of cohesin in general.

129

130 **DepMap and *in vivo* analyses reveal cooperation between *STAG2* and the *PAXIP1/PAGR1*  
131 complex in KRAS-driven lung tumors**

132 To understand mechanistically how STAG2-cohesin regulates tumorigenesis, we next  
133 sought to find genes that cooperate with STAG2 to mediate these effects. We analyzed the  
134 CRISPR/Cas9 human cancer cell line Dependency Map data for genes whose effects of cell growth  
135 (referred to as their gene effect) were most similar to that of *STAG2* (**Methods**) (40,41). We  
136 initially assessed the impact of inactivating each gene across all cell lines (excluding those with  
137 *STAG2* mutations, N=1018) and calculated the correlation (Pearson's r) with the *STAG2* gene  
138 effect. Consistent with previous analyses, *PAXIP1* and its obligate heterodimer partner *PAGR1*  
139 had the most highly correlated gene effects with *STAG2* (r=0.62 and 0.61, respectively) (8,42) and  
140 their correlation with *STAG2* was ~2-fold higher than any other gene, including core and auxiliary  
141 cohesin subunits (**Fig. 2a** and **Supplementary Fig. S1a-c**). Importantly, in oncogenic  
142 KRAS<sup>G12/G13</sup>-driven cell lines (N=120), as well as in oncogenic KRAS<sup>G12/G13</sup>-driven lung

143 adenocarcinoma cell lines (N=22), the gene effects of *PAXIP1* and *PAGR1* robustly correlated  
144 with the gene effect of *STAG2* (**Fig. 2b-c** and **Supplementary Fig. S1d-e**). Finally, the positive  
145 correlations between *STAG2* and *PAXIP1/PAGR1* gene effects extended to other cancer types  
146 where *STAG2* is frequently mutated, including acute myeloid leukemia, bladder cancer, and  
147 Ewing's sarcoma (**Supplementary Fig. S3f-i**) (9–14).

148 To determine whether *PAXIP1* and *PAGR1* are tumor suppressors in lung cancer, we used  
149 Tuba-seq<sup>Ultra</sup> to quantify the impact of inactivating *STAG2*, *PAXIP1*, or *PAGR1* on oncogenic  
150 KRAS-driven lung tumorigenesis (**Fig. 2d**). We initiated tumors *KT;H11<sup>LSL-Cas9</sup>* and *KT* mice with  
151 a pool of Lenti-U6<sup>BC</sup>-sgRNA/Cre vectors that contained sgRNAs targeting each gene of interest  
152 (**Fig. 2d** and **Supplementary Table S2**). After fifteen weeks of tumor growth, we quantified the  
153 number of neoplastic cells in each tumor using Tuba-seq<sup>Ultra</sup> (**Fig. 2d**). Inactivation of *Paxip1* and  
154 *Pagr1* significantly increased tumor size, on par with the effect of inactivating the canonical tumor  
155 suppressor *Rb1* (**Fig. 2e**). Inactivation of *Paxip1* and *Pagr1* showed modest effects on tumor  
156 number, on par with inactivation of *Stk11*, *Stag2*, and *Rb1* (**Supplementary Fig. S1j**). The tumor-  
157 suppressive effects of *PAXIP1* and *PAGR1* were less than that of *STAG2* (**Fig. 2e-f**). No vectors  
158 impacted tumorigenesis in the absence of Cas9 (**Supplementary Table S2**).

159 To validate the tumor-suppressive effects of *PAXIP1*, we initiated lung tumors with Lenti-  
160 sg*Inert*/Cre and Lenti-sg*Paxip1*/Cre in separate cohorts of *KT;H11<sup>LSL-Cas9</sup>* and *KT* mice (**Fig. 2g**).  
161 *PAXIP1* inactivation was confirmed by western blotting on sorted neoplastic cells  
162 (**Supplementary Fig. S1k**). Overall tumor burden as assessed by lung weight was significantly  
163 higher in *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1* mice compared to each of the three control cohorts (mice with  
164 *KT;H11<sup>LSL-Cas9</sup>* sg*Inert*, *KT* sg*Paxip1*, and *KT* sg*Inert* tumors) (**Fig. 2h-i**). Collectively, these

165 quantitative *in vivo* data and DepMap findings from human cancer cell lines show that STAG2-  
166 cohesin effects correlate with PAXIP1/PAGR1 effects in lung tumorigenesis (**Fig. 2a-f**).

167

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

170 To determine whether PAXIP1 and PAGR1 function in the same pathway as STAG2, as  
171 well as extend our understanding of how STAG2-deficiency changes the impact of coincident  
172 genetic alterations on KRAS-driven lung tumorigenesis, we quantify the impact of inactivating a  
173 curated list of genes on tumor initiation and growth in *Stag2*-proficient and *Stag2*-deficient tumors  
174 using Tuba-seq<sup>Ultra</sup> (**Fig. 3a-b**). We generated a pool of Lenti-U6<sup>BC</sup>-sgRNA/Cre vectors that  
175 targeted ~160 genes, including those that correlate with *STAG2* effects in DepMap (e.g. *Paxip1*  
176 and *Pagr1*), several canonical tumor suppressor genes, genes with protein-protein interactions with  
177 STAG2, and genes downstream of STAG2 based on preliminary molecular analyses (**Fig. 3a-b**,  
178 **Supplementary Table S3**, and **Methods**). Each gene was targeted with 3-6 sgRNAs, and this pool  
179 contained control Safe-cutting “inert” sgRNAs and sgRNAs targeting several control essential  
180 genes (**Fig. 3a,b**). We initiated lung tumors in *KT*, *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>*  
181 mice (**Fig. 3b**). After thirteen weeks of tumor growth, we used Tuba-seq<sup>Ultra</sup> to quantify the number  
182 of neoplastic cells in each tumor with each sgRNA (**Fig. 3b**).

183 *Stag2* deficiency in *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice significantly changes the impact of  
184 inactivating only a small number of genes, including several canonical tumor suppressor genes  
185 and genes identified through DepMap (**Fig. 3c**). As expected, sgRNAs targeting *Stag2* increased  
186 tumor growth *KT;H11<sup>LSL-Cas9</sup>* mice but not in *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice (**Fig. 3d**). Inactivation  
187 of *Setd2* increased the growth of *Stag2*-deficient tumors more than *Stag2*-proficient tumors, while

188 inactivation of *Stk11* or *Tsc1* increased the growth of *Stag2*-deficient tumors less than *Stag2*-  
189 proficient tumors (**Fig. 3e** and **Supplementary Fig. S1l**). Thus, *Stag2* inactivation modifies the  
190 overall tumor suppressor landscape in lung cancer.

191 Based on our analysis of DepMap data and the tumor-suppressive effects of STAG2,  
192 PAXIP1, and PAGR1 *in vivo*, we hypothesized that they may cooperate to suppress lung  
193 tumorigenesis. Using Tuba-seq<sup>Ultra</sup>, we found that inactivating *Paxip1* or *Parg1* increased the  
194 growth of *Stag2*-proficient tumors but had no impact on the growth of *Stag2*-deficient tumors (**Fig.**  
195 **3f** and **Supplementary Fig. S1m**). This epistasis suggests that PAXIP1/PAGR1 and STAG2-  
196 cohesin suppress lung tumorigenesis through partially overlapping the mechanisms.

197 Finally, *Stag1* was the only cohesin complex component for which inactivation had a  
198 significantly different effect in *Stag2*-proficient versus *Stag2*-deficient lung tumors (**Fig. 3c,g**).  
199 *Stag1* inactivation was significantly more deleterious for *Stag2*-deficient lung tumors with effects  
200 on tumor number approaching that of inactivating the known essential gene *Pole* (**Fig. 3g** and  
201 **Supplementary Fig. S1n**). These results are consistent with the well-established synthetic  
202 lethality of *Stag1* and *Stag2* and the essential role of cohesin in general in this tumor type (4,43).  
203 Collectively, these quantitative analyses demonstrate that STAG2 inactivation impacts the overall  
204 tumor suppression landscape of lung cancer, is synthetic lethal with STAG1, and is epistatic to the  
205 PAXIP1/PAGR1 complex.

206

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

209 To characterize the molecular effects of STAG2-deficiency, we initiated tumors in *KT* and  
210 *KT;Stag2*<sup>fl/fl</sup> mice and used FACS to isolate neoplastic cells (DAPI<sup>neg</sup>, lineage<sup>neg</sup>, Tomato<sup>pos</sup> cells)

211 for molecular analyses (**Fig. 4a**) (44). To capture any changes in the molecular output of STAG2  
212 deficiency during tumor development, we collected samples after 8 and 16 weeks of tumor growth.  
213 We initiated tumors using different viral titers such that mice would have similar overall tumor  
214 burden at time of analysis (**Fig. 4a** and **Supplementary Fig. S2a-b**).

215 *Stag2* inactivation in these samples was validated by mapping reads to the floxed exon of  
216 *Stag2* and measuring *Stag2* expression (**Supplementary Fig. S2c-d**). Based on hierarchical  
217 clustering and principle component analysis (PCA), samples from *KT* and *KT;Stag2<sup>flx</sup>* mice  
218 formed distinct clusters (**Fig. 4b** and **Supplementary Fig. S2e**). Comparisons between neoplastic  
219 cells from all *KT* and *KT;Stag2<sup>flx</sup>* samples uncovered >2500 differentially expressed genes  
220 (absolute value of log<sub>2</sub> fold change (|log<sub>2</sub>FC|) > 1, FDR <0.01; **Fig. 4b** and **Supplementary Table**  
221 **S5**). Using Gene set enrichment analysis (GSEA) and Gene Ontology (GO) term enrichment, we  
222 uncovered upregulation of pathways related to surfactant protein metabolism (exemplified by  
223 significantly increased expression of the transcription factor *Nkx2-1* and canonical NKX2-1-  
224 regulated surfactant protein genes), lung development, and several metabolic processes in *Stag2*-  
225 deficient neoplastic cells (**Fig. 4c,e** and **Supplementary Fig. S2f-g**). Consistent with this gene  
226 expression data, immunohistochemical staining for NKX2-1 showed that *Stag2*-deficient tumors  
227 were significantly more likely to have high expression for NKX2-1 than *Stag2*-proficient tumors  
228 (p < 0.001; **Supplementary Fig. S2h**). Significantly down regulated gene sets in *Stag2*-deficient  
229 neoplastic cells were related to epithelial response to immune cells and extracellular matrix  
230 organization (**Fig. 4d,f**). Across an independent dataset of lung tumorigenesis in oncogenic KRAS-  
231 driven models, *Nkx2-1* expression positively correlated with expression of genes that were  
232 upregulated by STAG2-deficiency ( $R^2 = 0.7162$ , p < 0.0001) and negatively correlated with

233 expression of genes that were downregulated by STAG2 inactivation ( $R^2 = 0.4589$ ,  $p < 0.0001$ )  
234 consistent with NKX2-1 driving some of these differences (**Fig. 4g-h**) (45).

235 To assess whether Stag2 deficiency also impacts the underlying chromatin state, we  
236 performed ATAC-seq on neoplastic cells isolated from *KT* and *KT;Stag2<sup>flx</sup>* mice. Neoplastic cells  
237 from *KT* and *KT;Stag2<sup>flx</sup>* mice clustered separately and had comparable library quality (**Fig. 4i**,  
238 **Supplementary Fig. S2i** and **Methods**). Stag2-deficient neoplastic cells have several thousand  
239 regions of increased and decreased chromatin accessibility relative to Stag2-proficient neoplastic  
240 cells ( $|\log_2\text{FC}| > 1$ , FDR  $< 0.05$ ) (**Fig. 4j**, **Supplementary Fig. S2j**, and **Supplementary Table**  
241 **S9**). Interestingly, regions with increased accessibility in *Stag2*-deficient neoplastic cells were  
242 enriched for NKX2-1 and GATA family transcription factor motifs, which are well-established  
243 regulators of lung-specific lineage programs and lung development (**Fig. 4j** and **Supplementary**  
244 **Fig. S2k**) (45–48). Additionally, there were regions with increased accessibility proximal to lung  
245 lineage genes, such as *Sftpd* (**Fig. 4k**). Regions with decreased accessibility in *Stag2*-deficient  
246 neoplastic cells were enriched for JUNB, AP-1, FOS, BATF, and FRA1/2 motifs (**Fig. 4j** and  
247 **Supplementary Fig. S2l**). Notably, the gene expression and chromatin accessibility changes in  
248 neoplastic cells from *KT* and *KT;Stag2<sup>flx</sup>* mice were positively correlated ( $R = 0.311$ ,  $p < 2.2 \times 10^{-16}$ )  
249 (**Fig. 4l**). Collectively, these data show that STAG2 inactivation rewires the epigenome and  
250 changes gene expression states.

251  
252 **STAG2 inactivation results in differential DNA looping that correlates with increased gene**  
253 **expression**

254 Cohesin regulates 3D genome structure which can facilitate enhancer-promoter  
255 interactions and impact gene expression (2,49). Thus, to determine whether STAG2 inactivation

256 impacts chromatin looping, we performed chromosome conformation capture (Hi-C) on neoplastic  
257 cells isolated from *KT* and *KT;Stag2<sup>flx</sup>* mice (**Fig. 5a** and **Supplementary Fig. 3a**). *Stag2*-  
258 deficient neoplastic cells gained and lost DNA loops compared to *Stag2*-proficient cells.  
259 Specifically, in addition to > 7,500 common loops there were ~3,000 loops that were lost/reduced  
260 in *Stag2*-deficient cells (*KT* unique loops) and >5,700 loops that were gained/stronger in *Stag2*-  
261 deficient cells (*KT;Stag2<sup>flx</sup>* unique loops; **Fig. 5b-c**). Furthermore, *KT;Stag2<sup>flx</sup>* unique loops were  
262 on average larger ( $p = 5.5 \times 10^{-28}$ ) than *KT* unique loops (**Fig. 5d** and **Supplementary Fig. S3b**).  
263 We identified unique and shared loop “anchors” between *KT* and *KT;Stag2<sup>flx</sup>* samples. Notably,  
264 many of the unique loops share one anchor site with loops in the corresponding *KT* or *KT;Stag2<sup>flx</sup>*  
265 samples (**Fig. 5e**). On average the loop anchors were more distal in *KT;Stag2<sup>flx</sup>* samples compared  
266 with *KT* suggesting loop extrusion may favor different anchor sites in the absence of STAG2  
267 (**Supplementary Fig. S3c**).

268 To better understand whether the differences in 3D chromatin looping contribute to the  
269 molecular changes caused by *Stag2* inactivation, we assessed whether the anchor sites of unique  
270 loops are associated with transcription start sites (TSSs) of differentially expressed genes. By  
271 overlapping differentially expressed genes with the coordinates of the anchor sites of unique loops,  
272 we found that genes upregulated in *KT;Stag2<sup>flx</sup>* samples are enriched proximal to the anchor sites  
273 of *KT;Stag2<sup>flx</sup>* unique loops (*i.e.* the gene TSS is within the anchor site) (85 genes,  $p = 5.3 \times 10^{-8}$ )  
274 (**Supplementary Fig. S3d**). Similarly, genes with higher expression in *KT* samples were enriched  
275 proximal to anchor sites of *KT* unique loops (67 genes,  $p = 5.6 \times 10^{-7}$ ) (**Supplementary Fig. S3e**  
276 and **Supplementary Table 11**). For differentially expressed genes that are associated with unique  
277 anchors sites in either *KT* or *KT;Stag2<sup>flx</sup>* samples, we analyzed whether the corresponding anchor  
278 (*i.e.* the other side of the unique loop) had changes in chromatin accessibility, which might occur

279 if the generation of novel enhancers led to the formation of the differential chromatin interactions.  
280 Interestingly, >75% of the corresponding anchor sites did not significantly change in chromatin  
281 accessibility. This suggests that the changes in 3D chromatin looping are the result of STAG2  
282 inactivation not because of enhancer re-programming. Additionally, many of the genes with  
283 increased expression and proximity to unique anchor sites in *Stag2*-deficient neoplastic cells, such  
284 as EREG and FGFR, are associated with tumor growth pathways (**Fig. 5f**). We also found two loci  
285 associated with unique loops in *Stag2*-deficient cells, that have multiple genes that are all  
286 upregulated, possible due to the different loop extrusion in the absence of STAG2  
287 (**Supplementary Fig. S3f-g**). These regions include genes implicated in lung lineage development  
288 and lung adenocarcinoma and progression (50–52). Altogether, this genome conformation data  
289 strongly suggests that STAG2 inactivation leads to changes in 3D chromatin looping and activation  
290 of downstream pathways that could promote tumor growth and alter differentiation.

291

## 292 **PAXIP1 and STAG2-cohesin control conserved gene expression programs**

293 As our genetic epistasis data suggest that the tumor-suppressive function of PAXIP1 and  
294 STAG2-cohesin are related, we next determined the extent to which the molecular programs driven  
295 by PAXIP1 and STAG2-cohesin overlap. To characterize the molecular effects of PAXIP1  
296 deficiency in lung tumors, neoplastic cells were FACS-isolated from tumors initiated in  
297 *KT;H11<sup>LSL-Cas9</sup>* mice with Lenti-sg*Inert*/Cre or Lenti-sg*Paxip1*/Cre (**Fig. 6a**). We performed bulk  
298 RNA-seq on 3-4 samples from each group (hereafter sg*Inert* and sg*Paxip1*). Samples cluster based  
299 on their genotype (**Fig. 6b** and **Supplementary Fig. S4a**). Many (~540) genes were differentially  
300 expressed between sg*Paxip1* and sg*Inert* neoplastic cells ( $|\log_2\text{FC}| > 1$  and  $\text{FDR} < 0.01$ ) (**Fig. 6b**  
301 and **Supplementary Table S6**).

302            Interestingly, genes that were regulated by both STAG2 and PAXIP1 were almost  
303 exclusively downregulated in *Stag2*-deficient and *Paxip1*-deficient neoplastic cells (**Fig. 6c**).  
304 Genes that were downregulated in *sgPaxip1* neoplastic cells (“PAXIP1 DOWN”), were  
305 downregulated in *KT;Stag2<sup>flx</sup>* neoplastic cells (p-value < 0.0036; **Fig. 6d**). However, genes that  
306 were downregulated by *Stag2* deficiency were either equally dependent on PAXIP1 or not  
307 dependent on PAXIP1. Pathways enriched in the genes downregulated in both *KT;Stag2<sup>flx</sup>* and  
308 *KT;H11<sup>LSL-Cas9</sup>* *sgPaxip1* tumors include those related to ion transport, metabolic processes, and  
309 angiogenesis (**Fig. 6e-f**). Genes that were upregulated in *KT;H11<sup>LSL-Cas9</sup>* *sgPaxip1* neoplastic cells  
310 were not upregulated in *KT;Stag2<sup>flx</sup>* neoplastic cells and vice versa (**Supplementary Fig. S4b-g**).

311            Given the broad changes in chromatin accessibility induced by *Stag2* deficiency, we  
312 performed ATAC-seq on *sgInert* and *sgPaxip1* neoplastic cells (**Fig. 6a**). These samples were of  
313 similar quality and clustered by genotype (**Supplementary Fig. S4h**). There were >4200 regions  
314 with decreased chromatin accessibility and >13000 regions with increased chromatin accessibility  
315 in neoplastic cells from *sgInert* versus *sgPaxip1* mice ( $|\log_2\text{FC}| > 1$ , FDR < 0.05) (**Supplementary**  
316 **Fig. S4i-j** and **Supplementary Table S9**). Motifs for NKX2-1, ZEB1, and GATA were enriched  
317 in regions with increased accessibility in *sgPaxip1* neoplastic cells, and motifs for TCF21 and  
318 BHLHA15 were enriched in regions with decreased accessibility (**Supplementary Fig. S4k-l**).  
319 Motifs for DLX1, SMAD2, and HOXA2 were enriched in regions with decreased accessibility  
320 that overlap between *sgPaxip1* neoplastic cells and STAG2-deficient cells (**Supplementary Fig.**  
321 **S4m**). Although there was motif enrichment for NKX2-1 in regions with increased accessibility  
322 in *sgPaxip1* neoplastic cells, neither *Nkx2-1* nor *Nkx2-1*-regulated genes were upregulated in  
323 *sgPaxip1* neoplastic cells suggesting that regulation of cell differentiation through NKX2-1 is a  
324 STAG2-specific mechanism (**Supplementary Fig. S4f-g** and **Supplementary Fig. S2f-g**).

