

# UCSF

## UC San Francisco Previously Published Works

### Title

LKB1 inactivation modulates chromatin accessibility to drive metastatic progression

### Permalink

<https://escholarship.org/uc/item/7nz19484>

### Journal

Nature Cell Biology, 23(8)

### ISSN

1465-7392

### Authors

Pierce, Sarah E

Granja, Jeffrey M

Corces, M Ryan

et al.

### Publication Date

2021-08-01

### DOI

10.1038/s41556-021-00728-4

Peer reviewed



# HHS Public Access

Author manuscript

*Nat Cell Biol.* Author manuscript; available in PMC 2022 February 02.

Published in final edited form as:

*Nat Cell Biol.* 2021 August ; 23(8): 915–924. doi:10.1038/s41556-021-00728-4.

## LKB1 inactivation modulates chromatin accessibility to drive metastatic progression

Sarah E. Pierce<sup>#1,\*</sup>, Jeffrey M. Granja<sup>#1,2</sup>, M. Ryan Corces<sup>2</sup>, Jennifer J. Brady<sup>1</sup>, Min K. Tsai<sup>1</sup>, Aubrey B. Pierce<sup>1</sup>, Rui Tang<sup>1</sup>, Pauline Chu<sup>7</sup>, David M. Feldser<sup>3</sup>, Howard Y. Chang<sup>1,2,4</sup>, Michael C. Bassik<sup>1,5</sup>, William J. Greenleaf<sup>1,2,\*</sup>, Monte M. Winslow<sup>1,5,6,\*</sup>

<sup>1</sup>Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA.

<sup>2</sup>Center for Personal and Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA.

<sup>3</sup>Department of Cancer Biology and Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA.

<sup>4</sup>HHMI, Stanford University School of Medicine, Stanford, CA, USA.

<sup>5</sup>Chemistry, Engineering, and Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA, USA.

<sup>6</sup>Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA.

<sup>7</sup>Department of Comparative Medicine, Stanford University School of Medicine, Stanford, CA, USA.

# These authors contributed equally to this work.

### Abstract

Metastasis is the leading cause of cancer-related deaths, enabling cancer cells to expand to secondary sites and compromise organ function. Given that primary tumors and metastases often share the same constellation of driver mutations, the mechanisms driving their distinct phenotypes are unclear. Here, we show that inactivation of the frequently mutated tumor suppressor gene, liver kinase B1 (LKB1), has evolving effects throughout lung cancer progression, leading to

\* **Contact Information:** William J. Greenleaf, PhD, Stanford University School of Medicine, 279 Campus Drive, Beckman Center B257, Stanford, CA 94305-5168, wjg@stanford.edu, Fax: 650-725-1534, Phone: 650-725-3672, Monte M. Winslow, PhD, Stanford University School of Medicine, 279 Campus Drive, Beckman Center B256, Stanford, CA 94305-5168, mwinslow@stanford.edu, Fax: 650-725-1534, Phone: 650-725-8696, S.E.P. (pierces@stanford.edu).

**Author contributions:** S.E.P., J.M.G., M.M.W., and W.J.G. conceived the project and designed the experiments. S.E.P. led the experimental data production together with contribution from J.M.G., M.R.C., J.J.B., M.K.T., A.B.P., R.T., and P.C.. S.E.P. and J.M.G. led the data analysis. S.E.P. performed the CRISPR screen analysis and RNA-seq analysis. J.M.G. and S.E.P. performed the ATAC-seq and scATAC-seq analysis. J.M.G. was supervised by H.Y.C. and W.J.G. S.E.P. was supervised by M.C.B., W.J.G., and M.M.W.. S.E.P., J.M.G., W.J.G., and M.M.W. wrote the manuscript with input from all authors.

**Competing interests:** W.J.G. and H.Y.C. are consultants for 10x Genomics who has licensed IP associated with ATAC-seq. W.J.G. has additional affiliations with Guardant Health (consultant) and Protillion Biosciences (co-founder and consultant). M.M.W. is a co-founder of, and holds equity in, D2G Oncology, Inc. H.Y.C. is a co-founder of Accent Therapeutics, Boundless Bio, and a consultant for Arsenal Biosciences and Spring Discovery. The remaining authors declare no competing interests.