325           Similar to gene expression, there were more shared regions with decreased rather than  
326 increased chromatin accessibility (**Fig. 6g, Supplementary Fig. S4j and Supplementary Table**  
327 **S9**). Interestingly, many downregulated genes conserved between *KT;Stag2<sup>flx</sup>* and *sgPaxip1* had  
328 decreased accessibility in Stag2-deficient and Paxip1-deficient neoplastic cells (**Fig. 6h-i** and  
329 **Supplementary Table S7**). Lastly, there was no overlap between genes regulated by PAXIP and  
330 genes regulated by STAG2-dependent chromatin looping (**Supplementary Fig. 4n**). Collectively,  
331 these results emphasize that there are distinct PAXIP1-dependent and PAXIP1-independent  
332 STAG2-mediated tumor suppressive programs, consistent with the overall tumor-suppressive  
333 effects of PAXIP1, PAGR1 and STAG2 *in vivo* (**Fig. 6j** and **Supplementary Fig. S5**).  
334

## 335 **DISCUSSION**

336

337           Here, we coupled multiplexed and quantitative functional genomics within autochthonous  
338 lung cancer models, *in vivo* genetic epistasis experiments, and molecular analyses to uncover  
339 mechanisms of STAG2-mediated tumor suppression and link it to PAXIP1/PARG1. Tumor  
340 suppression in lung cancer was specific to STAG2-cohesin rather than STAG1-cohesin or overall  
341 cohesin. Cohesin components beyond *STAG2* are also frequently mutated in human lung cancer,  
342 and our results show that heterozygous inactivation of a core cohesin gene increased tumor growth.  
343 We show that SMC3 exhibits both essential and haploinsufficient tumor suppressor gene  
344 characteristics, further emphasizing the versatility and power of *STAG2*-cohesin as a tumor  
345 suppressor. Thus, heterozygous or hypomorphic mutations in core or auxiliary cohesin  
346 components likely reduce STAG2-cohesin-mediated tumor suppression and make the fraction of

347 human lung tumors that are driven by this mechanism greater than only those with *STAG2*  
348 mutations (**Fig. 1** and **Supplementary Fig. S2**).

349 Using cancer cell line dependency data, multiple studies have revealed a correlation  
350 between the dependency scores for *STAG2* and the *PAXIP1-PAGR1* complex (8,42), which we  
351 show extends to cell lines driven by oncogenic KRAS as well as to the subset of those cell lines  
352 derived from lung adenocarcinoma (**Fig. 2b-c**). Previous studies using cancer cell lines have shown  
353 that inactivation of *STAG2* and *PAXIP1* reduces growth, leading to the suggestion that these genes  
354 promote cancer cell growth (42,53). However, our functional data show that *STAG2*, *PAXIP1*,  
355 and *PAGR1* normally constrain lung tumor growth *in vivo*, consistent with many studies that  
356 document *STAG2* as a tumor suppressor (9,11–14,19,30). Thus, while cell lines may correctly  
357 uncover correlated gene effects, their optimized growth in culture and/or lack of physiologic  
358 context limit their ability to identify functional tumor suppressors.

359 To identify genes that cooperate with *STAG2* to suppress tumor growth *in vivo*, we  
360 compared the tumor-suppressive effects of a broad panel of genes in *STAG2*-proficient and -  
361 deficient autochthonous lung tumors (**Fig. 3**). These data uncovered relatively few genes whose  
362 effects were impacted by *Stag2*-deficiency, emphasizing the uniqueness of the interaction between  
363 *STAG2* and *PAXIP1/PAGR1* (**Fig. 3c**). In this experiment, we did not address the role of  
364 inflammation in KRAS-driven tumor progression, which could serve as a critical dimension for  
365 future studies (54,55). *PAXIP1* and *PAGR1* have been implicated in several biological processes,  
366 including DNA damage responses (26,27,29,56). Thus, we initially considered whether the  
367 reduced impact of *Paxip1* and *Pagr1* inactivation on tumor growth relative to *Stag2* inactivation  
368 could be due to partially offsetting tumor-suppressive and essential DNA damage functions (**Fig.**  
369 **3f** and **Supplementary Fig. S5a-b**). However, *Paxip1* or *Pagr1* inactivation in the context of

370 *Stag2*-deficiency (*i.e.*, in  $KT;H11^{LSL-Cas9};Stag2^{flox}$  mice) did not reduce tumorigenesis, as would be  
371 expected if these two genes had a second essential function (**Fig. 3f** and **Supplementary Fig. S5a-**  
372 **b**). Furthermore, PAXIP1/PAGR1 are required for the expression of a subset of STAG2-regulated  
373 genes, some of which are likely responsible for PAXIP1/PAGR1-mediated tumor suppression  
374 (**Fig. 5** and **Supplementary Fig. S5a-b**).

375 STAG2-mediated tumor suppression is also mediated by PAXIP1/PAGR1-independent  
376 gene expression programs (**Fig. 5** and **Supplementary Fig. S3**). Interestingly, in contrast to many  
377 genetic alterations that increase tumorigenesis while decreasing or changing differentiation,  
378 STAG2-deficient tumors express higher levels of the lung lineage defining transcription factor  
379 NKX2-1 as well as canonical genes related to ATII differentiation (**Fig. 4g-h, j-k** and  
380 **Supplementary Fig. S2-3**). Additionally, there is increased DNA loop formation in STAG2-  
381 deficient tumors at loci that overlap with genes involved in lung differentiation and tumorigenesis,  
382 such as genes related to ATII differentiation and the EGF pathway (**Fig. 6d** and **Supplementary**  
383 **Fig. S4e-g**). Thus, it is interesting to note the potential stabilization of this cellular state by STAG2  
384 deficiency, which otherwise might be anticipated to be selected against (57). Genomic regions  
385 with increased accessibility in STAG2-deficient tumors are also enriched for NKX2-1 binding  
386 sites (**Fig. 4k**). Whether NKX2-1 interacts directly or indirectly with STAG2-cohesin-  
387 PAXIP1/PAGR1 and whether an NKX2-1-regulated oncogenic programs drive growth in this  
388 context will be important areas of future investigation (58).

389 While we show clear evidence for a genetic interaction between STAG2-cohesin and the  
390 PAXIP1/PAGR1 complex during lung tumor suppression this data does not distinguish whether  
391 PAXIP/PAGR1 functions upstream or downstream to STAG2-cohesin (**Fig. 3f**). Our data and  
392 existing literature on PAXIP1/PAGR1 and STAG2-cohesin are most consistent with a model in

393 which PAXIP1/PAGR1 is initially recruited to genome loci (perhaps through direct interactions  
394 with lineage-specific transcription factors), followed by STAG2-cohesin localization at those  
395 regions, and the formation of genomic contacts that ultimately impact gene expression to suppress  
396 lung tumor growth (42,59). A direct interaction between PAXIP1/PAGR1 and STAG2-cohesin is  
397 only weakly supported by existing mass spectrometry data on STAG2-, cohesin-, and PAXIP1-  
398 interacting proteins (25,60). PAXIP1/PAGR1 and STAG2-cohesin could interact weakly or  
399 transiently, through an intermediate protein, or be functionally linked through a less direct  
400 mechanism such as PAXIP1/PAGR1-mediated recruitment of histone modifying enzymes  
401 followed by preferential localization of STAG2-cohesin at these regions (28,61). Interestingly,  
402 while our Hi-C data uncovered unique loops that overlap with genes impacted by STAG2  
403 inactivation, these unique loops did not overlap with genes impacted by both STAG2 inactivation  
404 and PAXIP1 inactivation (**Supplementary Fig. S4n** and **Supplementary Table 7, 11**). Further  
405 biochemical, genetic, and genomic experiments including CTCF ChIA-PET and CUT&RUN  
406 should further clarify the interaction between STAG2-cohesin and the PAXIP1/PAGR1 complex  
407 as well as their genomic localization during tumor suppression.

408         Interestingly, the tumor-suppressive STAG2-PAXIP1/PAGR1 axis may transcend cancer  
409 type. In human cell lines, *STAG2*, *PAXIP1*, and *PAGR1* have highly correlated effects in bladder  
410 cancer, Ewing's sarcoma, acute myeloid leukemia (AML), and lung adenocarcinoma, all of which  
411 also have frequent mutations in *STAG2* (**Supplementary Fig. S1f-i**) (31–33). In particular, the  
412 correlation of *STAG2* with *PAXIP1* and *PAGR1* in AML was strikingly high ( $r = 0.86$  and  $r=0.90$ ,  
413 respectively) (**Supplementary Fig. S1f-g**). In AML, *STAG2* mutations and heterozygous  
414 mutations in *SMC3* are well-established regulators of the pre-leukemic state (6). Interestingly,  
415 cancer sequencing studies have also identified mutations in *PAXIP1* and *PAGR1* in AML and

416 down regulation of *PAXIP1* expression in AML (**Supplementary Fig. S5c-e**) (31–33,62). Thus,  
417 the PAXIP1/PAGR1 complex may represent an important unexplored driver in AML.

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

422

423 **ACKNOWLEDGEMENTS**

424 We thank the Stanford Shared FACS Facility for flow cytometry and cell sorting services,  
425 the Stanford Veterinary Animal Care Staff for expert animal care, Human Pathology/Histology  
426 Service Center, and Stanford Protein and Nucleic Acids Facility. We would like to acknowledge  
427 the American Association for Cancer Research and its financial and material support in the  
428 development of the AACR Project GENIE registry, as well as members of the consortium for their  
429 commitment to data sharing. Interpretations are the responsibility of study authors. We thank D.  
430 Solomon, D. Hargreaves, J. Sage, L. Attardi, P. Khavari, J. Carette, P. Du, B. Chick, N. Avina  
431 Ochoa, and members of the Winslow and Crabtree laboratories for helpful comments and  
432 experimental support. We thank A. Melnick for providing *Smc3<sup>flaxed</sup>* mice. E.L.A. was supported  
433 by the HHMI Gilliam Fellowship for Advanced Study (GT14928) and a National Cancer Institute  
434 Predoctoral to Postdoctoral Fellow Transition Award (F99CA284289). Y.J.T. was supported by  
435 the Canadian Institute of Health Research (CIHR) postdoctoral fellowship (MFE-176568). K.L.H.  
436 was supported by a Stanford Graduate Fellowship and a National Cancer Institute Predoctoral to  
437 Postdoctoral Fellow Transition Award (F99CA274692). J.B. was supported by the HHMI Hanna  
438 H. Gray Fellowship Award. H.C. was supported by a Tobacco-Related Disease Research Program

439 (TRDRP) fellowship (28FT-0019). D.N.D. was supported by the National Cancer Institute  
440 Predoctoral to Postdoctoral Fellow Transition Award (K00-CA245784). J.D.H was supported by  
441 an American Cancer Society postdoctoral fellowship (PF-21-112-01-MM) and a Tobacco-Related  
442 Disease Research Program (TRDRP) fellowship (T31FT1619). P.A.R was supported by the  
443 National Science Foundation Graduate Research Fellowship Program (DGE-2146755) and the  
444 Stanford Graduate Fellowship. This work was supported by NIH R01-CA231253 (to M.M.W) and  
445 R01-CA234349 (to M.M.W and D.A.P) and in part by the Stanford Cancer institute support grant  
446 (NIH P30-CA124435).

447

## 448 **CONTRIBUTIONS**

449 E.L.A. and M.M.W. conceptualized and designed the study. Y.J.T designed the Tuba-  
450 seq<sup>Ultra</sup> methodology. E.L.A., Y.J.T., and L.A. prepared Tuba-seq libraries. E.L.A., H.X., J.B.,  
451 Z.X., and K.L.H. designed the computational pipelines, methodology, and performed formal  
452 analysis. E.L.A., M.M.W., Z.X., and H.X. wrote the manuscript with comments from all authors.  
453 E.L.A., J.D.H., S.L., D.N.D., Z.X., J.R.D., M.M.W. and K.L.H. edited the manuscript. E.L.A.,  
454 Y.J.T., H.C., S.L., D.N.D., P.R., R.L., and L.A. acquired data. M.M.W., H.Y.C., J.R.D., and  
455 D.A.P. provided resources.

456

## 457 **DECLARATION OF INTERESTS**

458 M.M.W and D.A.P are founders and hold equity in Guide Oncology. H.Y.C. is a co-  
459 founder of Accent Therapeutics, Boundless Bio, Cartography Biosciences, 428 Orbital  
460 Therapeutics, and an advisor of 10x Genomics, Arsenal Biosciences, Chroma 429 Medicine, and  
461 Spring Discovery.

462 **Figure 1. Inactivation of *Stag2* and heterozygous inactivation of the cohesin subunit *Smc3***  
463 **uniquely increase lung tumor growth**

464 **a)** Schematic of the cohesin complex with subunits labeled.

465 **b)** Tumor initiation with a pool of Lenti-U6<sup>BC</sup>sgRNA/Cre vectors. Genotype and number of mice  
466 are indicated. Tuba-seq<sup>Ultra</sup> was performed on each tumor-bearing lung 15 weeks after tumor  
467 initiation, followed by analyses to quantify tumorigenesis.

468 **c)** Relative mean tumor size (normalized to sgInert). Mean +/- 95% confidence intervals are  
469 shown. Dotted line indicates no effect.

470 **d)** Relative tumor number (normalized to *KT* and sgInert). Mean +/- 95% confidence intervals are  
471 shown. Dotted line indicates no effect.

472 **e)** Relative tumor number (normalized to *KT* and sgInert) correlated with relative mean tumor size  
473 (normalized to sgInert). Data represents one replicate of two independent experiments.

474 **f)** Tumor initiation with Adeno-Spc-Cre in *KT* and *KT;Smc3*<sup>flx/+</sup> mice. Mouse number is indicated.

475 **g)** Representative fluorescence and histology images of lung lobes from the indicated genotypes  
476 of mice. Top scale bars, 5 mm, and bottom scale bars, 1 mm.

477 h) Lung weights from *KT* and *KT;Smc3*<sup>flx/+</sup> mice. Each dot represents a mouse and the bar is the  
478 mean. Data are representative of two independent experiments.

479 **i)** Percentage of Tomato-positive tumor area detected via histology. Each dot represents a mouse  
480 and the bar is the mean. \*p-value < 0.05, \*\* p-value < 0.01, via unpaired t-test.

481 Raw values and significance of each effect is shown in Supplementary Table 1.

482

483 **Figure 2. STAG2, PAXIP1, and PAGR1 effects are highly correlated in human cancer cell**  
484 **lines and PAXIP1/PAGR1 are tumor suppressive in KRAS-driven lung tumors *in vivo***

485 **a)** Genes with the most correlated effects with *STAG2* inactivation from the Dependency Map  
486 (DepMap). Cell lines with *STAG2* mutations were excluded. *PAXIP1* and *PAGR1* are colored bars.  
487 Core cohesin complex genes are white bars.

488 **b-c)** Gene effects for *PAXIP1* and *STAG2* inactivation (**b**) and for *PAGR1* and *STAG2*  
489 inactivation (**c**). Each dot represents a cell line with an oncogenic mutation at codons 12 or 13 of  
490 KRAS. Cell lines with *STAG2* mutations were excluded. Lung adenocarcinoma (LUAD) cell lines  
491 are shown as black dots. Spearman's r and Pearson rho for all cell lines and for LUAD cell lines  
492 are indicated.

493 **d)** Tumor initiation with a pool of barcoded Lenti-U6<sup>BC</sup>sgRNA/Cre Tuba-seq<sup>Ultra</sup> vectors.  
494 Genotype and number of mice are indicated. Tuba-seq<sup>Ultra</sup> was performed on each tumor-bearing  
495 lung 15 weeks after tumor initiation, followed by analyses to quantify tumorigenesis.

496 **e)** Relative mean tumor size (normalized to sgInert). Mean +/- 95% confidence intervals are  
497 shown. Dotted line indicates no effect.

498 **f)** Tumor sizes at the indicated percentiles for tumors with sgRNA targeting *Stag2*, *Paxip1*, or  
499 *Pagr1* (normalized to sgInert) in *KT;H11<sup>LSL-Cas9</sup>* mice. Each gene was targeted with three

500 sgRNAs. Error bars indicate 95% confidence intervals. Dotted line indicates no effect. Data  
501 represents one replicate of two independent experiments.

502 **g)** Tumor initiation with the indicated Lenti-sgRNA/Cre vectors in separate cohorts of *KT* and  
503 *KT;H11<sup>LSL-Cas9</sup>* mice. Number of mice in each group is indicated. Tumor burden was quantified 15  
504 weeks after tumor initiation.

505 **h)** Representative brightfield and fluorescence images of lung lobes and histology from the  
506 indicated groups of mice. Top scale bars and middle scale bars, 5 mm. Lower scale bars, 500  $\mu$ m.

507 **i)** Lung weights of mice in each group. Each dot represents a mouse and the bar is the mean. \*\*\*p-  
508 value <0.0001 by unpaired t-test. Data represents one replicate of two independent experiments.

509 Raw values and significance of each effect is shown in Supplementary Table 2.

510

511

512 **Figure 3. Genetic interactions with STAG2 include modification of overall tumor  
513 suppression and functional dependency with PAXIP1 and PAGR1**

514 **a)** List of candidate gene criteria for pool of 468 barcoded Lenti-sgRNA/Cre vectors.

515 **b)** Tumor initiation with Lenti-U6<sup>BC</sup>-sgRNA/Cre vectors in *KT;H11<sup>LSL-Cas9</sup>*, *KT;H11<sup>LSL-</sup>*  
516 *Cas9;Stag2<sup>fl/fl</sup>*, and *KT* mice. Number of mice in each group is indicated. Sequencing was performed  
517 on each tumor-bearing lung 14 weeks after tumor initiation, followed by analysis to quantify  
518 tumorigenesis.

519    **c)** Bar plot with percent of genes in each candidate category with differential effects between  
520     $KT;H11^{LSL-Cas9}$  and  $KT;H11^{LSL-Cas9};Stag2^{flox}$ .

521    **d)** Relative mean tumor size of tumors with sgRNAs targeting *Stag2* (normalized to *sgInert*) in  
522     $KT;H11^{LSL-Cas9}$  mice compared to  $KT;H11^{LSL-Cas9};Stag2^{flox}$  mice. Mean +/- 95% confidence  
523    intervals are shown. Dotted line indicates no effect.

524    **e)** Relative mean tumor size of tumors with sgRNAs targeting *Setd2*, *Stk11*, or *Tsc1* (normalized  
525    to *sgInert*) in  $KT;H11^{LSL-Cas9}$  mice compared to  $KT;H11^{LSL-Cas9};Stag2^{flox}$  mice. Mean +/- 95%  
526    confidence intervals are shown. Dotted line indicates no effect.

527    **f)** Relative mean tumor size of tumors with sgRNAs targeting *Paxip1* or *Pagr1* (normalized to  
528    *sgInert*) in  $KT;H11^{LSL-Cas9}$  mice compared to  $KT;H11^{LSL-Cas9};Stag2^{flox}$  mice. Mean +/- 95%  
529    confidence intervals are shown. Dotted line indicates no effect.

530    **g)** Comparison of average relative tumor number for sgRNAs targeting *Stag1*, *Stag2*, *Pole2*  
531    (essential gene), or *Inerts* in  $KT;H11^{LSL-Cas9}$  mice and  $KT;H11^{LSL-Cas9};Stag2^{flox}$  mice. Mean +/- 95%  
532    confidence intervals are shown.

533    Raw values and significance of each effect is shown in Supplementary Table 4. Data for  
534     $KT;H11^{LSL-Cas9}$  and  $KT$  mice represent one replicate of two independent experiments.

535

536    **Figure 4. STAG2 inactivation increases tumor-related metabolic processes and cell  
537    differentiation**

538   **a)** Schematic of tumor initiation in *KT* and *KT;Stag2<sup>flx</sup>* mice. Mice grew tumors until 8 weeks to  
539   represent large atypical adenomatous hyperplasia and small adenomas and 16 weeks to represent  
540   larger solid adenomas and early adenocarcinomas (Marjanovic *et al.*, 2020). Outline of tumor cell  
541   sorting and sample preparation for bulk RNA-seq and bulk ATAC-seq.

542   **b)** Upregulated and downregulated genes (n = 2545 genes) in *KT* relative to *KT;Stag2<sup>flx</sup>* tumors  
543   (absolute value of log<sub>2</sub> fold change (|log<sub>2</sub>FC|) > 1, FDR < 0.01).

544   **c-d)** Gene Set Enrichment Analysis (GSEA) pathways enriched in *KT;Stag2<sup>flx</sup>* relative to *KT*  
545   tumors (**c**) and enriched in *KT* relative to *KT;Stag2<sup>flx</sup>* tumors (**d**).

546   **e-f)** GO Term Gene Count Analysis with ClusterProfiler and EMBL-EBI GO:Term Category  
547   Analysis established pathways enriched from upregulated genes for *KT;Stag2<sup>flx</sup>/KT* mice (**e**) and  
548   from down-regulated genes for *KT;Stag2<sup>flx</sup>/KT* mice (**f**).

549   **g-h)** ES versus NKK2-1 gene expression (TPM) from Single Sample Gene Set Enrichment  
550   Analysis (ssGSEA) for *KPT* and *KT* samples (Chuang *et al.*, 2017) for gene set from genes  
551   upregulated in *KT;Stag2<sup>flx</sup>* versus *KT* (**g**) and genes downregulated in *KT;Stag2<sup>flx</sup>* versus *KT* (**h**).

552   **i)** Rank correlation of chromatin accessibility across *KT* and *KT;Stag2<sup>flx</sup>* samples. Samples cluster  
553   into two distinct groups.

554   **j)** Differential accessibility across 3576 significant peaks in *KT* and *KT;Stag2<sup>flx</sup>* mice (3  
555   mice/group). The x-axis represents the log<sub>2</sub> mean accessibility per peak and the y-axis represents  
556   the log<sub>2</sub> fold change in accessibility. Colored points are significant (|log<sub>2</sub>FC| > 1, FDR < 0.05). Red  
557   points are increased chromatin accessibility in *KT;Stag2<sup>flx</sup>* accompanied by transcription factor

558 hypergeometric motif enrichment in *KT;Stag2<sup>flx</sup>*, and blue points are decreased chromatin  
559 accessibility in *KT;Stag2<sup>flx</sup>* accompanied by transcription factor hypergeometric motif enrichment  
560 in *KT* tumors.

561 **k)** ATAC-seq signal tracks for the *Sftpd* gene locus in tumor from *KT* and *KT;Stag2<sup>flx</sup>* mice.

562 **l)** Comparison of log<sub>2</sub> fold change for *KT* and *KT;Stag2<sup>flx</sup>* mice for top genes from both bulk RNA-  
563 seq and bulk ATAC-seq. r = 0.311, p < 2.2 x 10-16.

564

565 **Figure 5. Stag2 deficiency impacts overall chromatin looping with effects on gene expression**

566 **a)** Schematic of tumor initiation with Adeno-Spc-Cre in *KT* and *KT;Stag2<sup>flx</sup>* mice. Outline of  
567 tumor cell sorting, crosslinking, and sample preparation for Hi-C.

568 **b)** Venn diagram of loops in *KT* and *KT;Stag2<sup>flx</sup>* samples.

569 **c)** Venn diagram of loop anchors in *KT* and *KT;Stag2<sup>flx</sup>* samples.

570 **d)** Sizes for *KT* unique loops, common loops, and *KT;Stag2<sup>flx</sup>* unique loops. Boxes show median  
571 +/- interquartile range. Whiskers show standard error. P-values (Wilcoxon rank test) are shown.

572 **e)** Loops in *KT* and *KT;Stag2<sup>flx</sup>* samples were compared and sorted into indicated categories based  
573 on their unique and shared loop anchors.

574 **f)** Differences in chromatin looping and gene expression (from RNA-seq data, log<sub>2</sub> fold change  
575 (*KT;Stag2<sup>flx</sup>/KT*); p-value) for *Ereg* and *Fgfr2*.

576

577 **Figure 6. PAXIP1 and STAG2-cohesin mechanisms of tumor suppression are conserved**

578   **a)** Schematic of tumor initiation with Lenti-sg*Paxip1*/Cre or Lenti-sg*Inert*/Cre in *KT;H11<sup>LSL-Cas9</sup>*  
579   mice (3-4 mice/group). Mice grew tumors for 15 weeks. Outline of tumor cell sorting and sample  
580   preparation for bulk RNA-seq and bulk ATAC-seq.

581   **b)** Upregulated and downregulated genes (n=540 genes) in *KT;H11<sup>LSL-Cas9</sup>* sg*Inert* relative to  
582   *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1* tumors ( $|\log_2\text{FC}| > 1$ , padj  $< 0.01$ ).

583   **c)** Positive correlation between *KT;Stag2<sup>flox</sup>/KT* ( $\log_2\text{FC}$ ) and *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1*/*KT;H11<sup>LSL-</sup>*  
584   *Cas9* sg*Inert* ( $\log_2\text{FC}$ ). Each dot is a gene. Dots  $|\log_2\text{FC}| > 1$ , padj  $< 0.05$  in both are maroon, dots  
585    $|\log_2\text{FC}| > 1$ , padj  $< 0.05$  in *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1*/*KT;H11<sup>LSL-Cas9</sup>* only are pink, and dots  
586    $|\log_2\text{FC}| > 1$ , padj  $< 0.05$  in *KT;Stag2<sup>flox</sup>/KT* are dark blue. Dots  $|\log_2\text{FC}| > 1$ , padj  $< 0.05$  in neither  
587   are grey.

588   **d)** Downregulated PAXIP1 gene signature enrichment in rank-ordered gene list for *KT* versus  
589   *KT;Stag2<sup>flox</sup>*.

590   **e)** Venn diagram of shared downregulated genes ( $\log_2\text{FC} < -2$ , padj  $< 0.01$ ) between  
591   *KT;Stag2<sup>flox</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1*/*KT;H11<sup>LSL-Cas9</sup>* sg*Inert*.

592   **f)** GO Term Gene Count Analysis with ClusterProfiler and EMBL-EBI GO:Term Category  
593   Analysis established from conserved downregulated pathways between *KT;Stag2<sup>flox</sup>/KT* and  
594   *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1*/*KT;H11<sup>LSL-Cas9</sup>* sg*Inert*.

595   **g)** Venn diagram of shared regions of decreased chromatin accessibility ( $\log_2\text{FC} < -1$ , padj  $< 0.01$ )  
596   between *KT;Stag2<sup>flox</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sg*Paxip1*/*KT;H11<sup>LSL-Cas9</sup>* sg*Inert*.

597 **h)** Accessibility z-score for regions with significantly different accessibility in both  
598 *KT;Stag2<sup>flox</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert* ATAC-seq samples ( $|\log_2$   
599 FC| > 1, padj < 0.01) that are also associated with genes with significantly different expression in  
600 both *KT;Stag2<sup>flox</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert* neoplastic cells ( $|\log_2$   
601 FC| > 1, padj < 0.01). Each row is a differentially accessible region.

602 **i)** Enrichment for regions with significant decreased accessibility in *KT;Stag2<sup>flox</sup>/KT* ( $|\log_2$ FC| > 1,  
603 padj < 0.01) samples associated with genes with significantly decreased expression in both  
604 *KT;Stag2<sup>flox</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert* neoplastic cells ( $\log_2$ FC < -  
605 1, padj < 0.01). Changes in accessibility and expression in *KT;Stag2<sup>flox</sup>/KT* is shown. Each dot is  
606 a differentially accessible region.

607 **j)** Model of lung tumor suppression regulated by STAG2-cohesin-PAXIP1/PAGR1 axis.

608

## 609 METHODS

### 610 611 Design, generation, barcoding, and production of lentiviral and adenoviral vectors

612       sgRNA sequences for each putative tumor suppressor gene from the Tuba-seq<sup>Ultra</sup> pools  
613 were chosen utilizing a combination of CRISPICK  
614 (<https://portals.broadinstitute.org/gppx/crispick/public>) and the Brie library (63). Firstly, the  
615 sgRNAs from the Brie library were filtered based on (1) their predicted on-target efficacy score  
616 (score > 0.2<sup>2</sup>) from CRISPICK and (2) off-target risk. To evaluate off-target risk, the sgRNA  
617 sequence + NGG were aligned to mouse reference genome assembly (GRCm39). sgRNAs that  
618 have a perfect match or a tolerable mismatch (a single mismatch occurring at a position located  
619 more than 10 bases away from the PAM sequence) in an off-target location were regarded as high

620 off-target risk and were excluded. For the Brie sgRNAs that successfully passed this filter, a  
621 maximum of two sgRNAs were selected based on their on-target efficacy score. To ensure the  
622 inclusion of three sgRNAs per gene, additional sgRNAs were designed using CRISPick, adhering  
623 to the aforementioned two criteria. This approach resulted in a final set of sgRNAs, comprising  
624 both tested sgRNAs from the Brie library and novel sgRNAs designed based on one of the most  
625 reliable sgRNA-predicting algorithms available (64). All sgRNA sequences used are shown in  
626 **Supplementary Tables 1, 2, 4.**

627 To generate vectors for Tuba-seq<sup>Ultra</sup>, barcode-containing oligo pools were cloned at the 3'  
628 end of the bU6 promoter in the Tuba-seq<sup>Ultra</sup> backbone vector by Gibson assembly and sgRNA  
629 libraries were added using Golden Gate assembly (35). Prior to adding the barcoded oligo pools,  
630 the last 20 bp of the bU6 promoter and 137 bp filler from the Tuba-seq<sup>Ultra</sup> backbone vector were  
631 removed, followed by DpnI digestion to remove remaining methylated parental vector. The single-  
632 stranded barcoded oligos were cloned into the linearized and DpnI-digested Tuba-seq<sup>Ultra</sup>  
633 backbone vector with Gibson assembly to diversify the sequence of the last 20 bp of the bU6  
634 promoter. PCR-amplified sgRNA pools (TWIST Biosciences) were cloned into the BsmBI-v2-  
635 digested barcoded Tuba-seq<sup>Ultra</sup> vector pool using Golden Gate assembly. The products were  
636 transformed into NEB 10-β electrocompetent bacteria and several million colonies were pooled,  
637 followed by plasmid extractions (Qiagen Plasmid Plus Midiprep kit) and sequencing (Novogene  
638 Corporation, Inc.) to determine the barcode and sgRNA representation (**Supplementary Fig. S2a**).  
639 To generate Lenti-sgRNA/Cre vectors encoding individual sgRNAs, individual sgRNAs were  
640 cloned downstream of the U6 promoter in a pLL3.3 backbone containing the PGK promoter and  
641 Cre using site-directed mutagenesis (34).

642           Lentiviral vectors were produced using polyethylenimine (PEI)-based transfection of 293T  
643    cells with delta8.2 and VSV-G packaging plasmids in 150 mm cell culture plates. Sodium butyrate  
644    (Sigma Aldrich, B5887) was added 8 hours after transfection to achieve a final concentration of  
645    20 mM. Medium was refreshed 24 hours after transfection. 20 mL of virus-containing supernatant  
646    was collected 36 and 48 hours after transfection. The two collections were then pooled and  
647    concentrated by ultracentrifugation (25,000 rpm for 1.5 hours), resuspended overnight in 100 µL  
648    PBS, and frozen at -80°C. Adeno-Spc-Cre was purchased from University of Iowa Core Web.

649           We generated a pool of barcoded lentiviral vectors encoding Cre along with sgRNAs  
650    targeting a panel of genes. Vectors with sgRNAs targeting essential genes (**Supplementary Table**  
651    **S1,2, and 4**) and ‘inert’ sgRNAs were also generated. In this lentiviral system, a diverse barcode  
652    is integrated into the bovine U6 (bU6) promoter directly 5' of the sgRNA. As a result, each clonal  
653    tumor is uniquely identified by a barcode-sgRNA element, and the sgRNA sequence indicates the  
654    specific gene targeted in each tumor.

655    **Mice and tumor initiation**

656           The use of mice for the current study has been approved by the Institutional Animal Care  
657    and Use Committee at Stanford University, protocol number 26696. *Kras*<sup>LSL-G12D/+</sup>  
658    (RRID:IMSR\_JAX:008179), *R26*<sup>LSL-*tdTomato*</sup> (RRID:IMSR\_JAX:007909), *H11*<sup>LSL-*Cas9*</sup>  
659    (RRID:IMSR\_JAX:027632), *Stag2*<sup>fl/fl</sup> (RRID:IMSR\_JAX:030902), and *Smc3*<sup>fl/fl</sup>  
660    (RRID:IMSR\_JAX:030559) mice have been previously described (19,65,66). All mice were on a  
661    C57BL/6 background. Lung tumors were initiated by intratracheal delivery of Lenti-sgRNA/Cre  
662    vectors or Adeno-Spc-Cre vectors (19,64). Adenoviral vectors were CaCl<sub>2</sub>-precipitated prior to  
663    delivery.

664 In our initial screen for cohesin complex members, which was part of a larger screen of  
665 regulators of lung tumorigenesis (35), cohorts of *KT* (n=14) and *KT;H11<sup>LSL-Cas9</sup>* (n=68) were  
666 transduced with  $3 \times 10^5$  and  $1 \times 10^5$  infectious units (IFU) per mouse, respectively (**Fig. 1a**). To  
667 generate lung tumors that were heterozygous mutant for cohesin, tumors were initiated in *KT* and  
668 *KT;Smc3<sup>flox/wt</sup>* mice (n=13-15 mice/group) with  $1 \times 10^9$  IFU Adeno-Spc-Cre (**Fig. 1f**). For our  
669 screen that included sg*Stag2*, sg*Paxip1*, and sg*Pagr1*, which was part of a larger screen of  
670 regulators of lung tumorigenesis, tumor cohorts of *KT* (n=5) and *KT;H11<sup>LSL-Cas9</sup>* (n=31) (35) were  
671 transduced with  $3 \times 10^5$  IFU and  $1 \times 10^5$  IFU, respectively, of the Lenti-U6<sup>BC</sup>-sgRNA/Cre pool. To  
672 generate tumors with inactivation of PAXIP1 for RNA-seq and ATAC-seq, cohorts of *KT* (n=13)  
673 and *KT;H11<sup>LSL-Cas9</sup>* (n=32) mice were transduced with Lenti-sg*Paxip1*/Cre and Lenti-sg*Inert*/Cre  
674 at  $1.5 \times 10^5$  IFU and  $3 \times 10^5$  IFU per mouse, respectively (**Fig. 2d**). For our Tuba-seq<sup>Ultra</sup> screen  
675 of STAG2-mediated tumor suppression, cohorts of *KT* (n=12), *KT;H11<sup>LSL-Cas9</sup>* (n=33), and  
676 *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flox</sup>* (n=35) mice were transduced with  $3 \times 10^5$  IFU,  $1 \times 10^5$  IFU, and  $5 \times 10^4$   
677 IFU of Lenti-U6<sup>BC</sup>-sgRNA/Cre per mouse, respectively (**Fig. 3a-b**). To generate lung tumors for  
678 RNA-seq and ATAC-seq, tumors were initiated in *KT* (n=20) with  $5 \times 10^9$  and  $1 \times 10^9$  IFU per  
679 mouse for collection at the 8- and 16-week timepoints, respectively, and in *KT;Stag2<sup>flox</sup>* (n=20)  
680 with  $2.5 \times 10^8$  and  $2.5 \times 10^7$  IFU per mouse for the 8- and 16-week timepoints, respectively (**Fig.**  
681 **4a**).

682

### 683 **Tuba-seq<sup>Ultra</sup> library generation**

684 Genomic DNA was isolated from bulk tumor-bearing lung tissue from each mouse as  
685 previously described (19,35,67,68). Following homogenization and overnight proteinase K  
686 digestion, genomic DNA was extracted from the lung lysates using standard phenol-chloroform

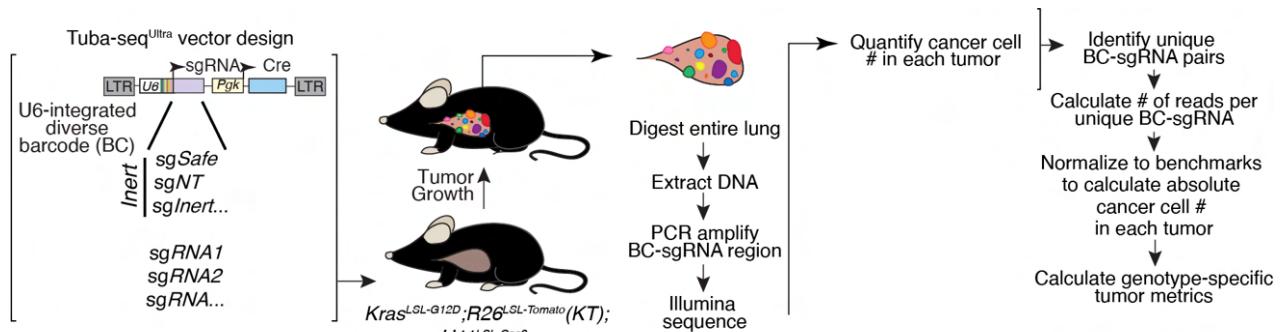
687 methods. Subsequently, Q5 High-Fidelity 2x Master Mix (New England Biolabs, M0494X) was  
688 used to amplify the U6-BC-sgRNA region from 32 µg of genomic DNA in a total reaction volume  
689 of 800 µL per sample. The unique dual-indexed primers used were Forward: AAT GAT ACG  
690 GCG ACC ACC GAG ATC TAC AC-8 nucleotides for i5 index-ACA CTC TTT CCC TAC ACG  
691 ACG CTC TTC CGA TCT-6 to 9 random nucleotides for increased diversity-GCG CAC GTC  
692 TGC CGC GCT G and Reverse: CAA GCA GAA GAC GGC ATA CGA GAT-6 nucleotides for  
693 i7 index- GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC T-9 to 6 random  
694 nucleotides for increased diversity-CAG GTT CTT GCG AAC CTC AT. The PCR products were  
695 purified with Agencourt AMPure XP beads (Beckman Coulter, A63881) using a double size  
696 selection protocol. The concentration and quality of the purified libraries were determined using  
697 the Agilent High Sensitivity DNA kit (Agilent Technologies, 5067-4626) on the Agilent 2100  
698 Bioanalyzer (Agilent Technologies, G2939BA). The libraries were pooled based on lung weight,  
699 cleaned up using AMPure XP beads, and sequenced (read length 2x150bp) on the Illumina HiSeq  
700 2500 or NextSeq 500 platform (Novogene Corporation, Inc.).

701 **Processing of paired-end reads to identify the U6-barcode and sgRNA**

702 Paired ends were first merged using AdapterRemoval (69), and merged reads were parsed  
703 using regular expressions to identify the sgRNA sequence and clonal barcode. When identifying  
704 sgRNA sequences, we strictly required a perfect match with the designed sequences. The 14-  
705 nucleotide random barcode sequence possesses a high theoretical diversity of approximately  $4^{14}$   
706 ( $> 10^8$ ). In each mouse, there are typically fewer than 100 unique tumors associated with each  
707 sgRNA. This indicates that the probability of two genuine unique clonal barcodes being within a  
708 hamming distance of each other is extremely low. Consequently, when we encounter low-  
709 frequency clonal barcodes within a 1-hamming distance of high-frequency clonal barcodes, we

710 attribute them to sequencing or PCR errors. These low-frequency barcodes were merged with  
711 barcodes of higher frequencies.

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



718  
719 Schematic of Tuba-seq<sup>Ultra</sup> vector design and tumor initiation. U6-integrated diverse barcode ( $U6^{BC}$ ) contains a 21-  
720 nucleotide region at the 3' end of the bovine U6 promoter immediately downstream of the TATA box. The sgRNA-  
721 Pool/Cre consists of barcoded lenti-sgRNA/Cre vectors that contained three sgRNAs targeting each gene of interest,  
722 Safe cutting "inert" sgRNAs, and sgRNAs targeting an essential gene. *KT;H11<sup>LSL-Cas9</sup>* are transduced with the lenti-  
723 U6<sup>BC</sup>sgRNA-Pool/Cre. Fourteen-fifteen weeks after tumor initiation, we extract DNA from bulk tumor-bearing lungs  
724 and used Tuba-seq<sup>Ultra</sup> to quantify the impact of targeting each regulator on tumor growth, and tumor initiation for  
725 each tumor of each genotype.  
726

## 727 Summary statistics for overall growth rate

728 To quantify the impact of each gene on tumor growth, we employed a normalization  
729 process that involved calculating statistics for tumors produced by a given sgRNA  $X$  (sg $X$  tumors)  
730 and comparing them to the corresponding statistic of tumors generated by control sgRNAs (sg $Inert$   
731 tumors). Two key statistical measures were employed to characterize these distributions: the size  
732 of tumors at defined percentiles of the distribution (specifically the 50<sup>th</sup>, 70<sup>th</sup>, 80<sup>th</sup>, 90<sup>th</sup>, and 95<sup>th</sup>

733 percentile tumor sizes), and the log-normal mean (LN mean) size. The percentile sizes are  
734 nonparametric summary statistics of the tumor size distribution. By focusing on percentiles  
735 corresponding to the upper tail of the distribution, we specifically examined the growth of larger  
736 tumors, thereby mitigating potential confounding factors arising from variations in cutting  
737 efficiency among guides. The LN mean is the maximum-likelihood estimate of mean tumor size  
738 assuming a log-normal distribution. We normalized these two statistics calculated on tumors of  
739 each genotype to the corresponding *sgInert* statistic. The resulting ratios reflect the growth  
740 advantage or disadvantage associated with each tumor genotype relative to the growth of *sgInert*  
741 tumors.