**Material availability:** Plasmids generated in this study are available from the Lead Contact upon reasonable request.

**Code availability:** All custom code used in this work is available upon request. We additionally are hosting a Github website that includes the main analysis code used in this study ([https://github.com/GreenleafLab/LKB1\\_2021](https://github.com/GreenleafLab/LKB1_2021))<sup>43</sup>.

the differential epigenetic re-programming of early-stage primary tumors compared to late-stage metastases. By integrating genome-scale CRISPR/Cas9 screening with bulk and single-cell multi-omic analyses, we unexpectedly identify LKB1 as a master regulator of chromatin accessibility in lung adenocarcinoma primary tumors. Using an in vivo model of metastatic progression, we further reveal that loss of LKB1 activates the early endoderm transcription factor SOX17 in metastases and a metastatic-like sub-population of cancer cells within primary tumors. SOX17 expression is necessary and sufficient to drive a second wave of epigenetic changes in LKB1-deficient cells that enhances metastatic ability. Overall, our study demonstrates how the downstream effects of an individual driver mutation can appear to change throughout cancer development, with implications for stage-specific therapeutic resistance mechanisms and the gene regulatory underpinnings of metastatic evolution.

---

The serine/threonine kinase LKB1 (also known as STK11) is frequently inactivated in many cancer types, including pancreatic, ovarian, and lung carcinomas, and germline heterozygous LKB1 mutations cause Peutz-Jeghers familial cancer syndrome<sup>1-3</sup>. Loss of LKB1 leads to both increased primary tumor growth and the acquisition of metastatic ability in lung adenocarcinoma, the most common subtype of lung cancer<sup>4-8</sup>. However, beyond its well-established role as an activator of AMPK-related kinases<sup>6,9-12</sup>, the mechanisms by which LKB1 constrains metastatic ability and cell state are unclear.

In addition to exhibiting high rates of LKB1 mutations, human lung adenocarcinomas frequently harbor mutations in chromatin modifying genes, such as *SETD2*, *ARID1A*, and *SMARCA4*<sup>1</sup>, suggesting that genetic alterations can drive tumor progression by influencing epigenetic state. This interplay between genetic and epigenetic mechanisms is starting to be characterized in other cancer types; for example, H3K27M mutations in diffuse midline gliomas suppress epigenetic repressive capacity and differentiation<sup>13</sup> and SDH-deficiency in gastrointestinal stromal tumors initiates global DNA hyper-methylation and unique oncogenic programs<sup>14</sup>. Chromatin accessibility profiling of mouse lung adenocarcinoma tumors has also started to unveil the epigenomic state transitions that occur during lung cancer development.<sup>15</sup> Genetic profiling of primary tumors and matched metastases has revealed a lack of metastasis-specific driver mutations, opening up the possibility of an epigenetic mechanism of metastasis<sup>16-18</sup>. However, the role of chromatin dynamics in regulating the progression of different genotypes of lung adenocarcinoma, particularly with regards to metastatic spread, remains uncharacterized.

## Lung adenocarcinoma cell lines with restorable alleles of *Lkb1*

To establish a tractable platform to assess LKB1 function in cancer, we generated cell lines from oncogenic KRAS-driven, TP53-deficient (*Kras*<sup>G12D</sup>; *p53*<sup>-/-</sup>; *KP*) murine lung tumors harboring homozygous restorable alleles of *Lkb1* (*Lkb1*<sup>TR/TR</sup>) and a tamoxifen-inducible FlpOER allele (*Rosa26*<sup>FlpOER</sup>) (Extended Data Figs. 1a, b; see Methods)<sup>19</sup>. In LKB1-restorable cell lines (LR1 and LR2), a gene trap cassette within intron 1 of *Lkb1* introduces a splice acceptor site and premature transcription-termination signal before any sequences encoding functional domains of LKB1. Treatment with 4-hydroxytamoxifen (4-OHT) results in FlpOER nuclear translocation and excision of the FRT-flanked gene

trap cassette, thereby restoring full-length expression of *Lkb1* (Extended Data Figs. 1c–e). Restoring LKB1 decreased proliferation in cells in culture and decreased tumor growth after transplantation into mice, whereas treating LKB1-unrestorable, FlpOER-negative cell lines (LU1 and LU2) with 4-OHT had no effect (Extended Data Figs 1f–i).