742  
743 For example, the relative  $i^{th}$  percentile size for tumors of genotype X was calculated as:

744                   *Relative tumor size at  $i^{th}$  percentile for sgX tumors*  
745                   = 
$$\frac{\text{tumor size at } i^{th} \text{ percentile for sgX tumors}}{\text{tumor size at } i^{th} \text{ percentile for sgInert tumors}}$$

746  
747  
748  
749  
750 Likewise, the relative LN mean size for tumors of genotype X was calculated as:

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

753 **Summary statistics for relative tumor number**

754                   In addition to the tumor size metrics described above, we characterized the effects of gene  
755 inactivation on tumorigenesis in terms of the number of tumors (“tumor number”) associated with  
756 each genotype. Unlike the aforementioned metrics of tumor size, tumor number and burden are  
757 linearly affected by lentiviral titer and are thus sensitive to underlying differences in the  
758 representation of each Lenti-sgRNA/Cre vector in the viral pool. Critically, each Tuba-seq<sup>Ultra</sup>

759 experiment included a cohort of *KT* control mice. *KT* mice lack expression of Cas9, rendering all  
760 Lenti-sgRNA/Cre vectors functionally equivalent to non-targeting sgRNA vectors in these mice.  
761 Therefore, the observed tumor number and burden associated with each sgRNA reflect their  
762 respective representation within the viral pool. This control allows us to accurately assess the  
763 specific effects of gene inactivation on tumor formation and growth, accounting for any biases  
764 introduced by the viral vector composition.

765 To assess the extent to which a given gene (*X*) affects tumor number, we therefore first  
766 normalized the number of sg*X* tumors in *KT;H11<sup>LSL-Cas9</sup>* mice by the number of sg*X* tumors in the  
767 *KT* mice:

$$768 \quad \text{Tumor number for sg}X \text{ tumors} = \frac{\text{Number of sg}X \text{ tumors in } KT; H11^{LSL-Cas9} \text{ mice}}{\text{Number of sg}X \text{ tumors in } KT \text{ mice}}$$

769  
770 As with the tumor size metrics, we then calculated a relative tumor number by normalizing this  
771 statistic to the corresponding statistic calculated using sg*Inert* tumors:

$$772 \quad \text{Relative tumor number for sg}X \text{ tumors} = \frac{\text{Tumor number for sg}X \text{ tumors}}{\text{Tumor number for sg}Inert \text{ tumors}}$$

773 Genes that influence relative tumor number modify the probability of tumor initiation  
774 and/or the very early stages of oncogene-driven epithelial expansion, which prior work suggests  
775 are imperfectly correlated with tumor growth at later stages (44). Relative tumor number thus  
776 captures an additional and potentially important aspect of tumor suppressor gene function.

777  
778 **Calculation of confidence intervals and P-values for tumor growth and number metrics**

779 Confidence intervals and *P*-values were calculated using bootstrap resampling for each  
780 sample statistic. To account for both mouse-to-mouse and within mouse variability, we adopted a  
781 two-step, nested bootstrap approach where we first resampled mice, and then resampled tumors  
782 within each mouse to generate resampled data. 10,000 times of bootstrapping was performed to

783 calculate 10,000 values of each statistic. 95% confidence intervals were calculated using the 2.5<sup>th</sup>  
784 and 97.5<sup>th</sup> percentiles of the bootstrapped statistics. Because we calculate metrics of tumor growth  
785 that are normalized to the same metrics in *sgInert* tumors, under the null model where genotype  
786 does not affect tumor growth, the test statistic is equal to 1. Two-sided *P*-values were thus  
787 calculated as followed:

788 
$$p = 2 \cdot \min\{Pr(T > 1), Pr(T < 1)\}$$

789

790 Where T is the test statistic and  $Pr(T > 1)$  and  $Pr(T < 1)$  were calculated empirically as the proportion  
791 of bootstrapped statistics that were more extreme than the baseline of 1. To account for multiple  
792 hypothesis testing, *P*-values were FDR-adjusted using the Benjamini-Hochberg procedure (70).  
793 Summarized statistics of all Tuba-seq<sup>Ultra</sup> experiments in this study can be found in Supplementary  
794 Table 1-4.

795

## 796 Analysis of DepMap data

797 Cancer cell line dependency data (DepMap Public 22Q2) and mutation data (Cancer Cell  
798 Line Encyclopedia) were acquired from the Broad Institute DepMap Portal  
799 (<https://depmap.org/portal/>). First, cell lines without non-silent mutations in *STAG2* were selected.  
800 Then the remaining cell lines were further categorized into five overlapping sets: (1) cell lines with  
801 *KRAS* mutations (with G12 or G13 substitutions), (2) cell lines derived from adenocarcinoma or  
802 non-small cell lung cancers (NSCLC), (3) cell lines obtained from bladder cancer, (4) cell lines  
803 derived from Ewing's sarcoma, (5) cell lines derived from acute myeloid leukemia. Pearson's and  
804 Spearman's correlation coefficients were calculated to measure the correlation between *STAG2*  
805 and other genes using gene effect scores across the aforementioned cell line sets.

806

807 **Tumor dissociation and cancer cell sorting**

808 Tumors were dissociated using collagenase IV, dispase, and trypsin at 37 degrees C for 30  
809 minutes as previously described (44). Cells were stained with DAPI and antibodies against CD45  
810 (30-F11), CD31 (390), F4/80 (BM8), and Ter119 (all from BioLegend) to exclude hematopoietic  
811 and endothelial cells. FACS Aria sorters (BD Biosciences) were used for cell sorting.

812

813 **Western blotting on sorted cancer cells**

814 Cells were lysed in RIPA buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 1% Nonidet  
815 P-40 and 0.1% SDS) and incubated at 4 °C with continuous rotation for 30 min, followed by  
816 centrifugation at 12,000 rcf for 10 min. The supernatant was collected, and the protein  
817 concentration was determined by BCA assay (Thermo Fisher Scientific, 23250). Protein extracts  
818 (10–50 µg) were separated on 4–12% SDS-PAGE and transferred onto polyvinylidene fluoride  
819 membranes. The membranes were blocked with 5% non-fat milk in TBS with 0.1% Tween 20  
820 (TBST) at room temperature for 1 h, cut according to the molecular weight of the target protein  
821 (with at least two flanking protein markers), followed by incubation with primary antibodies  
822 diluted in TBST (1:1,000) at 4 °C overnight. After three 10 min washes with TBST, the membranes  
823 were incubated with the appropriate secondary antibody conjugated to HRP diluted in TBST  
824 (1:10,000) at room temperature for 1 h. After three 10 min washes with TBST, protein expression  
825 was quantified with enhanced chemiluminescence reagents (Fisher Scientific, PI80196).

826 Antibodies used in this study: HSP90 (BD Biosciences, 610418), GAPDH (Cell Signaling,  
827 5174S), PAXIP1 (Sigma-Aldrich, ABE1877), STAG2 (Santa Cruz Biotech., SC-81852), goat-  
828 anti-rabbit IgG antibody, HRP-conjugate (Sigma-Aldrich, 12-348), and goat-anti-mouse IgG  
829 antibody, HRP-conjugate (Thermo Fisher Scientific, 62-6520).

830

831 **RNA-seq on sorted cancer cells**

832 Total RNA was prepared from FACS-sorted neoplastic cells ranging from  $5 \times 10^4$  to  $4 \times$   
833  $10^5$  cells. RNA quality was assessed using the RNA 6000 Pico Assay Kit on the Agilent 2100  
834 Bioanalyzer (Agilent). RNA used for RNA-seq had a mean RIN of 6.6. RNA-seq libraries were  
835 generated by Novogene, Inc. Total RNA (from 2.5-10 ng/sample) was used for cDNA synthesis  
836 using the Ovation RNA-seq System (NuGEN Technologies, Inc.). Two micrograms of NuGEN-  
837 amplified double-stranded cDNA was sheared using a Covaris sonicator to an average length of  
838 490 bp and subjected to library preparation using the Takara V4 Ultra low input mRNA non-  
839 directional library preparation according to the manufacturer's protocol and sequenced on a  
840 NovaSeq PE150 (~6G raw data per sample).

841

842 **Analysis of RNA-seq data sets**

843 Paired-end RNA-seq reads were aligned to the mm10 mouse genome using STAR under  
844 standard input parameters. Two *KT;Stag2<sup>flx</sup>* samples were excluded from subsequent analysis  
845 based on two parameters: 1) >20% the average number of nucleotide counts in Exon 8 of STAG2,  
846 and 2) >40% mRNA expression of *Stag2* from RNA-seq results (TPM). The differentially  
847 expressed genes between different tumor genotypes were called by DESeq2 using the HTSeq-  
848 derived counts as input (71). Unsupervised hierarchical clustering and heatmap visualization of  
849 DEGs were performed using the ‘pheatmap’ package in R (v.1.0.12, [\(1\)](#)).

850 The DESeq2-calculated fold changes were used to generate ranked gene lists for input into  
851 GSEA (v.4.0.3). GSEA was performed using normalized RNA-seq counts against gene signatures  
852 from the MSigDB database (72). Default parameters were used with the following exception: max

853 size= 20,000. Plots were made using the Rtoolbox package in R  
854 (<https://github.com/PeeperLab/Rtoolbox>). clusterProfiler (v.4.3.1.900) was used to perform GO  
855 analyses, GO term enrichment, and functional enrichment analysis (73). GO terms were then  
856 categorized using EMBL's European Bioinformatics Institute (EMBL-EBI) goslim\_mouse  
857 annotations and statistics analysis by “all” relationships (“occurs\_in”, “has\_input”, etc.).

858

### 859 **ssGSEA Analysis of previously published data set**

860 *KT* and *KT;p53<sup>flx,flx</sup>(KPT)* bulk RNA-seq samples were analyzed at various stages of  
861 tumor progression (55 samples) (44). Single sample GSEA was performed on all samples using  
862 the GSVA package v1.46.0 in R (74), and linear regression and unpaired t-tests were performed  
863 in GraphPad Prism.

864

### 865 **ATAC-seq library preparation and data analysis**

866 After cells were sorted and counted, 50-100,000 cells were resuspended in 1 mL of cold  
867 ATAC-seq resuspension buffer (RSB; 10mM Tris-HCl pH 7.4, 10 mM NaCl, and 3 mM MgCl<sub>2</sub> in  
868 water) (75). Cells were centrifuged 500 x g for 5 min at 4 °C. After centrifugation, 900 µL of  
869 supernatant was aspirated. The remaining 100 µL of supernatant was carefully aspirated with a  
870 P20 pipette tip to avoid disturbing the cell pellet. Cell pellets were then resuspended in 50 µL of  
871 ATAC-seq RSB containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin and mixed by  
872 pipetting up and down five times. This reaction was incubated on ice for 3 minutes, and after lysis,  
873 1 mL of ATAC-seq RSB containing 0.1% Tween-20 was added, and tubes were inverted six times  
874 to mix. Nuclei were then centrifuged for 10 min at 500 x g at 4 °C. Supernatant was again removed  
875 using the two-pipette steps, as described before, and nuclei were resuspended in 50 µL of

876 transposition mix (25 µL 2X TD buffer, 2.5 µL transposase (100 nM final), 16.5 µL PBS, 0.5 µL  
877 10% Tween-20, and 5 µL water) by pipetting up and down six times. Transposition reactions were  
878 incubated at 37 °C for 30 minutes at 1000 rpm. Reactions were purified using Qiagen MinElute  
879 Reaction Cleanup Kit. ATAC-seq library preparation was performed as described using the  
880 primers listed in Supplementary Table 10 and sequenced on the NovaSeq 6000 platform (Illumina)  
881 with 2 × 75 bp reads. Adapter-trimmed reads were aligned to the mm10 genome using Bowtie2  
882 (2.1.0). Aligned reads were filtered for quality using SAMtools (v.1.9), duplicate fragments were  
883 removed using Picard (v.2.21.9-SNAPSHOT) and peaks were called using MACS2  
884 (v.2.1.0.20150731) with a q-value cut-off of 0.01 and with a no-shift model. Peaks from replicates  
885 were merged, read counts were obtained using bedtools (v.2.17.0) and normalized using DESeq2  
886 (v.1.26.0). For the *KT* versus *KT;Stag2<sup>flx</sup>* samples all libraries had a TSS enrichment greater than  
887 10. For the *KT;H11<sup>LSL-Cas9</sup>* sgInert and sgPaxip1 samples all libraries had a TSS enrichment greater  
888 than 8.

889

890 **In situ Hi-C library construction on sorted cancer cells**

891 *In situ* Hi-C libraries were prepared from FACS-sorted neoplastic cells with low input  
892 ranging from  $2.5 \times 10^5$  to  $3.5 \times 10^5$  cells. In detail, cells were first pelleted at 300 G and resuspended  
893 in 1 mL Room Temp PBS containing 3% BSA, then fixed as previously described, (76) only with  
894 a higher centrifuge speed at 2,500 G for pelleting the fixed cells. Fixed cell pellets were processed  
895 through the standard Arima Hi-C kit (Catalog # A510008), until the step of Covaris DNA  
896 sonication to an average size of 400 bp. Sheared DNA was purified and size-selected by a two-  
897 step AMPure XP bead (Beckman Coulter #A63882) cleanup at 0.6x and 1.0x, and then inputted  
898 into the 2S Plus DNA Library Kit from Integrated DNA Technologies (Catalog #10009878) for

899 low-input library preparation, until the step before library PCR amplification. Libraries underwent  
900 a biotin pulldown using streptavidin to enrich chromatin interactions before an initial five-cycle  
901 amplification was performed and a qPCR quantification determining the number of additional  
902 cycles required for each library to reach a final concentration around 20 nM in 20 µL. Amplified  
903 libraries were then examined by Agilent 4200 TapeStation for confirmation of size distribution  
904 before submission for sequencing. Libraries were first sequenced using the Illumina MiniSeq PE50  
905 for quality control (analysis described in the next section), and then deeply sequenced using the  
906 Illumina NovaSeq PE50 aiming at  $>3 \times 10^8$  read pairs for each library that passed the MiniSeq  
907 QC.

908

#### 909 **Analysis of in situ Hi-C data**

910 Hi-C data was aligned to the mm10 reference genome using MWA-MEM (77). Reads were  
911 filtered (MAPQ  $\geq 30$ ) and paired using the previously described pipeline (78). PCR duplicate  
912 reads were detected and removed by Picard (<https://broadinstitute.github.io/picard/>). For MiniSeq  
913 quality control analysis, we sorted unique read pairs into three categories: 1) cis-reads that aligned  
914 to the same chromosome with a distance over 1 kb, 2) cis-reads that aligned to the same  
915 chromosome with a distance equal or less than 1 kb, and 3) trans-reads that aligned to two different  
916 chromosomes. Hi-C libraries with a ratio of cis-reads with distances over 1 kb larger than 50% and  
917 trans-reads ratio less than 25% were proceeded with deep sequencing using the NovaSeq.

918 For deeply sequenced Hi-C data, .hic files were generated and normalized (-w 5000) using  
919 JuiceBox (79) after alignment and duplication removal described above. Hi-C data from replicates  
920 within the same treatment were merged before visualization and loop-calling using JuiceBox. For  
921 loop-calling, genome-wide KR normalization and three resolutions were used (-k GW\_KR -r 5000,

922 10000, 25000). Called loop lists from wildtype and *Stag2*-deficient neoplastic cells were compared  
923 to each other and sorted into indicated categories: 1) common loops that both loop anchors have  
924 overlap between *KT* and *KT;Stag2<sup>fl</sup>ox* samples, 2) unique loops that neither loop anchors have  
925 overlap between *KT* and *KT;Stag2<sup>fl</sup>ox* samples, 3) loops that are unique to *KT* that have one loop  
926 anchor that overlaps with a loop unique to *KT;Stag2<sup>fl</sup>ox* samples, 5) loops that are unique to  
927 *KT;Stag2<sup>fl</sup>ox* that have one loop anchor that overlaps with common loops between *KT* and  
928 *KT;Stag2<sup>fl</sup>ox* samples, 6) loops that are unique to *KT;Stag2<sup>fl</sup>ox* that have one loop anchor that over-  
929 laps with a loop unique to *KT* samples (**Fig. 5e**). For loop intensity analysis, loops from both *KT*  
930 and *KT;Stag2<sup>fl</sup>ox* samples were concatenated and removed for redundancy, before using the juicer  
931 dump function to output the ratio of observed/expected for every loop in each separate sample  
932 (**Supplementary Table 11**).

933

### 934 Histology and IHC

935 Lung lobes were preserved in 4% formalin and embedded in paraffin. Hematoxylin and  
936 eosin stains were performed by Stanford Pathology/Histology Service Center (Stanford, CA) or  
937 Histo-Tec Laboratories (Hayward, CA). Total tumor burden (tumor area/total area × 100%) and  
938 individual tumor sizes were determined using ImageJ. Immunohistochemistry (IHC) was  
939 conducted on 4 µm sections. Antigen retrieval was performed with 10mM citrate buffer in a  
940 pressurized decloaking chamber. The slides were washed with 1X PBST and prepared with  
941 VECTASTAIN ABC-HRP Kit (Vector Laboratories, PK-4000). The following primary antibody  
942 and dilution was used: NKX2-1 (1:250, Abcam, ab7013). Sections were developed with DAB  
943 (Vector Laboratories, SK-4100) and counterstained with hematoxylin. To assess NKX2-1  
944 expression in tumors while taking into account potential differences in staining between samples

945 and across sections, we compared NKX2-1 staining intensity between tumors and adjacent normal  
946 tissue. Tumors were binned as having 1) higher expression relative to adjacent normal tissue, 2)  
947 equal expression relative to adjacent normal tissue, or 3) lower expression relative to adjacent  
948 normal tissue.

949

## 950 **Data Availability Statement**

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

955

## 956 **Supplementary Material**

957 **Figure S1** shows the effect of *STAG2* inactivation is highly correlated with that of *PAXIP1* and  
958 *PAGR1* inactivation in human cancer cell lines and genetic interactions with *STAG2* are robust  
959 to different measures of tumorigenesis and tumor growth . **Figure S2** shows validation of  
960 *STAG2* inactivation, clustering of *KT* and *KT;Stag2<sup>flx</sup>* samples, and gene expression of  
961 surfactant genes regulated by *STAG2*, TSS enrichment, and TF enrichment for *KT and*  
962 *KT;Stag2<sup>flx</sup>* samples. **Figure S3** shows DNA loops and DNA anchors that correlate with gene  
963 expression but not chromatin accessibility when comparing *KT* and *KT;Stag2<sup>flx</sup>* samples.  
964 **Figure S4** shows *PAXIP1* and *STAG2*-cohesin regulation of gene expression is conserved in  
965 downregulation but not upregulation of genes, *Paxip1* clustering, TSS enrichment, and TF  
966 enrichment for *KT;H11<sup>LSL-Cas9</sup> sgInert* and *KT;H11<sup>LSL-Cas9</sup>*. **Figure S5** shows summary of the *in*  
967 *vivo* genetic experiments as specific tumor suppressive function of *STAG2*-cohesin and

968 PAXIP1/PAGR1, as STAG2-cohesin and PAXIP1/PAGR1 mediated tumor suppression, and  
969 mutations in *PAXIP1* and *PAGR1* and downregulation of *PAXIP1* mRNA expression in a subset  
970 of acute myeloid leukemias. **Table S1** shows tumor size and number metrics from the cohesin  
971 complex member Tuba-seq<sup>Ultra</sup> analysis. **Table S2** shows tumor size and number metrics from  
972 *Paxip1*, *Pagr1*, *Cdca5*, and various TSGs and essential genes Tuba-seq<sup>Ultra</sup> analysis. **Table S3**  
973 shows genes Selected for Tuba-seq<sup>Ultra</sup> analysis of putative genetic interactors of STAG2. **Table**  
974 **S4** shows tumor size and number metrics from Tuba-seq<sup>Ultra</sup> analysis of putative genetic  
975 interactors of STAG2. **Table S5** shows metrics and RLE-normalized (DESeq2) counts for  
976 *KT;Stag2<sup>flx</sup>* versus *KT* 8 weeks and 16 weeks after tumor initiation RNA-seq experiment. **Table**  
977 **S6** shows Metrics and RLE-normalized (DESeq2) counts for *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* vs  
978 *KT;H11<sup>LSL-Cas9</sup> sgInert* RNA-seq experiment. **Table S7** shows metrics for differentially  
979 expressed genes (DEGs) altered by both STAG2 and PAXIP1 with Log2FoldChange > (abs) 1.  
980 **Table S8** shows metrics and RLE-normalized (DESeq2) counts for *KT;Stag2<sup>flx</sup>* versus *KT* 16  
981 weeks after tumor initiation ATAC-seq experiment. **Table S9** shows metrics and RLE-  
982 normalized (DESeq2) counts for *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* versus *KT;H11<sup>LSL-Cas9</sup> sgInert* 16  
983 weeks after tumor initiation ATAC-seq experiment. **Table S10** shows a list of forward and  
984 reverse primers used to generate ATACseq libraries for *KT;Stag2<sup>flx</sup>* versus *KT* and *KT;H11<sup>LSL-</sup>*  
985 *Cas9 sgPaxip1* versus *KT;H11<sup>LSL-Cas9</sup> sgInert* ATACseq experiments. **Table S11** shows loci for all  
986 loops and normalized metrics for *KT;Stag2<sup>flx</sup>* versus *KT* 16 weeks after tumor initiation Hi-C  
987 experiment.