## A genetic link between LKB1/SIK and chromatin regulation

To identify genes and pathways that contribute to LKB1-mediated tumor suppression, we performed a proliferation-based genome-scale CRISPR/Cas9 knock-out screen in both LKB1-deficient and LKB1-restored cells (Fig. 1a). We first transduced a Cas9-expressing LKB1-restorable cell line (LR1;Cas9) with a lentiviral library containing ~10 sgRNAs per gene in the genome as well as ~13,000 inert controls<sup>20</sup>. After selecting for transduced cells, we treated the cells with 4-OHT or vehicle for 12 days and sequenced the sgRNA region of the integrated lentiviral vectors (Supplementary Table 1 and Extended Data Fig. 1j). As expected, the most highly enriched sgRNA target in LKB1-restored cells compared to LKB1-deficient cells was *Lkb1* itself (Fig. 1b and Extended Data Fig. 1k). Gene Ontology (GO) term enrichment analysis<sup>21</sup> of the remaining top targets surprisingly revealed a strong enrichment of chromatin-related processes (Fig. 1c). In particular, six of the top 20 targets were chromatin modifiers (*Suv39h1*, *Arid1a*, *Eed*, *Suz12*, *Trim28*, and *Smarca1*) (Fig. 1b and Extended Data Fig. 1l), suggesting that the LKB1 pathway engages chromatin regulatory mechanisms to limit growth in lung cancer.

To understand how LKB1 expression affects chromatin accessibility, we performed the assay for transposase-accessible chromatin using sequencing (ATAC-seq) on LKB1-deficient and LKB1-restored cells from two cell lines (LR1;Cas9 and LR2;Cas9) (Fig. 1d, Extended Data Fig. 2a–c, and Supplementary Table 2)<sup>22,23</sup>. Remarkably, LKB1 restoration resulted in consistent, large-scale chromatin accessibility changes, with >14,000 regions increasing and >16,000 regions decreasing in accessibility (Fig. 1d and Extended Data Fig. 2d). LKB1-induced chromatin changes were of similar magnitude to the overarching chromatin accessibility differences between cancer sub-types, such as basal and luminal breast cancer (Extended Data Fig. 2e)<sup>24,25</sup>. In addition, the majority of LKB1-induced chromatin changes occurred within 24–48 hours of LKB1 restoration (Extended Data Fig. 2f–j), suggesting rapid regulation by the LKB1 pathway. Genomic regions with increased accessibility in LKB1-restored cells were enriched for TEAD and RUNX transcription factor binding motifs, whereas genomic regions with increased accessibility in LKB1-deficient cells were enriched for SOX and FOXA motifs (Fig. 1d and Extended Data Fig. 2h). Interestingly, inactivating the top chromatin modifier hits from the screen (*Eed*, *Suz12*, *Trim28*, *Suv39h1*) in the LR1;Cas9 cell line appears to delay, but not prevent, LKB1-induced chromatin accessibility changes (Extended Data Fig. 3), suggesting compensation between chromatin regulatory pathways.

The canonical tumor suppressive role for LKB1 involves the phosphorylation and activation of AMPK-related kinases, including the AMPK, SIK, NUA, and MARK families<sup>9</sup>. To evaluate whether the downstream substrates of LKB1 contribute to LKB1-induced chromatin changes, we knocked out each family with multiple arrays of sgRNAs and performed ATAC-seq with and without LKB1 restoration in the LR1;Cas9 cell line



(Fig. 1e and Extended Data Fig. 4a). Knocking out the *Sik* family (*Sik1*, *Sik2*, and *Sik3* simultaneously) almost entirely abrogated the ability of LKB1 to induce chromatin accessibility changes (Fig. 1f and Extended Data Figs. 4b–g), whereas inactivation of the *Ampk*, *Nuak*, or *Mark* families or the individual *Sik* paralogs (*Sik1*, *Sik2*, or *Sik3* independently) had no effect (Fig. 1f and Extended Data Figs. 4b–j). Therefore, the SIK family of kinases act redundantly, but collectively mediate LKB1-induced chromatin changes.

## LKB1 mutation status defines chromatin sub-types of lung adenocarcinoma

Given the strength of LKB1/SIK-induced chromatin accessibility changes in the murine restoration model, we next evaluated whether LKB1 mutation status correlates with chromatin accessibility differences across human lung adenocarcinoma primary tumors. De novo hierarchical clustering of the 21 lung adenocarcinoma samples from the TCGA ATAC-seq dataset<sup>25</sup> revealed two chromatin sub-types of lung cancer (annotated as Chromatin Type 1 and Chromatin Type 2) (Fig. 2a). Of the top ~200 mutated genes in lung adenocarcinoma, LKB1 was the most significantly enriched mutated gene in Chromatin Type 2 tumors compared to Chromatin Type 1 tumors (FDR = 0.088) (Extended Data Fig. 5a).