988

989

990

991 REFERENCES

- 992 1. Ren G, Jin W, Cui K, Rodriguez J, Hu G, Zhang Z, et al. CTCF-Mediated Enhancer-Promoter  
993 Interaction Is a Critical Regulator of Cell-to-Cell Variation of Gene Expression. *Mol Cell*.  
994 2017;67:1049-1058.e6.
- 995 2. Alonso-Gil D, Losada A. NIPBL and cohesin: new take on a classic tale. *Trends Cell Biol*.  
996 Elsevier; 2023;33:860–71.
- 997 3. Arruda NL, Carico ZM, Justice M, Liu YF, Zhou J, Stefan HC, et al. Distinct and  
998 overlapping roles of STAG1 and STAG2 in cohesin localization and gene expression in  
999 embryonic stem cells. *Epigenetics Chromatin*. 2020;13:32.
- 1000 4. van der Lelij P, Lieb S, Jude J, Wutz G, Santos CP, Falkenberg K, et al. Synthetic lethality  
1001 between the cohesin subunits STAG1 and STAG2 in diverse cancer contexts. Musacchio A,  
1002 editor. eLife. eLife Sciences Publications, Ltd; 2017;6:e26980.
- 1003 5. Cuadrado A, Giménez-Llorente D, Kojic A, Rodríguez-Corsino M, Cuartero Y, Martín-  
1004 Serrano G, et al. Specific Contributions of Cohesin-SA1 and Cohesin-SA2 to TADs and  
1005 Polycomb Domains in Embryonic Stem Cells. *Cell Rep*. 2019;27:3500-3510.e4.
- 1006 6. Viny AD, Bowman RL, Liu Y, Lavallée V-P, Eisman SE, Xiao W, et al. Cohesin Members  
1007 Stag1 and Stag2 Display Distinct Roles in Chromatin Accessibility and Topological  
1008 Control of HSC Self-Renewal and Differentiation. *Cell Stem Cell*. 2019;25:682-696.e8.
- 1009 7. Richart L, Lapi E, Pancaldi V, Cuenca-Ardura M, Pau EC-S, Madrid-Mencía M, et al.  
1010 STAG2 loss-of-function affects short-range genomic contacts and modulates the basal-  
1011 luminal transcriptional program of bladder cancer cells. *Nucleic Acids Res*.  
1012 2021;49:11005–21.
- 1013 8. Bailey ML, Tieu D, Habsid A, Tong AHY, Chan K, Moffat J, et al. Paralogous synthetic  
1014 lethality underlies genetic dependencies of the cancer-mutated gene STAG2. *Life Sci  
1015 Alliance*. 2021;4:e202101083.
- 1016 9. De Koninck M, Lapi E, Badía-Careaga C, Cossío I, Giménez-Llorente D, Rodríguez-Corsino  
1017 M, et al. Essential Roles of Cohesin STAG2 in Mouse Embryonic Development and Adult  
1018 Tissue Homeostasis. *Cell Rep*. 2020;32:108014.
- 1019 10. Solomon DA, Kim J-S, Bondaruk J, Shariat SF, Wang Z-F, Elkahloun AG, et al. Frequent  
1020 truncating mutations of STAG2 in bladder cancer. *Nat Genet*. Nature Publishing Group;  
1021 2013;45:1428–30.
- 1022 11. Taylor CF, Platt FM, Hurst CD, Thygesen HH, Knowles MA. Frequent inactivating  
1023 mutations of STAG2 in bladder cancer are associated with low tumour grade and stage and  
1024 inversely related to chromosomal copy number changes. *Hum Mol Genet*. 2014;23:1964–  
1025 74.

- 1026 12. Aquila L, Ohm J, Woloszynska-Read A. The role of STAG2 in bladder cancer. *Pharmacol*  
1027 *Res.* 2018;131:143–9.
- 1028 13. Kim MS, Kim SS, Je EM, Yoo NJ, Lee, SH. Mutational and expressional analyses of  
1029 STAG2 gene in solid cancers. *Neoplasma.* 2012;59:524–9.
- 1030 14. Adane B, Alexe G, Seong BKA, Lu D, Hwang EE, Hnisz D, et al. STAG2 loss rewires  
1031 oncogenic and developmental programs to promote metastasis in Ewing sarcoma. *Cancer*  
1032 *Cell.* 2021;39:827-844.e10.
- 1033 15. Evers L, Perez-Mancera PA, Lenkiewicz E, Tang N, Aust D, Knösel T, et al. STAG2 is a  
1034 clinically relevant tumor suppressor in pancreatic ductal adenocarcinoma. *Genome Med.*  
1035 2014;6:9.
- 1036 16. Bailey ML, O’Neil NJ, van Pel DM, Solomon DA, Waldman T, Hieter P. Glioblastoma cells  
1037 containing mutations in the cohesin component, STAG2, are sensitive to PARP inhibition.  
1038 *Mol Cancer Ther.* 2014;13:724–32.
- 1039 17. Ball AR, Chen Y-Y, Yokomori K. Mechanisms of cohesin-mediated gene regulation and  
1040 lessons learned from cohesinopathies. *Biochim Biophys Acta BBA - Gene Regul Mech.*  
1041 2014;1839:191–202.
- 1042 18. Crompton BD, Stewart C, Taylor-Weiner A, Alexe G, Kurek KC, Calicchio ML, et al. The  
1043 genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* 2014;4:1326–41.
- 1044 19. Cai H, Chew SK, Li C, Tsai MK, Andrejka L, Murray CW, et al. A Functional Taxonomy of  
1045 Tumor Suppression in Oncogenic KRAS–Driven Lung Cancer. *Cancer Discov.*  
1046 2021;11:1754–73.
- 1047 20. Blair LM, Juan JM, Sebastian L, Tran VB, Nie W, Wall GD, et al. Oncogenic context shapes  
1048 the fitness landscape of tumor suppression. *Nat Commun.* Nature Publishing Group;  
1049 2023;14:6422.
- 1050 21. Kong X, Ball Jr. AR, Pham HX, Zeng W, Chen H-Y, Schmiesing JA, et al. Distinct  
1051 Functions of Human Cohesin-SA1 and Cohesin-SA2 in Double-Strand Break Repair. *Mol*  
1052 *Cell Biol.* Taylor & Francis; 2014;34:685–98.
- 1053 22. Mondal G, Stevers M, Goode B, Ashworth A, Solomon DA. A requirement for STAG2 in  
1054 replication fork progression creates a targetable synthetic lethality in cohesin-mutant  
1055 cancers. *Nat Commun.* Nature Publishing Group; 2019;10:1686.
- 1056 23. Ding S, Diep J, Feng N, Ren L, Li B, Ooi YS, et al. STAG2 deficiency induces interferon  
1057 responses via cGAS-STING pathway and restricts virus infection. *Nat Commun.* Nature  
1058 Publishing Group; 2018;9:1485.
- 1059 24. Shen C-H, Kim SH, Trousil S, Frederick DT, Piris A, Yuan P, et al. Loss of cohesin complex  
1060 components STAG2 or STAG3 confers resistance to BRAF inhibition in melanoma. *Nat*  
1061 *Med.* Nature Publishing Group; 2016;22:1056–61.

- 1062 25. Kim J-S, He X, Liu J, Duan Z, Kim T, Gerard J, et al. Systematic proteomics of endogenous  
1063 human cohesin reveals an interaction with diverse splicing factors and RNA-binding  
1064 proteins required for mitotic progression. *J Biol Chem.* 2019;294:8760–72.
- 1065 26. Liu B, Li Z. PTIP-Associated Protein 1: More Than a Component of the MLL3/4 Complex.  
1066 *Front Genet [Internet].* 2022 [cited 2024 Feb 19];13. Available from:  
1067 <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2022.889109>
- 1068 27. Gong Z, Cho Y-W, Kim J-E, Ge K, Chen J. Accumulation of Pax2 Transactivation Domain  
1069 Interaction Protein (PTIP) at Sites of DNA Breaks via RNF8-dependent Pathway Is  
1070 Required for Cell Survival after DNA Damage\*. *J Biol Chem.* 2009;284:7284–93.
- 1071 28. Grimley E, Dressler GR. Are Pax proteins potential therapeutic targets in kidney disease and  
1072 cancer? *Kidney Int.* 2018;94:259–67.
- 1073 29. Lechner MS, Levitan I, Dressler GR. PTIP, a novel BRCT domain-containing protein  
1074 interacts with Pax2 and is associated with active chromatin. *Nucleic Acids Res.*  
1075 2000;28:2741–51.
- 1076 30. Solomon DA, Kim T, Diaz-Martinez LA, Fair J, Elkahloun AG, Harris BT, et al. Mutational  
1077 inactivation of STAG2 causes aneuploidy in human cancer. *Science.* 2011;333:1039–43.
- 1078 31. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis  
1079 of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Sci Signal.*  
1080 American Association for the Advancement of Science; 2013;6:p11–p11.
- 1081 32. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, et al. The cBio cancer  
1082 genomics portal: an open platform for exploring multidimensional cancer genomics data.  
1083 *Cancer Discov.* 2012;2:401–4.
- 1084 33. de Bruijn I, Kundra R, Mastrogiacomo B, Tran TN, Sikina L, Mazor T, et al. Analysis and  
1085 Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE  
1086 Biopharma Collaborative in cBioPortal. *Cancer Res.* 2023;83:3861–7.
- 1087 34. Rogers ZN, McFarland CD, Winters IP, Naranjo S, Chuang C-H, Petrov D, et al. A  
1088 quantitative and multiplexed approach to uncover the fitness landscape of tumor  
1089 suppression in vivo. *Nat Methods.* Nature Publishing Group; 2017;14:737–42.
- 1090 35. Tang, Yuning, Xu, Haiqing, Lopez, Steven, Hughes, Nicholas, Kim, Samuel, Ruiz, Paloma,  
1091 et al. Functional mapping of epigenetic regulators uncovers coordinated tumor suppression  
1092 by HBO1 and MLL1 complexes. *bioRxiv* 2024.08.19.607671
- 1093 36. Canudas S, Smith S. Differential regulation of telomere and centromere cohesion by the Scc3  
1094 homologues SA1 and SA2, respectively, in human cells. *J Cell Biol.* 2009;187:165–73.
- 1095 37. Rivas MA, Durmaz C, Kloetgen A, Chin CR, Chen Z, Bhinder B, et al. Cohesin Core  
1096 Complex Gene Dosage Contributes to Germinal Center Derived Lymphoma Phenotypes  
1097 and Outcomes. *Front Immunol.* 2021;12:688493.

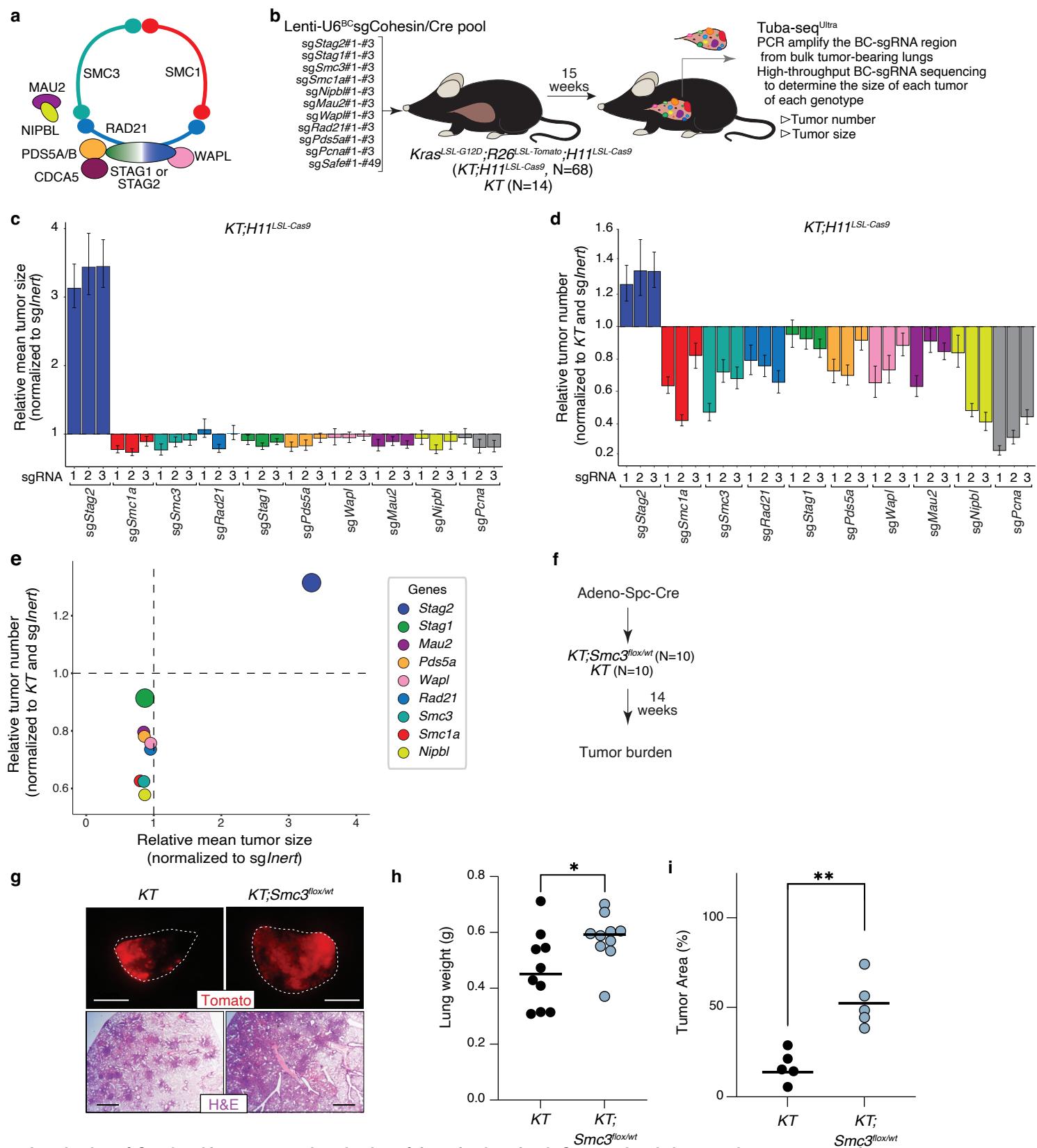
- 1098 38. Rivas MA, Meydan C, Chin CR, Challman MF, Kim D, Bhinder B, et al. Smc3 dosage  
1099 regulates B cell transit through germinal centers and restricts their malignant  
1100 transformation. *Nat Immunol.* Nature Publishing Group; 2021;22:240–53.
- 1101 39. Viny AD, Ott CJ, Spitzer B, Rivas M, Meydan C, Papalex E, et al. Dose-dependent role of  
1102 the cohesin complex in normal and malignant hematopoiesis. *J Exp Med.* 2015;212:1819–  
1103 32.
- 1104 40. Kim E, Dede M, Lenoir WF, Wang G, Srinivasan S, Colic M, et al. A network of human  
1105 functional gene interactions from knockout fitness screens in cancer cells. *Life Sci Alliance*  
1106 [Internet]. Life Science Alliance; 2019 [cited 2024 Feb 19];2. Available from:  
1107 <https://www.life-science-alliance.org/content/2/2/e201800278>
- 1108 41. Pan J, Meyers RM, Michel BC, Mashtalir N, Sizemore AE, Wells JN, et al. Interrogation of  
1109 Mammalian Protein Complex Structure, Function, and Membership Using Genome-Scale  
1110 Fitness Screens. *Cell Syst.* 2018;6:555–568.e7.
- 1111 42. Mayayo-Peralta I, Gregoricchio S, Schuurman K, Yavuz S, Zaalberg A, Kojic A, et al.  
1112 PAXIP1 and STAG2 converge to maintain 3D genome architecture and facilitate  
1113 promoter/enhancer contacts to enable stress hormone-dependent transcription. *Nucleic*  
1114 *Acids Res.* 2023;51:9576–93.
- 1115 43. van der Lelij P, Newman JA, Lieb S, Jude J, Katis V, Hoffmann T, et al. STAG1  
1116 vulnerabilities for exploiting cohesin synthetic lethality in STAG2-deficient cancers. *Life*  
1117 *Sci Alliance.* 2020;3:e202000725.
- 1118 44. Chuang C-H, Greenside PG, Rogers ZN, Brady JJ, Yang D, Ma RK, et al. Molecular  
1119 definition of a metastatic lung cancer state reveals a targetable CD109–Janus kinase–Stat  
1120 axis. *Nat Med.* Nature Publishing Group; 2017;23:291–300.
- 1121 45. Hwang DH, Sholl LM, Rojas-Rudilla V, Hall DL, Shivdasani P, Garcia EP, et al. KRAS and  
1122 NKX2-1 Mutations in Invasive Mucinous Adenocarcinoma of the Lung. *J Thorac Oncol.*  
1123 2016;11:496–503.
- 1124 46. Yamaguchi T, Hosono Y, Yanagisawa K, Takahashi T. NKX2-1/TTF-1: An Enigmatic  
1125 Oncogene that Functions as a Double-Edged Sword for Cancer Cell Survival and  
1126 Progression. *Cancer Cell.* 2013;23:718–23.
- 1127 47. Yang L, Lin M, Ruan W, Dong L, Chen E, Wu X, et al. Nkx2-1: a novel tumor biomarker of  
1128 lung cancer. *J Zhejiang Univ Sci B.* 2012;13:855–66.
- 1129 48. Zheng R, Blobel GA. GATA Transcription Factors and Cancer. *Genes Cancer.* 2010;1:1178–  
1130 88.
- 1131 49. Rao SSP, Huang S-C, Glenn St Hilaire B, Engreitz JM, Perez EM, Kieffer-Kwon K-R, et al.  
1132 Cohesin Loss Eliminates All Loop Domains. *Cell.* 2017;171:305–320.e24.

- 1133 50. Ingram K, Samson SC, Zewdu R, Zitnay RG, Snyder EL, Mendoza MC. NKX2-1 controls  
1134 lung cancer progression by inducing DUSP6 to dampen ERK activity. *Oncogene*. Nature  
1135 Publishing Group; 2022;41:293–300.
- 1136 51. He H, Bell SM, Davis AK, Zhao S, Sridharan A, Na C-L, et al. PRDM3/16 Regulate  
1137 Chromatin Accessibility Required for NKX2-1 Mediated Alveolar Epithelial  
1138 Differentiation and Function. *bioRxiv*. Cold Spring Harbor Laboratory Preprints
- 1139 52. Toba-Ichihashi Y, Yamaoka T, Ohmori T, Ohba M. Up-regulation of Syndecan-4 contributes  
1140 to TGF- $\beta$ 1-induced epithelial to mesenchymal transition in lung adenocarcinoma A549  
1141 cells. *Biochem Biophys Rep*. 2016;5:1–7.
- 1142 53. van Schie JJM, de Lint K, Molenaar TM, Moronta Gines M, Balk JA, Rooimans MA, et al.  
1143 CRISPR screens in sister chromatid cohesion defective cells reveal PAXIP1-PAGR1 as  
1144 regulator of chromatin association of cohesin. *Nucleic Acids Res*. 2023;51:9594–609.
- 1145 54. Serresi M, Siteur B, Hulsman D, Company C, Schmitt MJ, Lieftink C, et al. Ezh2 inhibition  
1146 in Kras-driven lung cancer amplifies inflammation and associated vulnerabilities. *J Exp  
1147 Med*. 2018;215:3115–35.
- 1148 55. Serresi M, Gargiulo G, Proost N, Siteur B, Cesaroni M, Koppens M, et al. Polycomb  
1149 Repressive Complex 2 Is a Barrier to KRAS-Driven Inflammation and Epithelial-  
1150 Mesenchymal Transition in Non-Small-Cell Lung Cancer. *Cancer Cell*. Elsevier;  
1151 2016;29:17–31.
- 1152 56. Callen E, Faryabi RB, Luckey M, Hao B, Daniel JA, Yang W, et al. The DNA damage- and  
1153 transcription-associated protein paxip1 controls thymocyte development and emigration.  
1154 *Immunity*. 2012;37:971–85.
- 1155 57. Winslow MM, Dayton TL, Verhaak RGW, Kim-Kiselak C, Snyder EL, Feldser DM, et al.  
1156 Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature*. 2011;473:101–4.
- 1157 58. Weir BA, Woo MS, Getz G, Perner S, Ding L, Beroukhim R, et al. Characterizing the cancer  
1158 genome in lung adenocarcinoma. *Nature*. 2007;450:893–8.
- 1159 59. Kojic A, Cuadrado A, De Koninck M, Giménez-Llorente D, Rodríguez-Corsino M, Gómez-  
1160 López G, et al. Distinct roles of cohesin-SA1 and cohesin-SA2 in 3D chromosome  
1161 organization. *Nat Struct Mol Biol*. 2018;25:496–504.
- 1162 60. Hara K, Zheng G, Qu Q, Liu H, Ouyang Z, Chen Z, et al. Structure of cohesin subcomplex  
1163 pinpoints direct shugoshin-Wapl antagonism in centromeric cohesion. *Nat Struct Mol Biol*.  
1164 Nature Publishing Group; 2014;21:864–70.
- 1165 61. Yan J, Chen S-AA, Local A, Liu T, Qiu Y, Dorighi KM, et al. Histone H3 lysine 4  
1166 monomethylation modulates long-range chromatin interactions at enhancers. *Cell Res*.  
1167 2018;28:204–20.

- 1168 62. AACR Project GENIE Consortium. AACR Project GENIE: Powering Precision Medicine  
1169 through an International Consortium. *Cancer Discov.* 2017;7:818–31.
- 1170 63. DeWeirdt PC, McGee AV, Zheng F, Nwolah I, Hegde M, Doench JG. Accounting for small  
1171 variations in the tracrRNA sequence improves sgRNA activity predictions for CRISPR  
1172 screening. *Nat Commun.* Nature Publishing Group; 2022;13:5255.
- 1173 64. Doench JG, Fusi N, Sullender M, Hegde M, Vaimberg EW, Donovan KF, et al. Optimized  
1174 sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat  
1175 Biotechnol.* Nature Publishing Group; 2016;34:184–91.
- 1176 65. Jackson EL, Willis N, Mercer K, Bronson RT, Crowley D, Montoya R, et al. Analysis of  
1177 lung tumor initiation and progression using conditional expression of oncogenic K-ras.  
1178 *Genes Dev.* 2001;15:3243–8.
- 1179 66. DuPage M, Dooley AL, Jacks T. Conditional mouse lung cancer models using adenoviral or  
1180 lentiviral delivery of Cre recombinase. *Nat Protoc.* Nature Publishing Group; 2009;4:1064–  
1181 72.
- 1182 67. Rogers ZN, McFarland CD, Winters IP, Seoane JA, Brady JJ, Yoon S, et al. Mapping the in  
1183 vivo fitness landscape of lung adenocarcinoma tumor suppression in mice. *Nat Genet.*  
1184 Nature Publishing Group; 2018;50:483–6.
- 1185 68. Yousefi M, Boross G, Weiss C, Murray CW, Hebert JD, Cai H, et al. Combinatorial  
1186 Inactivation of Tumor Suppressors Efficiently Initiates Lung Adenocarcinoma with  
1187 Therapeutic Vulnerabilities. *Cancer Res.* 2022;82:1589–602.
- 1188 69. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming,  
1189 identification, and read merging. *BMC Res Notes.* 2016;9:88.
- 1190 70. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful  
1191 Approach to Multiple Testing. *J R Stat Soc Ser B Methodol.* 1995;57:289–300.
- 1192 71. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-  
1193 seq data with DESeq2. *Genome Biol.* 2014;15:550.
- 1194 72. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular  
1195 Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1:417–25.
- 1196 73. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological  
1197 Themes Among Gene Clusters. *OMICS J Integr Biol.* 2012;16:284–7.
- 1198 74. Foroutan M, Bhuva DD, Lyu R, Horan K, Cursons J, Davis MJ. Single sample scoring of  
1199 molecular phenotypes. *BMC Bioinformatics.* 2018;19:404.
- 1200 75. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying  
1201 Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol.* 2015;109:21.29.1-21.29.9.

- 1202 76. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D  
1203 map of the human genome at kilobase resolution reveals principles of chromatin looping.  
1204 Cell. 2014;159:1665–80.
- 1205 77. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM  
1206 [Internet]. arXiv; 2013 [cited 2024 May 11]. Available from: <http://arxiv.org/abs/1303.3997>
- 1207 78. Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, et al. Integrative detection and analysis  
1208 of structural variation in cancer genomes. Nat Genet. 2018;50:1388–98.
- 1209 79. Durand NC, Robinson JT, Shamim MS, Machol I, Mesirov JP, Lander ES, et al. Juicebox  
1210 Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. Cell Syst.  
1211 2016;3:99–101.
- 1212

Ashkin et al.



**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<sup>BC</sup>sgRNA/Cre vectors. Genotype and number of mice are indicated. Tuba-seq<sup>Ultra</sup> was performed on each tumor-bearing lung 15 weeks after tumor initiation, followed by analyses to quantify tumorigenesis.

**c**) Relative mean tumor size (normalized to sgInert). Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect.

**d**) Relative tumor number (normalized to KT and sgInert). Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect.

**e**) Relative tumor number (normalized to KT and sgInert) versus relative mean tumor size (normalized to sgInert). Data represents one replicate of two independent experiments.

**f**) Tumor initiation with Adeno-Spc-Cre in KT and KT; Smc3<sup>flx/wt</sup> mice. Mouse number is indicated.

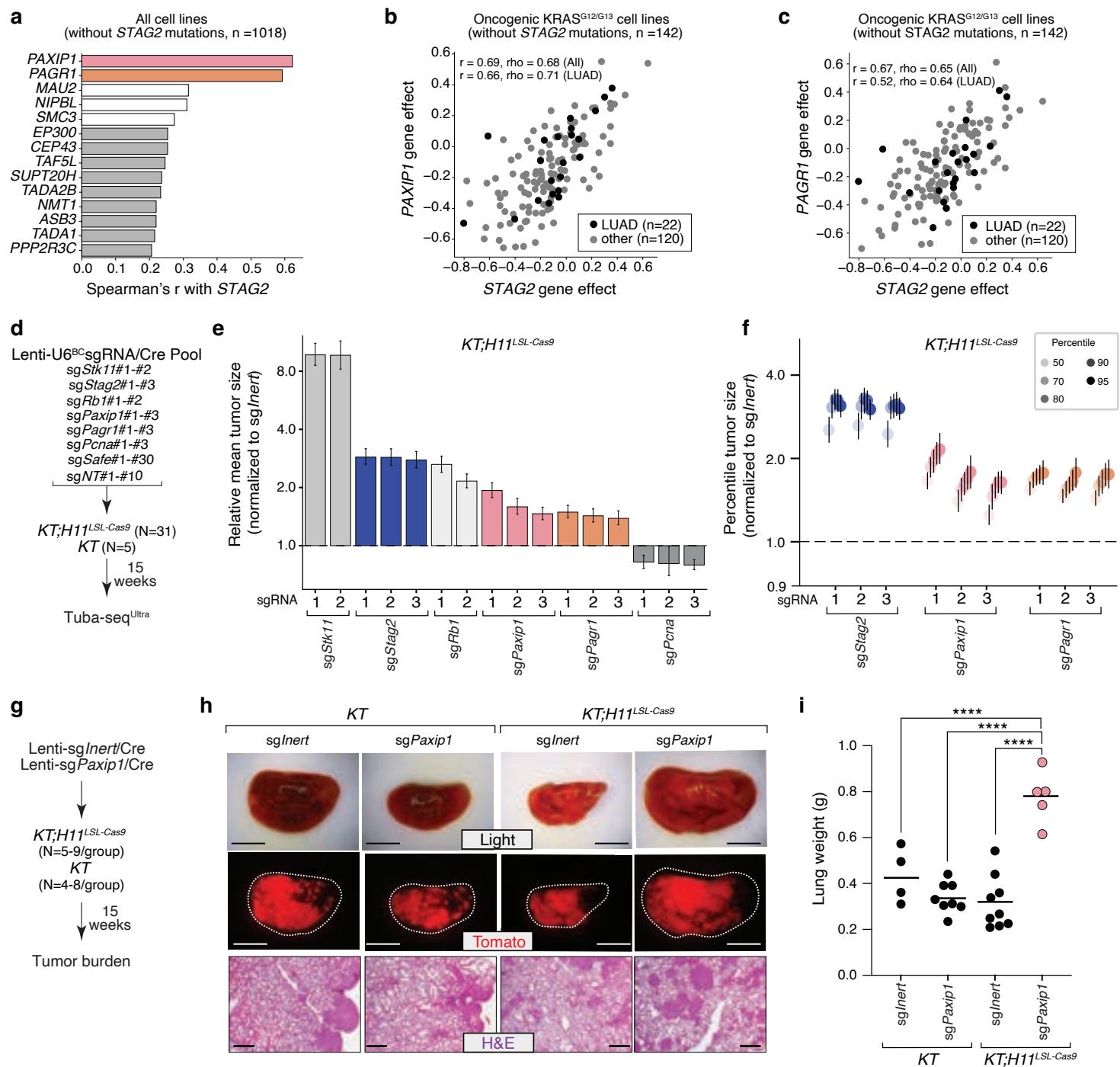
**g**) Representative fluorescence and histology images of lung lobes from the indicated genotypes of mice. Top scale bars, 5 mm, and bottom scale bars, 1 mm.

**h**) Lung weights from KT and KT; Smc3<sup>flx/wt</sup> 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 is shown in Supplementary Table 1.

Ashkin et al.



**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 Dependency Map (DepMap). Cell lines with *STAG2* mutations were excluded. *PAXIP1* and *PAGR1* are colored bars. Core cohesin complex genes are white bars.

**b-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<sup>BC</sup>sgRNA/Cre Tuba-seq<sup>Ultra</sup> vectors. Genotype and number of mice are indicated. Tuba-seq<sup>Ultra</sup> was performed on each tumor-bearing lung 15 weeks after tumor initiation, followed by analyses to quantify tumorigenesis.

**e**) Relative mean tumor size (normalized to sg*Inert*). Mean +/- 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 sg*Inert*) in *KT;H11*<sup>LSL-Cas9</sup> mice. Each gene was targeted with three sgRNAs. Error bars indicate 95% confidence intervals. Dotted line indicates no effect.

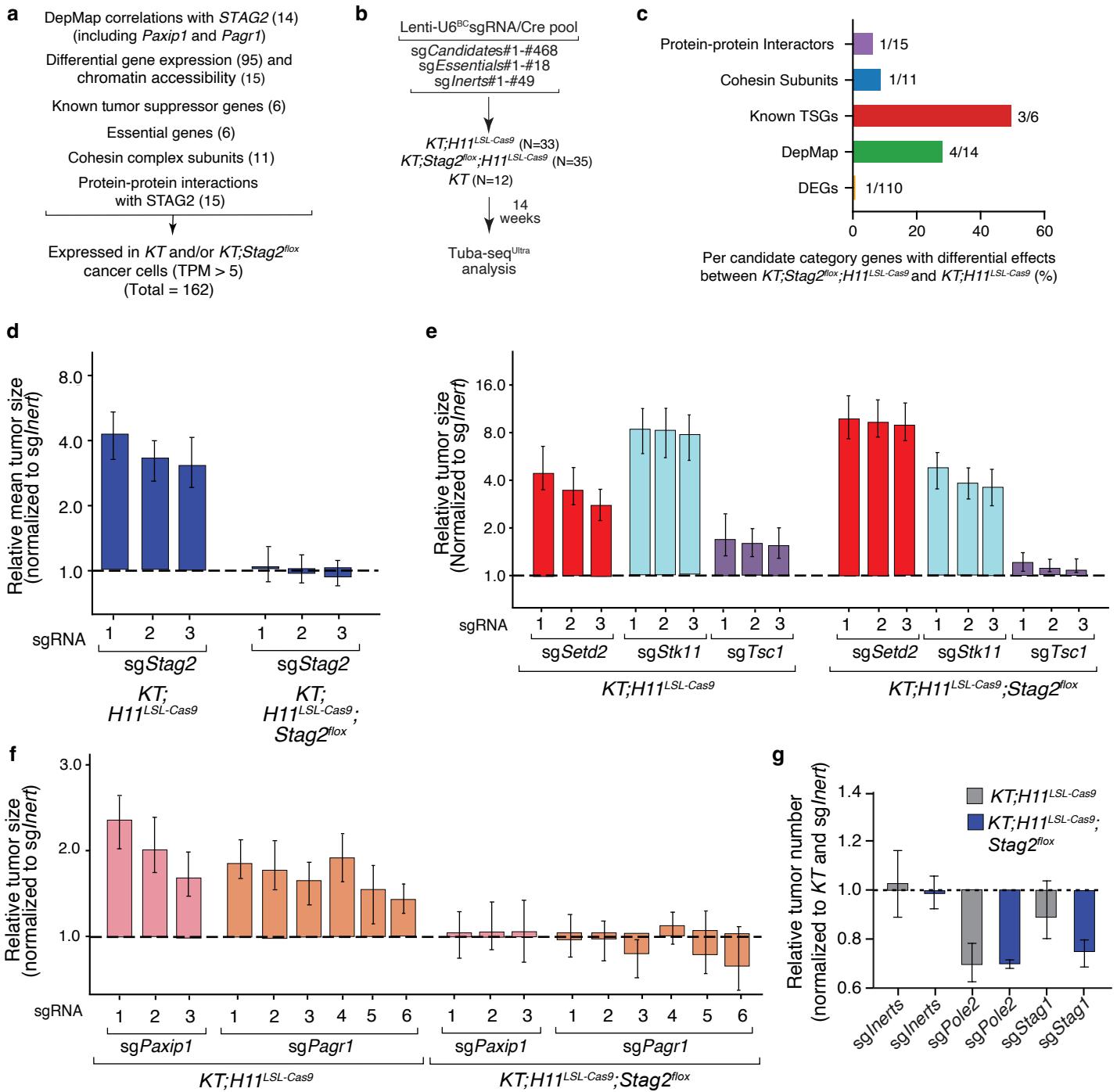
Raw values and significance of each effect is shown in Supplementary Table 2. Data represents one replicate of two independent experiments.

**g**) Tumor initiation with the indicated Lenti-sgRNA/Cre vectors in separate cohorts of *KT* and *KT;H11*<sup>LSL-Cas9</sup> mice. Number of mice in each group is indicated. Tumor burden was quantified 15 weeks 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 µm.

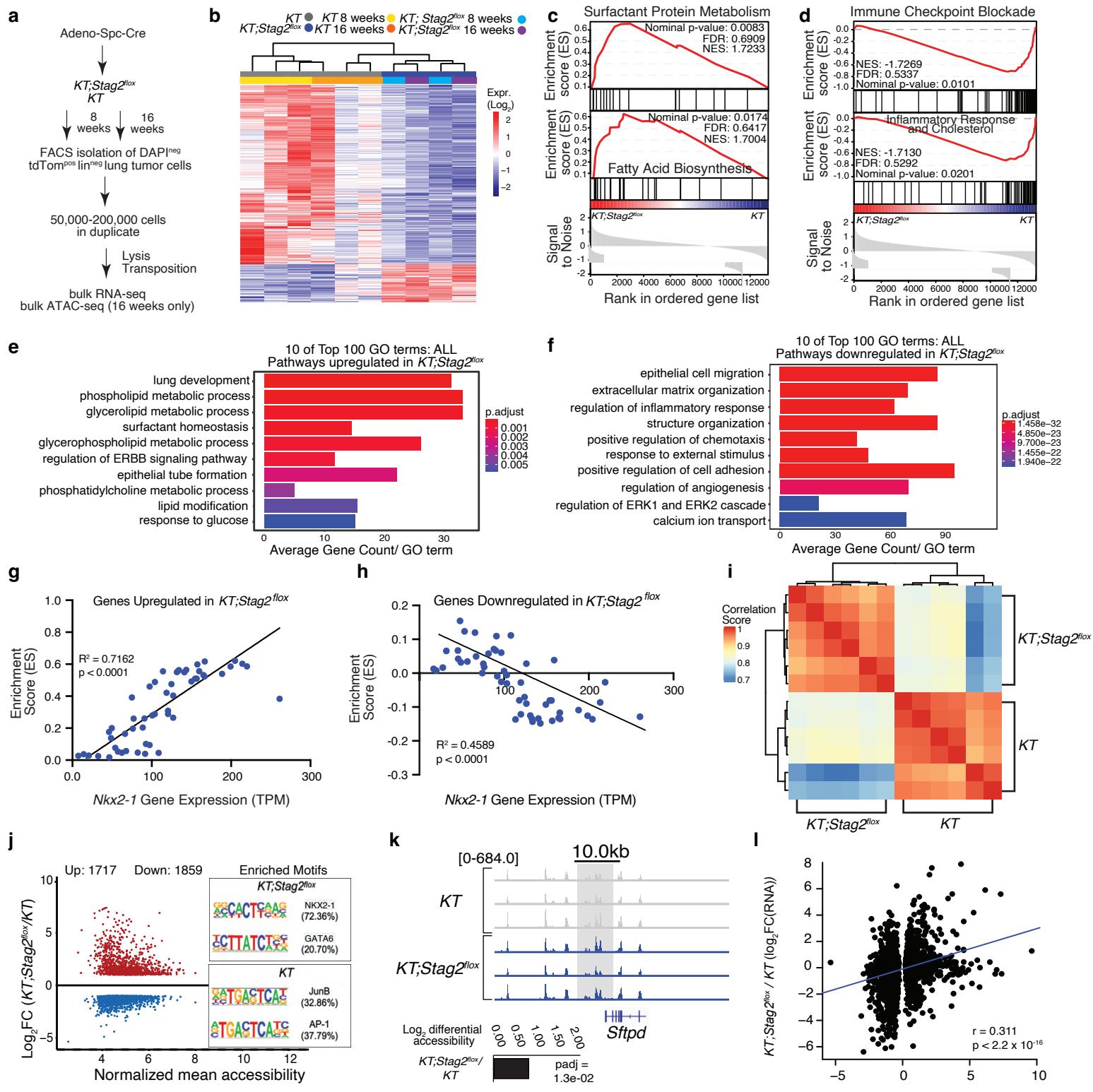
**i**) Lung weights of mice in each group. Each dot represents a mouse and the bar is the mean. \*\*\*p-value <0.0001 by unpaired t-test.

Data represents one replicate of two independent experiments.



**Figure 3. Genetic interactions with STAG2 include modification of overall tumor suppression and functional dependency with PAXIP1 and PAGE1**

- List of candidate gene criteria for pool of 468 barcoded Lenti-sgRNA/Cre vectors.
  - Tumor initiation with Lenti-U6<sup>BC</sup>-sgRNA/Cre vectors in *KT;H11<sup>LSL-Cas9</sup>*, *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>*, and *KT* mice. Number of mice in each group is indicated. Sequencing was performed on each tumor-bearing lung 14 weeks after tumor initiation, followed by analysis to quantify tumorigenesis.
  - Bar plot with percent of genes in each candidate category with differential effects between *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>*.
  - Relative mean tumor size of tumors with sgRNAs targeting *Stag2* (normalized to *sgInert*) in *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice. Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect.
  - Relative mean tumor size of tumors with sgRNAs targeting *Setd2*, *Stk11*, or *Tsc1* (normalized to *sgInert*) in *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice. Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect.
  - Relative mean tumor size of tumors with sgRNAs targeting *Paxip1* or *Page1* (normalized to *sgInert*) in *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice. Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect.
  - Relative tumor number for sgRNAs targeting *Stag1*, *Stag2*, *Pole2* (essential gene), or *Inerts* in *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice. Mean +/- 95% confidence intervals are shown.
- Raw values and significance of each effect is shown in Supplementary Table 4. Data for *KT;H11<sup>LSL-Cas9</sup>* and *KT* mice represent one replicate of two independent experiments.