We next evaluated how the distinct chromatin accessibility states of Chromatin Type 1 and Chromatin Type 2 human primary tumors compared to the acute chromatin accessibility changes induced by LKB1 restoration in murine cells. We first calculated the differential accessibility of transcription factor binding motifs between Chromatin Type 1 and Type 2 human tumors and between LKB1-restored and LKB1-deficient murine cells using chromVAR<sup>26</sup>. For motifs that were conserved across murine and human datasets, we then compared their differential motif deviation scores (Fig. 2b). Overall the differences between Chromatin Type 1 and Type 2 primary tumors were highly concordant with the differences between LKB1-restored and LKB1-deficient murine lung cancer cells. In particular, genomic regions containing TEAD and RUNX motifs were more accessible in Chromatin Type 1 tumors and LKB1-restored murine cells, and genomic regions containing SOX and FOXA motifs were more accessible in Chromatin Type 2 tumors and LKB1-deficient murine cells (Fig. 2b, c and Extended Data Fig. 5b). These results suggest that LKB1 mutations are not only enriched in Chromatin Type 2 tumors, but also that inactivation of LKB1 is likely a defining feature that divides lung adenocarcinoma into two chromatin accessibility sub-types.

To further evaluate LKB1-dependent effects on chromatin, we performed ATAC-seq on a panel of eight human non-small cell lung cancer cell lines (H1650, H1975, H358, H2009, H1437, A549, H460, H1355). Principle component analysis (PCA) and hierarchical clustering unbiasedly stratified LKB1-wild-type and LKB1-mutant cell lines based on their chromatin profiles (Fig. 2d and Extended Data Fig. 6a). Genomic regions containing RUNX and TEAD motifs were more accessible in LKB1-wild-type cell lines, whereas genomic regions containing SOX and FOXA motifs were more accessible in LKB1-mutant cell lines (Extended Data Figs. 6b–c). Furthermore, similar to the murine restoration model, expressing wild-type LKB1 in LKB1-mutant human lung cancer cells dramatically altered chromatin accessibility, with on average >15,000 regions increasing and >10,000

regions decreasing in accessibility (Fig. 2e and Extended Data Fig. 6d). The magnitude of differential accessibility changes was positively correlated with the baseline LKB1-deficiency gene expression score of each cell line ( $R=0.96$ ) (Extended Data Fig. 6e)<sup>27</sup>. Expression of an orthogonal tumor suppressor KEAP1 in KEAP1-mutant cell lines (A549, H460, and H1355) induced very minor chromatin changes (Extended Data Figs. 6f–h), emphasizing the specificity of the LKB1 tumor suppressor pathway in regulating chromatin accessibility states in lung cancer.

## LKB1-driven chromatin accessibility states in murine primary tumors and metastases

LKB1-deficiency cooperates with oncogenic KRAS in mouse models of lung adenocarcinoma to promote both early-stage tumor growth and late-stage metastasis<sup>4</sup>. To determine whether LKB1 loss has stage-specific effects on tumor progression, we generated an *in vivo* model system to directly compare LKB1-proficient and LKB1-deficient primary tumors and metastases. We incorporated homozygous *Lkb1* floxed alleles into the metastatic, *Kras*<sup>LSL-G12D/+</sup>; *p53*<sup>flx/flx</sup>; *Rosa26*<sup>LSL-tdTomato</sup> (*KPT*) mouse model to maintain a common genetic background between LKB1-proficient and LKB1-deficient tumors. Lentiviral Cre administration into the lungs of *KPT* and *KPT*; *Lkb1*<sup>flx/flx</sup> mice led to the development of aggressive primary tumors capable of seeding spontaneous metastases within 4–7 months. Overall, LKB1-deficiency increased the rate of metastatic progression (Supplementary Table 5 and Extended Data Fig. 7a–d;  $p = 0.00016$ ). We FACS-isolated cancer cells from individual primary tumors and metastases and performed ATAC-seq ( $n=12$  *KPT* primary tumors, 13 *KPT*; *Lkb1*<sup>-/-</sup> primary tumors, 4 *KPT* metastases, and 5 *KPT*; *Lkb1*<sup>-/-</sup> metastases; Fig. 3a). PCA of the 25 primary tumors stratified samples based on LKB1 status, similar to the stratification of Chromatin Type 1 and Chromatin Type 2 human primary tumors (Fig. 3b). In addition, the motif accessibility differences between LKB1-proficient and LKB1-deficient murine samples were consistent in directionality with our previous datasets (Extended Data Fig. 7e, f), with SOX motifs more accessible in LKB1-deficient samples and TEAD, RUNX, and MEF2 binding sites more accessible in LKB1-proficient samples. These results underscore the robustness of LKB1-driven chromatin accessibility states across species and model systems.

## LKB1-deficient metastases activate the transcription factor SOX17

To evaluate genotype- and metastasis-specific epigenetic features, we compared the chromatin accessibility profiles of LKB1-proficient and LKB1-deficient metastases after correcting for their related primary tumor chromatin accessibility profiles. Downregulation of the transcription factor *Nkx2.1* has previously been shown to increase metastatic ability in lung adenocarcinoma<sup>28</sup>; similarly, all metastases had decreased local accessibility at the *Nkx2.1* locus, decreased *Nkx2.1* mRNA expression, and decreased accessibility of genomic regions containing NKX2 motifs compared to primary tumors (Fig. 3c–e and Extended Data Fig. 7g). In contrast, the most prominent genotype-specific difference was that LKB1-deficient metastases had high accessibility of genomic regions containing SOX motifs (Fig. 3f). Of all the SOX family members, LKB1-deficient metastases specifically expressed

high levels of the early endoderm transcription factor *Sox17*, whereas LKB1-deficient primary tumors expressed low levels of *Sox17* and LKB1-proficient primary tumors and metastases did not express *Sox17* (Extended Data Fig. 7h). Similarly, the *Sox17* locus was highly accessible in LKB1-deficient metastases, weakly accessible in LKB1-deficient primary tumors, and inaccessible in LKB1-proficient samples (Fig. 3e). Thus, high SOX17 expression and increased accessibility of genomic regions containing SOX motifs correlate with metastatic progression in LKB1-deficient lung adenocarcinoma. While SOX17 has not previously been associated with the LKB1 pathway, expression of SOX17 in mature lung epithelial cells is sufficient to inhibit differentiation and induce hyperplastic clusters of diverse cell types<sup>29</sup>, suggesting that SOX17 can have strong effects on cell state and behavior.

### LKB1-deficient primary tumors harbor metastatic-like, SOX17+ cells

To characterize the heterogeneity and level of SOX17 protein expression in lung adenocarcinoma, we performed SOX17 immunohistochemistry on LKB1-proficient and LKB1-deficient primary tumors and metastases. LKB1-proficient primary tumors and metastases were universally SOX17-negative, while all LKB1-deficient metastases contained SOX17+ cancer cells (Fig. 4a and Extended Data Figs. 8a, b). In addition, a fraction of LKB1-deficient primary tumors (63/203 tumors) harbored sub-populations of SOX17+ cells, primarily located within invasive acinar structured areas (Extended Data Figs. 8a, b). In support of the hypothesis that LKB1 signaling regulates SOX17 expression, we also found that *LKB1* mRNA expression was negatively correlated with *SOX17* mRNA expression in metastatic lung adenocarcinoma cells derived from human tumors<sup>30</sup> (Extended Data Fig. 8c;  $R = -0.81$ ). In addition, LKB1-deficient Chromatin Type 2 human primary tumors had higher accessibility at the *SOX17* locus compared to Chromatin Type 1 human primary tumors (Extended Data Fig. 8d).

To evaluate the epigenetic profiles of SOX17+ primary tumor cells, we performed droplet-based single-cell ATAC-seq (scATAC-seq)<sup>31,32</sup> on cancer cells from LKB1-proficient (n=4) and LKB1-deficient (n=3) primary tumors (Fig. 4b and Extended Data Figs. 9a, b). We identified 12 distinct clusters of cells (Fig. 4c and Extended Data Fig. 9c; see Methods)<sup>33</sup>, with clusters 1–5 primarily composed of LKB1-proficient cells and clusters 6–12 primarily composed of LKB1-deficient cells. However, cells in cluster 12 (n=112 cells) stood out as a potential source of metastatically competent LKB1-deficient cells, exhibiting the highest accessibility near the *Sox17* locus as well as the lowest accessibility near the *Nkx2.1* locus (Figs. 4d–f). Cluster 12 is primarily composed of cells from two LKB1-deficient primary tumors derived from mouse 13 (13A and 13B). Motif enrichment and transcription factor footprinting<sup>25</sup> revealed high flanking accessibility of SOX-containing genomic regions and a loss of the NKX2 footprint in cells in cluster 12 (Fig. 4g, h and Extended Data Fig. 9d). Furthermore, genomic regions with the highest accessibility in LKB1-deficient primary tumors had the lowest average accessibility in cells in cluster 12 compared to clusters 1–11 (Extended Data Fig. 9e), while genomic regions with the highest accessibility in LKB1-deficient metastases had the highest average accessibility in cells in cluster 12 compared to clusters 1–11 (Extended Data Fig. 9f). Thus, sub-populations of cancer cells within LKB1-

deficient primary tumors exhibit chromatin features suggestive of a SOX17+, metastatic-like state.