**Figure 4. STAG2 inactivation increases tumor-related metabolic processes and cell differentiation**

**a**) Schematic of tumor initiation in *KT* and *KT; Stag2<sup>flx</sup>* mice. Mice grew tumors until 8 weeks to represent large atypical adenomas and small adenomas and 16 weeks 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 ( $n = 2545$  genes) in *KT* relative to *KT; Stag2<sup>flx</sup>* tumors (absolute value of  $\log_2$  fold change ( $\log_2FC$ )  $> 1$ ,  $FDR < 0.01$ ).

**c-d**) Gene Set Enrichment Analysis (GSEA) pathways enriched in *KT; Stag2<sup>flx</sup>* relative to *KT* tumors (**c**) and enriched in *KT* relative to *KT; Stag2<sup>flx</sup>* tumors (**d**).

**e-f**) GO Term Gene Count Analysis with ClusterProfiler and EMBL-EBI GO:Term Category Analysis established pathways enriched from upregulated genes for *KT; Stag2<sup>flx</sup>* / *KT* mice (**e**) and from down-regulated genes for *KT; Stag2<sup>flx</sup>* / *KT* mice (**f**).

**g-h**) ES versus NKX2-1 gene expression (TPM) from Single Sample Gene Set Enrichment Analysis (ssGSEA) for *KPT* and *KT* samples (Chuang et al., 2017) for gene sets from genes upregulated in *KT; Stag2<sup>flx</sup>* versus *KT* (**g**) and genes downregulated in *KT; Stag2<sup>flx</sup>* versus *KT* (**h**).

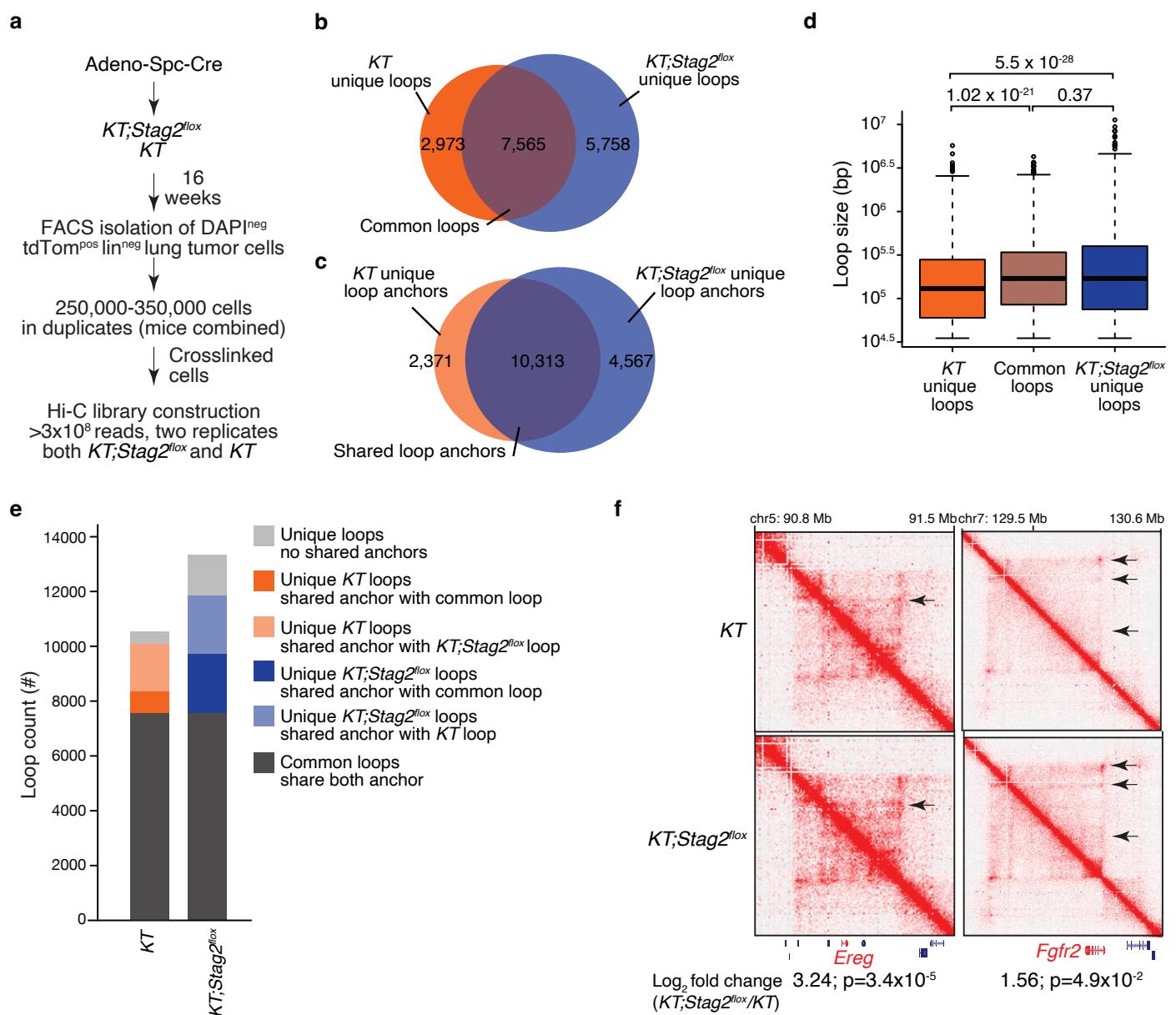
**i**) Rank correlation of chromatin accessibility from ATAC-seq on *KT* and *KT; Stag2<sup>flx</sup>* samples in duplicate. Samples cluster into two distinct groups.

**j**) Differential chromatin accessibility (3 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 ( $\log_2FC > 1$ ,  $FDR < 0.05$ ). Red points are increased chromatin accessibility in *KT; Stag2<sup>flx</sup>* accompanied by transcription factor hypergeometric motif enrichment in *KT; Stag2<sup>flx</sup>*, and blue points are decreased chromatin accessibility in *KT; Stag2<sup>flx</sup>* accompanied by transcription factor hypergeometric motif enrichment in *KT* tumors.

**k**) ATAC-seq signal tracks for the *Sftpd* gene locus in tumor from *KT* and *KT; Stag2<sup>flx</sup>* mice.

**l**) Comparison of  $\log_2$  fold change for *KT* and *KT; Stag2<sup>flx</sup>* samples for top genes from both bulk RNA-seq and bulk ATAC-seq.  $r = 0.311$ ,  $p < 2.2 \times 10^{-16}$ .

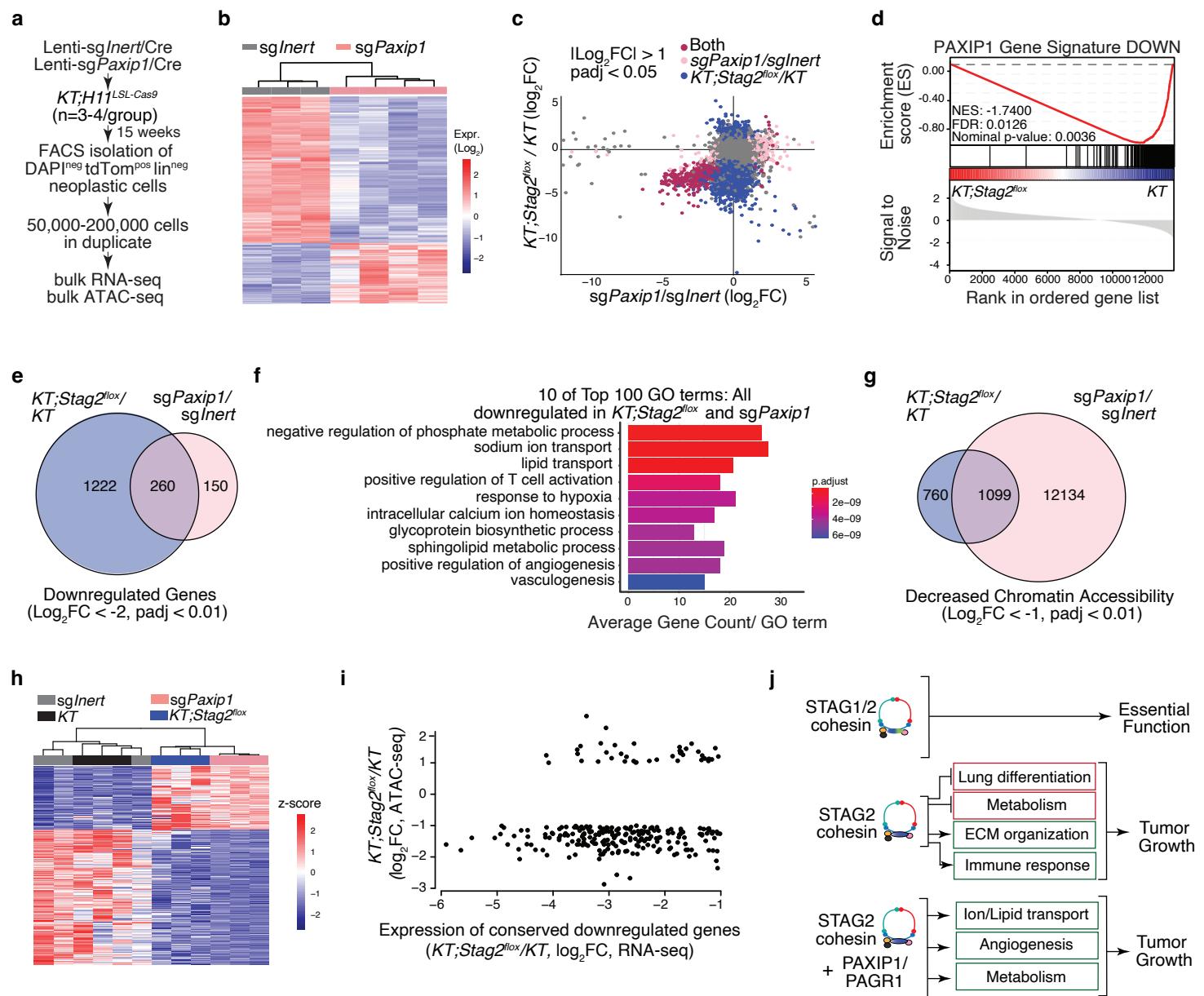
Ashkin et al.



**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<sup>flx</sup>* mice. Outline of tumor cell sorting, crosslinking, and sample preparation for Hi-C.
- b) Venn diagram of loops in *KT* and *KT;Stag2<sup>flx</sup>* samples.
- c) Venn diagram of loop anchors in *KT* and *KT;Stag2<sup>flx</sup>* samples.
- d) Sizes for *KT* unique loops, common loops, and *KT;Stag2<sup>flx</sup>* unique loops. Boxes show median +/- interquartile range. Whiskers show standard error. P-values (Wilcoxon rank test) are shown.
- e) Loops in *KT* and *KT;Stag2<sup>flx</sup>* samples were compared and sorted into the indicated categories based on their unique and shared loop anchors.
- f,g) Differences in chromatin looping and gene expression (from RNA-seq data, log<sub>2</sub> fold change (*KT;Stag2<sup>flx</sup>* / *KT*); p-value) for *Ereg* and *Fgfr2*.

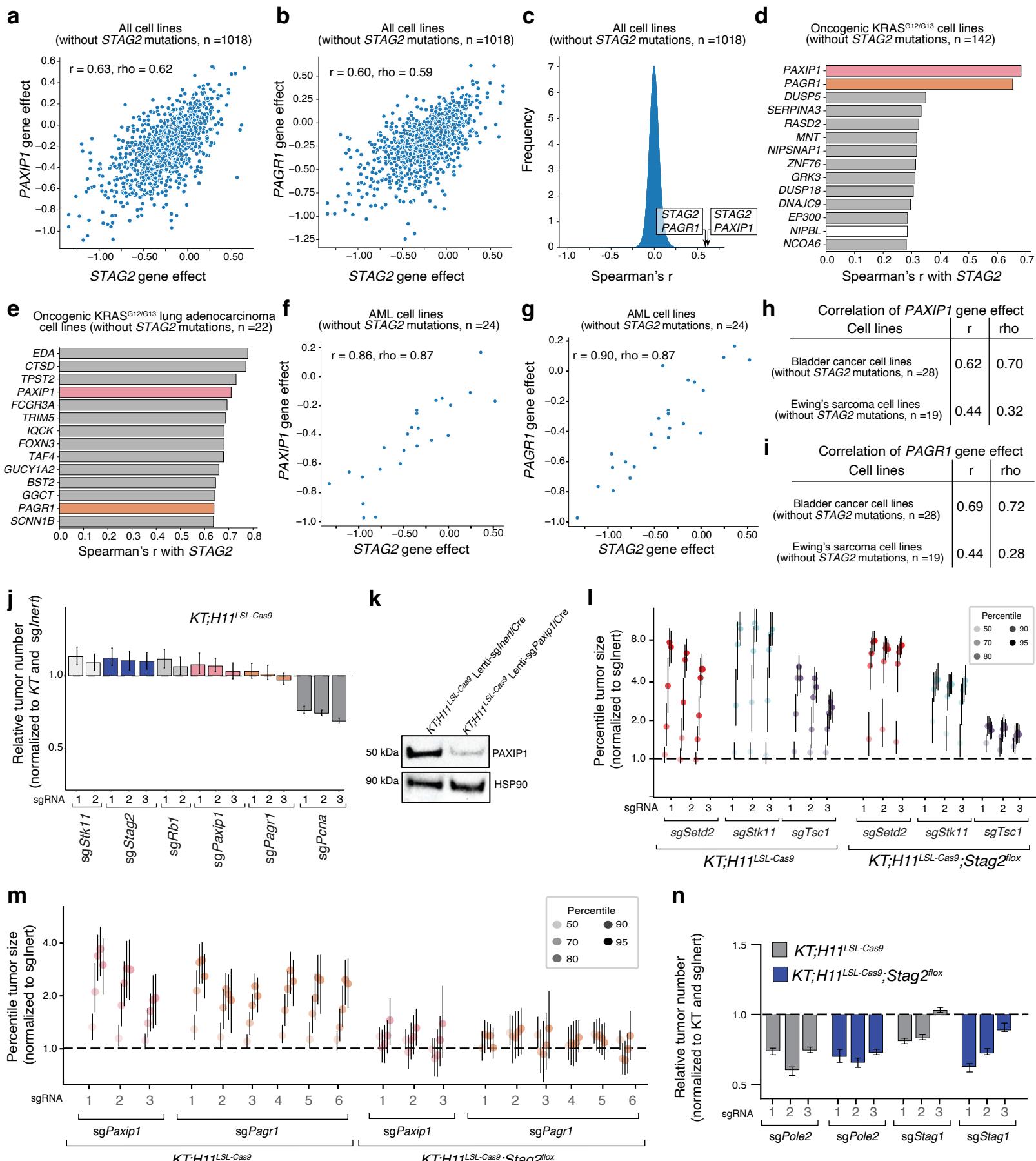
Ashkin et al.



**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<sup>LSL-Cas9</sup>* mice (3-4 mice/group). Mice grew tumors for 15 weeks. 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<sup>LSL-Cas9</sup>* sgInert relative to *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1 tumors ( $I\log_2 FCI > 1$ ,  $padj < 0.01$ ).
- c) Gene expression in *KT;Stag2<sup>flx</sup>/KT* ( $\log_2 FC$ ) and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert ( $\log_2 FC$ ). Each dot is a gene. Dots  $I\log_2 FCI > 1$ ,  $padj < 0.05$  in both are maroon, dots  $I\log_2 FCI > 1$ ,  $padj < 0.05$  in *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* only are pink, and dots  $I\log_2 FCI > 1$ ,  $padj < 0.05$  in *KT;Stag2<sup>flx</sup>/KT* are dark blue. Dots  $I\log_2 FCI > 1$ ,  $padj < 0.05$  in neither are grey.
- d) Downregulated PAXIP1 gene signature enrichment in rank-ordered gene list for *KT* versus *KT;Stag2<sup>flx</sup>*.
- e) Venn diagram of shared downregulated genes ( $\log_2 FC < -2$ ,  $padj < 0.01$ ) between *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert.
- f) GO Term Gene Count Analysis with ClusterProfiler and EMBL-EBI GO:Term Category Analysis established from conserved downregulated pathways between *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert.
- g) Venn diagram of shared regions of decreased chromatin accessibility ( $\log_2 FC < -1$ ,  $padj < 0.01$ ) between *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert.
- h) Accessibility z-score for regions with significantly different accessibility in both *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert ATAC-seq samples ( $I\log_2 FCI > 1$ ,  $padj < 0.01$ ) that are also associated with genes with significantly different expression in both *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert neoplastic cells ( $I\log_2 FCI > 1$ ,  $padj < 0.01$ ). Each row is a differentially accessible region.
- i) Enrichment for regions with significant decreased accessibility in *KT;Stag2<sup>flx</sup>/KT* ( $I\log_2 FCI > 1$ ,  $padj < 0.01$ ) samples associated with genes with significantly decreased expression in both *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup>* sgPaxip1/*KT;H11<sup>LSL-Cas9</sup>* sgInert neoplastic cells ( $\log_2 FC < -1$ ,  $padj < 0.01$ ). Changes in accessibility and expression in *KT;Stag2<sup>flx</sup>/KT* is shown. Each dot is a differentially accessible region.
- j) Model of lung tumor suppression regulated by STAG2-cohesin-PAXIP1/PAGR1 axis.

Data are representative of at least two independent experiments.



Supplementary Figure 1. 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* (Legend on next page)

**Supplementary Figure 1. 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-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-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-g)** Gene effects for *PAXIP1* and *STAG2* inactivation (**f**) and for *PAGR1* and *STAG2* inactivation (**g**) in Acute Myeloid Leukemia (AML). Each dot represents a cell line. Cell lines with *STAG2* mutations were excluded. Spearman's r and Pearson rho are indicated.

**h-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 *siginert*). Mean +/- 95% confidence intervals are shown. Dotted line indicates no effect. Raw values and significance of each effect is shown in Supplementary Table 2 (**j**).