## SOX17 maintains accessibility of genomic regions containing SOX binding sites

To further establish a link between LKB1 and SOX17 during metastatic progression, we evaluated the effect of LKB1 restoration on SOX17 expression in our metastatic, LKB1-restorable cell lines (LR1 and LR2). Restoring LKB1 was sufficient to dramatically reduce *Sox17* mRNA expression and local accessibility at cis-regulatory sites near the *Sox17* locus (Fig. 5a, b and Extended Data Fig. 10a). Restoring LKB1 was also associated with a global loss of accessibility at genomic regions containing SOX binding sites in human and murine cell lines following LKB1 restoration (Fig. 1d and Extended Data Fig. 2h–i, 6f). GO term enrichment analysis of the genes closest to these genomic regions revealed decreased accessibility near genes related to the positive regulation of epithelial cell adhesion and extracellular matrix assembly, with implications for how cancer cells interact with the microenvironment and surrounding cell types (Extended Data Fig. 10b). Notably, inactivating the SIK family of kinases prior to LKB1 restoration was sufficient to maintain high *Sox17* mRNA expression and local chromatin accessibility at the *Sox17* locus (Extended Data Fig. 10c, d). These results suggest that LKB1-deficient metastases not only express higher levels of SOX17 compared to LKB1-proficient metastases, but also that the LKB1-SIK pathway actively inhibits the expression and thus activity of SOX17.

To evaluate whether SOX17 is required to maintain accessibility at genomic regions containing SOX binding sites, we inactivated *Sox17* with two sgRNAs in the LR2;Cas9 cell line and performed ATAC-seq with and without LKB1 restoration (Fig. 5c and Extended Data Fig. 10e). In LKB1-deficient metastatic cells, *Sox17* inactivation decreased accessibility at SOX-containing genomic regions to levels approaching that of LKB1-restored cells (Fig. 5d–e and Extended Data Fig. 10f). Next, we overexpressed *Sox17* cDNA and performed ATAC-seq with and without LKB1 restoration (Extended Data Fig. 10h). In LKB1-restored cells, *Sox17* expression led to the maintenance of accessibility at genomic regions containing SOX binding sites (Extended Data Figs. 10i; cluster 4). We confirmed these results in a second independent cell line (LR1;Cas9) (Extended Data Fig. 10g). Furthermore, expression of a sgRNA-resistant *Sox17* cDNA abrogated the effects of knocking out endogenous *Sox17* (Figure 5g). Therefore, SOX17 is necessary and sufficient to maintain accessibility at genomic regions containing SOX binding sites in LKB1-deficient, metastatic cells.

## SOX17 drives tumor growth of metastatic, LKB1-deficient cells

To further evaluate whether SOX17 regulates the growth of metastatic lung cancer cells, we inactivated *Sox17* with sgRNAs or overexpressed *Sox17* cDNA in the LR2;Cas9 LKB1-restorable cell line, restored LKB1, and injected each cell population intravenously into recipient mice (Fig. 6a). After three weeks of growth, we evaluated the colonization and growth of cells in the lung. Knocking out *Sox17* in LKB1-deficient cells resulted in a significantly reduced tumor burden relative to an sgSafe control (Fig. 6b, c and Extended

Data Fig. 10g). In contrast, overexpressing *Sox17* in LKB1-restored cells increased tumor burden (Fig. 6b, c and Extended Data Fig. 10h). In addition, to evaluate the ability of *Sox17*-overexpressing cells to both leave the primary tumor and establish metastases, we injected LKB1-restored cells with and without overexpressed *Sox17* cDNA subcutaneously into recipient mice (Fig. 6d). After five weeks of growth, we evaluated cells that had left the subcutaneous “primary” tumor and colonized in the lung (Fig. 6e). While overexpressing *Sox17* did not change subcutaneous tumor growth (Fig. 6f, top), *Sox17*-overexpressing cells had a significantly greater ability to colonize the lung (Fig. 6f, bottom). Further, to evaluate the ability of *Sox17*-overexpressing cells to form metastases elsewhere in the body, we injected LKB1-restored cells with and without overexpressed *Sox17* cDNA intrasplenically into recipient mice (Extended Data Fig. 10i). After three weeks of growth, *Sox17*-overexpressing cells had a greater ability to colonize to the liver ( $p=0.055$ ) (Extended Data Fig. 10j–k). Thus, SOX17 drives a genotype-specific epigenetic program that promotes the metastatic competency of LKB1-deficient cells.

## Discussion:

Here we show that inactivation of LKB1, a tumor suppressive kinase, drives widespread chromatin accessibility changes in lung adenocarcinoma that evolve throughout cancer progression. While LKB1 has been well-studied for its metabolic roles in cancer, LKB1-induced chromatin changes are surprisingly AMPK-independent and depend almost exclusively on expression of the SIK family of kinases. Recent studies have additionally revealed that deleting AMPK hurts rather than helps lung cancer growth, and AMPK1 is preferentially amplified in lung adenocarcinoma, suggesting that AMPK is not a classic tumor suppressor in this cancer type<sup>5,33</sup>. Therefore, SIKs are emerging as the main drivers of LKB1-mediated tumor suppression and epigenetic regulation in lung cancer. Interestingly, the SIK family of kinases has a known role in the inhibition of class IIa histone deacetylases (HDACs)<sup>34</sup>. In contrast to other classes of HDACs, Class IIa HDACs do not have the typical core enzymatic domain required for deacetylating histones; however, they form multiprotein complexes with transcription factors to interact with chromatin<sup>35</sup>. Thus, the SIK-HDAC relationship might be relevant for future studies attempting to dissect the regulation of SOX17 and the overall chromatin accessibility states of LKB1-deficient and LKB1-proficient cells.

Overall, our findings reveal that inactivation of LKB1/SIK signaling drives two separate waves of epigenetic re-modeling, with the first set of changes occurring within lung primary tumors and the second set of changes mediated by cis-regulatory activation of the transcription factor SOX17 in metastatic cells (Fig. 6g). Thus, the downstream effects of a driver mutation can change throughout tumor development and subsequently enhance metastatic ability. While the LKB1 pathway likely constitutively represses SOX17, the consequences of this repression are not observed until SOX17 expression is activated during metastatic transformation. However, as not all LKB1-deficient cancer cells express SOX17, there must be a second signal that initiates the metastatic program, which currently remains unknown. Regulation of the strong endodermal transcription factor SOX17 could also have implications for the diverse histological sub-types observed in LKB1-deficient

lung tumors<sup>4,36</sup>, and further work to understand the plasticity of LKB1-deficient cells in the context of such widespread chromatin accessibility changes is warranted.

By resolving the epigenetic landscape of lung adenocarcinoma primary tumors at single-cell resolution, we further discovered sub-populations of cancer cells in primary tumors that share a common epigenetic state with the cancer cells in metastases. This result suggests that primary tumors harbor rare and epigenetically distinct cells that are “poised” to seed distant metastases, rather than evolving a specialized cell state after metastatic colonization. An early mechanism of epigenetic transformation opens up the possibility of identifying biomarkers to predict which tumors have already seeded micrometastases before detection is possible. In addition, we anticipate that genotype-driven epigenetic differences between primary tumors and metastases will likely inform how patients respond to personalized therapies. Overall, these findings help to explain the paradox wherein primary tumors and metastases share the same genetic mutations yet exhibit extremely different behaviors, and we anticipate that an evolving mechanism of tumor suppression is more broadly applicable to other commonly mutated driver genes and cancer types.

## Materials and Methods:

### Murine cell lines

Murine cell lines were generated from individual primary tumors and metastases from *Kras<sup>LSL-G12D</sup>;Trp53<sup>flox/flox</sup>;Lkb1<sup>XTR/XTR</sup>;Rosa26<sup>LSL-tdTomato</sup>* (cell lines LU1 and LU2), *Kras<sup>LSL-G12D</sup>;Trp53<sup>flox/flox</sup>;Lkb1<sup>XTR/XTR</sup>;Rosa26<sup>FlpOER/LSL-tdTomato</sup>* (cell line LR2), and *Kras<sup>LSL-G12D</sup>;Trp53<sup>flox/flox</sup>;Lkb1<sup>XTR/XTR</sup>;Rosa26<sup>FlpOER/+</sup>* (cell line LR1) mice previously transduced with lentiviral Cre. The *Lkb1<sup>XTR/XTR</sup>* mouse allele has been deposited at The Jackson Laboratory (034052) and was generated using the same design and methods as outlined for the *Trp53<sup>XTR/XTR</sup>* allele<sup>19</sup>. All cell lines have gene expression patterns consistent with being in a metastatic state (*Nkx2.1<sup>low</sup>*, *Hmga2<sup>high</sup>*) (Supplementary Table 3). To derive cell lines, tumors were excised from the lungs or lymph nodes of mice, minced into pieces using scissors, and directly cultured in DMEM media supplemented with 10% FBS, 1% penicillin-streptomycin-glutamate, and 0.1% amphotericin at 37°C with 5% CO<sub>2</sub> until cell line establishment. Cells were authenticated for genotype. All human cell lines tested negative for mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza).