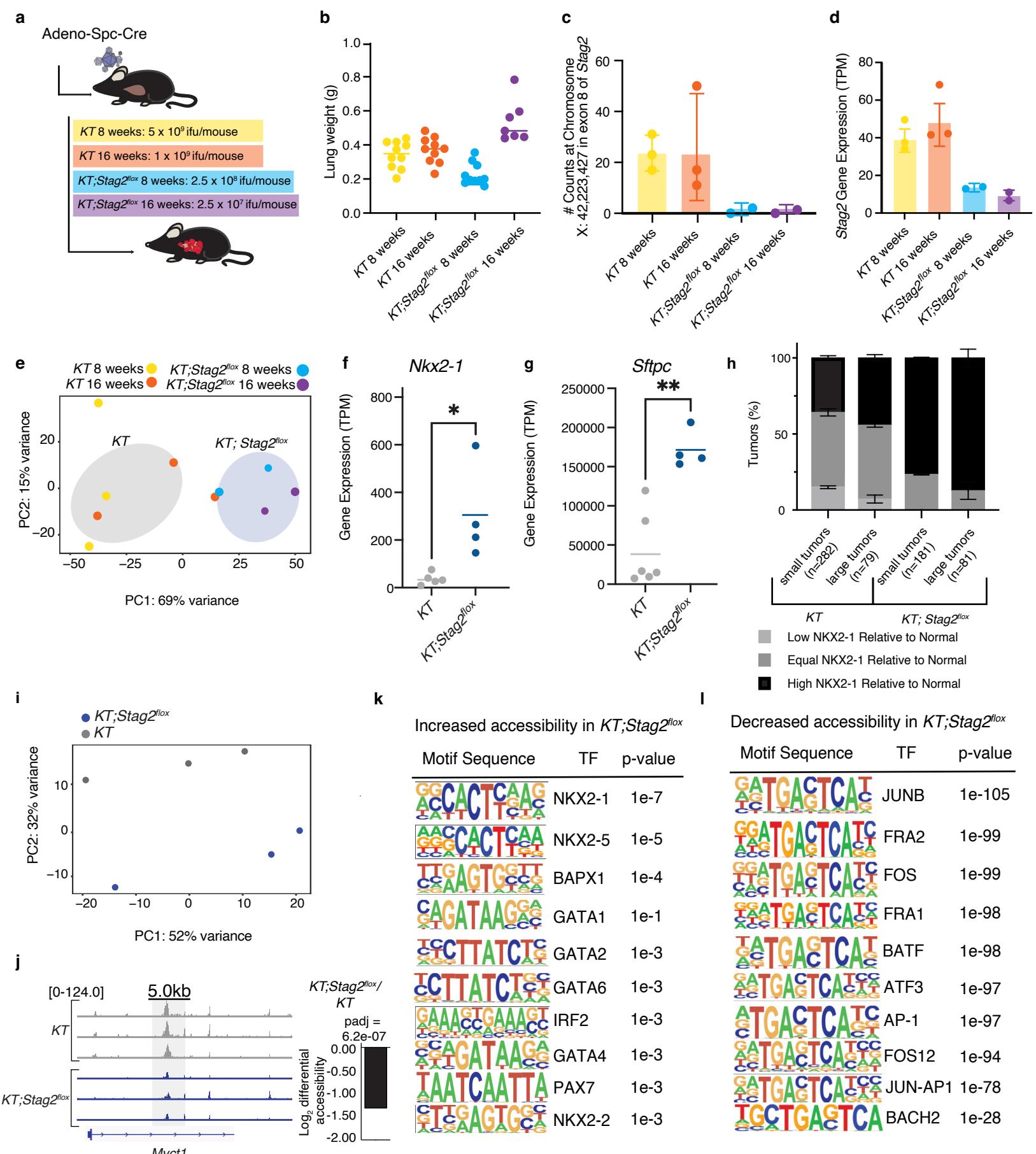
**k)** Western blot on sorted neoplastic cells from *KT;H11<sup>LSL-Cas9</sup>* mice with tumors initiated with the indicated Lenti-sgRNA/Cre vectors. Data represents one replicate of three independent experiments.

**l)** Tumor sizes at the indicated percentiles for tumors with sgRNA targeting *Setd2*, *Stk11*, or *Tsc1* (normalized to *sgInert*) in *KT;H11<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* 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<sup>LSL-Cas9</sup>* and *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* 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<sup>LSL-Cas9</sup>* mice compared to *KT;H11<sup>LSL-Cas9</sup>;Stag2<sup>flx</sup>* mice. Mean +/- 95% confidence intervals are shown.

Raw values and significance of each effect is shown in Supplementary Table 4 (**l-n**).

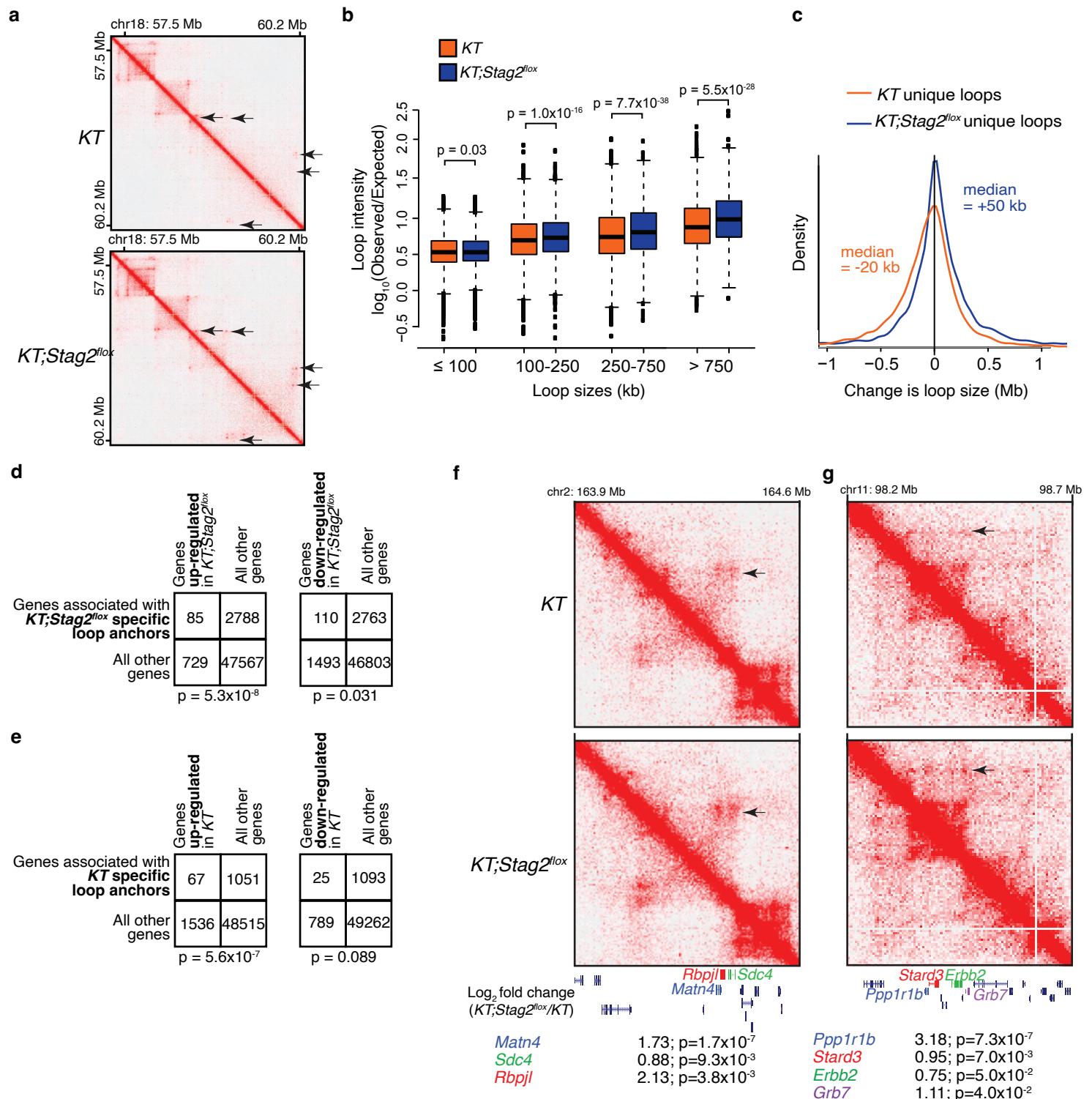


Supplementary Figure 2. Validation of STAG2 inactivation, clustering of KT and KT;Stag2<sup>flx</sup> samples, gene expression of surfactant genes, and differential accessibility regulated by STAG2 (Legend on next page)

**Supplementary Figure 2. Validation of STAG2 inactivation, clustering of *KT* and *KT;Stag2<sup>flx</sup>* 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<sup>flx</sup>* mice to generate tumor samples to analyze at 8 weeks and 16 weeks 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 +/- SEM is shown.
- d)** *Stag2* RNA expression in TPM via RNA-seq in each group. Bar is the standard error. Mean +/- SEM is shown.
- e)** Principal component analysis (PCA) of *KT;Stag2<sup>flx</sup>* and *KT* tumors processed with RNA-seq libraries at 8 weeks and 16 weeks clustered by genotype.
- f-g)** Gene expression (TPM) for *Nkx2-1* (**f**) and *Sftpc* (**g**) for *KT* and *KT;Stag2<sup>flx</sup>* 8 weeks and 16 weeks. \*\* 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<sup>flx</sup>* (n=2 mice) and *KT* (n=2 mice) tumors at 8 weeks after initiation. Tumor NKX2-1 expression quantified by comparison to NKX2-1 expression in adjacent normal tissue. Tumor number quantified labeled in graph. Mean +/- SD is shown (p-value <0.001 by Ordinary one-way ANOVA)
- i)** Principal component analysis (PCA) of *KT* and *KT;Stag2<sup>flx</sup>* tumors processed with ATAC-seq libraries at 16 weeks clustered by genotype.
- j)** ATAC-seq signal tracks for the *Myct1* gene locus in tumors from *KT* and *KT;Stag2<sup>flx</sup>* 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<sup>flx</sup>* 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<sup>flx</sup>* mice.

Ashkin et al.



**Supplementary Figure 3. KT and KT;Stag2<sup>flx</sup> 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<sup>flx</sup> (bottom) neoplastic cells. Contact matrices were visualized in Juicebox with the same color scale. Arrows indicate novel or stronger loop contacts in KT;Stag2<sup>flx</sup>.

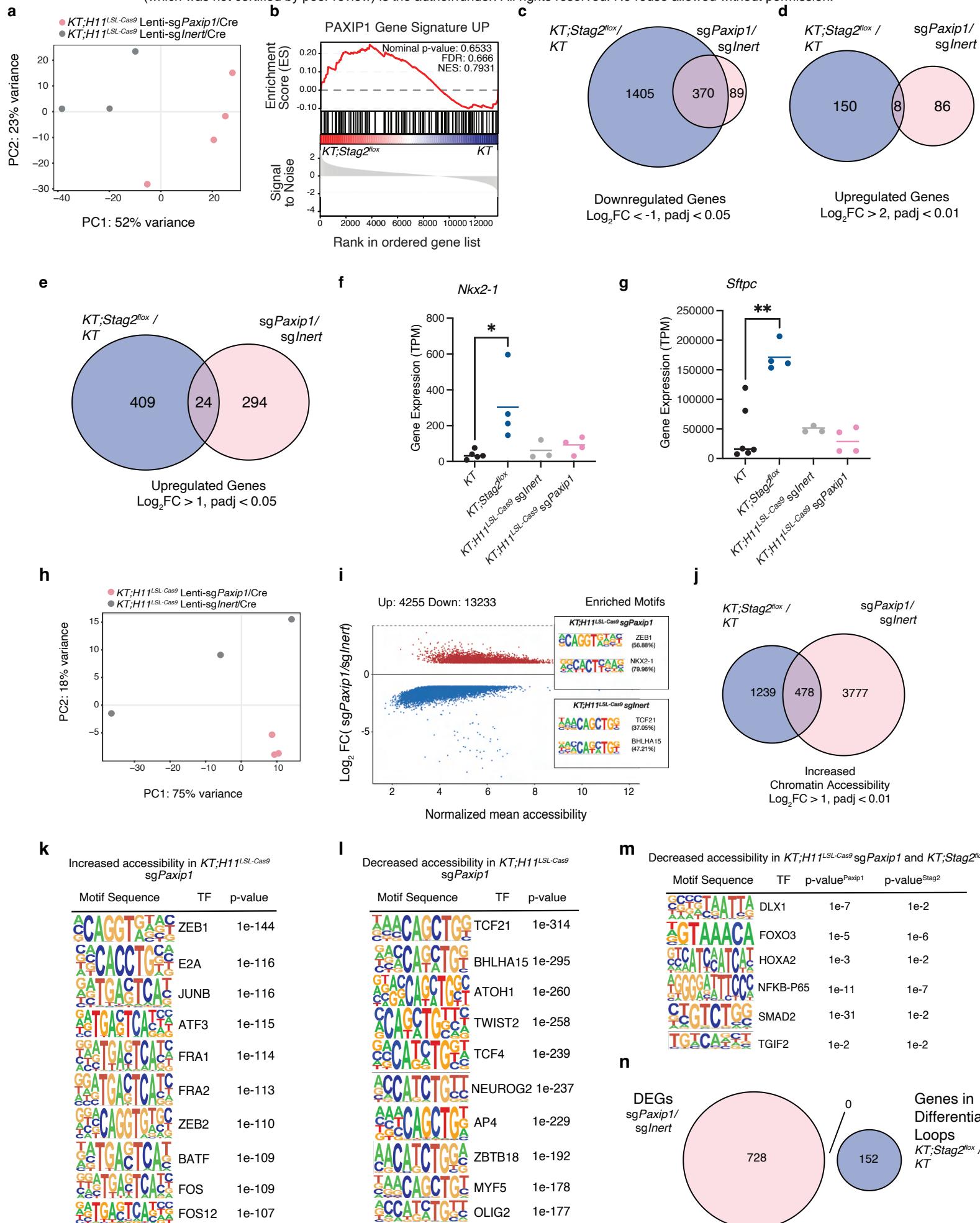
**b** Comparison of loop intensity ( $\log_{10}(\text{Observed}/\text{Expected})$ ) of the indicated loop sizes between all loops in KT and KT;Stag2<sup>flx</sup> samples. Boxes show median +/- interquartile range. Whiskers show standard error. P-values (Wilcoxon rank test) are shown.

**c** Change in loop size between the KT;Stag2<sup>flx</sup> 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<sup>flx</sup> unique loops with which they share one anchor (orange).

**d** Number of genes associated with KT;Stag2<sup>flx</sup> unique loops that are up-regulated (left) or down-regulated (right) in KT;Stag2<sup>flx</sup> samples versus all other genes that are and are not associated with KT;Stag2<sup>flx</sup> unique loops. p-value (chi-square test) is shown.

**e** Number of genes associated with KT unique loops that are up-regulated (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,g** Differences in chromatin looping and gene expression (from RNA-seq data,  $\log_2$  fold change (KT;Stag2<sup>flx</sup> / KT); p-value) of several differentially expressed genes proximal to unique anchor sites.



Supplementary Figure 4. PAXIP1 and STAG2-cohesin regulation of gene expression and chromatin accessibility is conserved in downregulation but not upregulation of genes (Legend on next page)

**Supplementary Figure 4. PAXIP1 and STAG2-cohesin regulation of gene expression and chromatin accessibility is conserved in downregulation but not upregulation of genes**

- a) Principal component analysis (PCA) of RNA-seq on *KT;H11<sup>LSL-Cas9</sup> sgInert* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* tumors.
- b) Upregulated PAXIP1 gene signature enrichment in rank-ordered gene list for *KT* versus *KT;Stag2<sup>flx</sup>*.
- c-e) Venn diagram of shared downregulated genes ( $\log_2 FC < -1$ ,  $padj < 0.05$ ) (c), shared upregulated genes ( $\log_2 FC > 2$ ,  $padj < 0.01$ ) (d), and shared upregulated genes ( $\log_2 FC > 1$ ,  $padj < 0.05$ ) (e) between *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert*.
- f-g) Gene expression (TPM) for *Nkx2-1* (f) and *Sitpc* (g) in neoplastic cells from tumors from *KT*, *KT;Stag2<sup>flx</sup>*, *KT;H11<sup>LSL-Cas9</sup> sgInert*, and *KT;H11<sup>LSL-Cas9</sup> 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<sup>flx</sup>* data are same as Supplementary Fig. S2).
- h) Principal component analysis (PCA) of *KT;H11<sup>LSL-Cas9</sup> sgInert* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* tumors processed with ATAC-seq libraries.
- i) Differential accessibility across 17488 significant peaks in *KT;H11<sup>LSL-Cas9</sup> sgInert* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* mice (3 mice/group). The x-axis represents the  $\log_2$  mean accessibility per peak and the y-axis represents the  $\log_2$  FC in accessibility. Colored dots are significant ( $\log_2 FC > 1$ , FDR < 0.05). Red dots are increased chromatin accessibility in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* accompanied by transcription factor hypergeometric motif enrichment in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1*, and blue dots are decreased chromatin accessibility in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* accompanied by transcription factor hypergeometric motif enrichment in *KT;H11<sup>LSL-Cas9</sup> sgInert*.
- j) Venn diagram of shared regions of increased chromatin accessibility ( $\log_2 FC > 1$ ,  $padj < 0.01$ ) between *KT;Stag2<sup>flx</sup>/KT* and *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert*.
- k-l) Motif sequences and corresponding p-values for transcription factors with motif enrichment in regions with increased accessibility in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* (k) and with decreased accessibility in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* (l).
- m) Motif sequences and corresponding p-values for transcription factors with motif enrichment in regions with decreases accessibility in *KT;H11<sup>LSL-Cas9</sup> sgPaxip1* and *KT;Stag2<sup>flx</sup>*.
- n) Venn diagram of differentially expressed genes regulated by *KT;H11<sup>LSL-Cas9</sup> sgPaxip1/KT;H11<sup>LSL-Cas9</sup> sgInert* ( $|log_2 FCI| > 1$ ,  $padj < 0.05$ ; RNA-seq) versus genes found in differential loops regulated by *KT;Stag2<sup>flx</sup>/KT* (HiC).

**a**

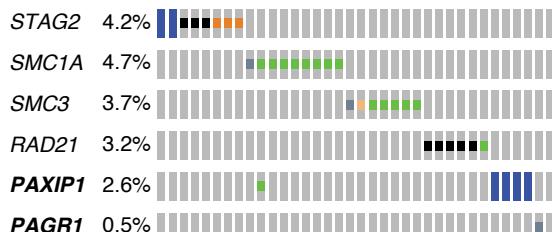
	Effect on lung tumorigenesis <i>in vivo</i>	Essential Function	Tumor Suppressive Function
<i>Stag2</i> inactivation (Fig.1)	increased ↑	perhaps reduced but compensation by <i>Stag1</i>	eliminated
<i>Stag1</i> homozygous inactivation (Fig. 1)	reduced ↓	reduced (partial redundancy with <i>Stag2</i> )	unaffected
<i>Stag1</i> and <i>Stag2</i> inactivation (Fig. 3)	reduced ↓	eliminated	eliminated
<i>Smc3</i> heterozygous inactivation (Fig. 1)	increased ↑	perhaps reduced but not below a critical threshold	reduced below a critical threshold
<i>Smc3</i> homozygous inactivation (Fig. 1)	reduced ↓	eliminated	eliminated
<i>Paxip1</i> or <i>Pagr1</i> homozygous inactivation (Fig. 2,3)	increased ↑	unaffected	reduced

**b**

	Consistent with Genetic Epistasis	Consistent with RNA-seq/ATAC-seq
<b>Model 1</b> STAG2-cohesin regulated tumor suppressive genes: <b>All/most are partially dependent on PAXIP1/PAGR1</b>		Yes No
<b>Model 2</b> STAG2-cohesin regulated tumor suppressive genes: <b>All/most are dependent on PAXIP1/PAGR1, but PAXIP1/PAGR1 also have an offsetting essential function either as part of STAG1-cohesin or any other essential function</b>	 	No No
<b>Model 3</b> STAG2-cohesin regulated tumor suppressive genes: <b>A subset are highly dependent on PAXIP1/PAGR1 while others are not</b>	 	Yes Yes

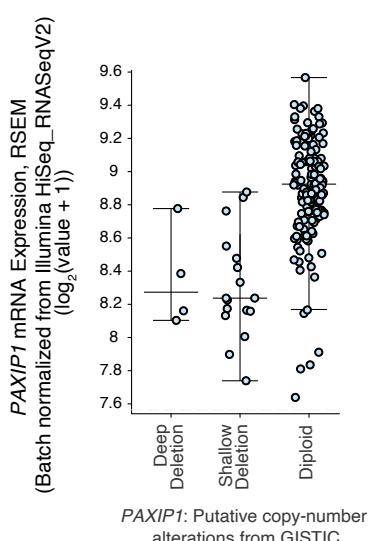
**c**

Acute Myeloid Leukemia  
TCGA, PanCancer Atlas  
190 samples with mutation and copy number data



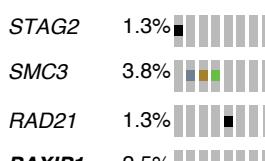
- Missense Mutation (unknown significance)
- Truncating Mutation (unknown significance)
- Splice Mutation (putative driver)
- Splice Mutation (unknown significance)
- Truncating Mutation (putative driver)
- Deep Deletion
- No alterations

**d**



**e**

Acute Myeloid Leukemia  
GENIE Cohort v15.1-public  
DFCI-Oncopanel-3  
80 samples



- Inframe Mutation (unknown significance)
- Truncating Mutation (unknown significance)
- Missense Mutation (unknown significance)
- Truncating Mutation (putative driver)
- No alterations

**Supplementary Figure 5. 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 acute myeloid leukemias (Legend on next page)**

**Supplementary Figure 5. 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 acute myeloid leukemias**

**a)** Interpretations of results from Figures 1-3 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 and 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 & Smith *et al.*, 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 (Figure 3) and molecular analyses (Figures 4-5) 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 acute myeloid leukemias from TCGA accessed through cBioPortal. 190 samples with mutation and copy number data. Mutation type is indicated.

**d)** mRNA expression of PAXIP1 in acute myeloid leukemias 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 acute myeloid leukemias 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.

## Figure S1K.