All four murine cell lines (LR1, LR2, LU1, and LU2) were grown in DMEM media supplemented with 10% FBS, 1% penicillin-streptomycin-glutamate, and 0.1% amphotericin. LR1 and LR2 cell lines were then transduced with an SpCas9 lentiviral vector with a Blasticidin selection marker (Addgene #52962) and selected with Blasticidin (10ug/mL) for >5 days. To be able to test Cas9 cutting efficiency, site-directed mutagenesis was used to delete a loxP site in the pMCB306 backbone (Addgene #89360), since these cell lines were previously transduced with Cre recombinase to initiate tumor growth in mice. This plasmid is a self-GFP cutting reporter with both expression of GFP and a sgRNA against GFP on the same backbone. Polyclonal Cas9+ populations with high cutting efficiency were established and used for subsequent experiments (referred to as LR1;Cas9 and LR2;Cas9 in the text). For LKB1 restoration induction, cells were treated with either



1 $\mu$ M 4-hydroxytamoxifen (4-OHT; Sigma Aldrich) dissolved in 100% ethanol or a vehicle (1:2000 100% ethanol) for the indicated time-points.

### **Proliferation doubling assays**

For population doubling assays, cell lines were treated with 4-OHT or vehicle for twelve days. Every other day, cells were trypsinized, counted, and re-plated with 50,000 cells per well of a 6-well in triplicate. The number of population doublings was assessed by taking the total number of cells (N) for that day and normalizing to the original 50,000 cells plated i.e.  $\log_2(N/50000)$ . Two-tailed t-tests were performed to determine statistical significance.

### **Clonogenic growth assays**

For clonogenic growth assays, cell lines were pre-treated with a vehicle control or 4-OHT for six days. Cells were trypsinized, counted, and re-plated at 500 cells/well of a 6-well plate in triplicate. Plates were incubated at 37°C with 5% CO<sub>2</sub> for six days. For analysis, cells were rinsed with room temperature PBS, fixed with ethanol for 5 minutes at room temperature, and stained with 1% crystal violet solution in water (Millipore-Sigma) for an additional 5 minutes. Plates were rinsed with water, scanned into the computer, and analyzed using ImageJ. The % area of the plate covered by cells was normalized to the average % area of the plate covered by cells treated with a vehicle control. Two-tailed t-tests were performed to determine statistical significance.

### **RNA-sequencing library preparation for cell lines**

Cell lines were treated with 4-OHT or vehicle for six days, rinsed with PBS, trypsinized, spun down, and cell pellets were frozen at -80°C. Cell pellets were processed to total RNA using the RNeasy Plus Mini Kit (Qiagen). RNA quality was assessed using the Bioanalyzer 2100 (Agilent). All samples had an RNA integrity number (RIN) of 10.0. 500ng total RNA for each sample was processed into libraries using the TruSeq RNA Library Prep Kit v2 (Illumina) and sequenced according to standard protocols.

### **RNA-sequencing data processing and alignment**

RNA-seq data was trimmed with CutAdapt and aligned with kallisto<sup>37</sup>. We downloaded pre-compiled transcriptome indices from <https://github.com/pachterlab/kallisto-transcriptome-indices/releases> for mm10 and hg38. We aligned with kallisto quant using the following parameters: “kallisto quant -genomebam -gtf -chromosomes -threads -index”. This generated a transcript count file that was converted to gene counts using tximport. We then created a SummarizedExperiment in R containing a matrix of the samples by genes with the gene coordinates. We used the genomebam created by kallisto to validate the number of reads per exon in LKB1 (the trapped configuration of the XTR allele causes early termination of transcription after exon 1, Extended Data Fig. 1e). All gene expression matrices (count and tpm) are made available in Supplementary Table 3 and Supplementary Table 6.



## RNA-seq data analysis – Differential Expression

To compute differential gene expression, we used edgeR's glmQLFTest. We used as input two groups with a simple design with a 0 intercept “~0 + Group”. We first calculated normFactors using the TMM normalization “calcNormFactors(y, method = “TMM”)”. Next, we estimated dispersions with robustness “estimateDisp(y, design = design, robust = TRUE)”. Then we fitted the generalized linear model using “glmQLFit(y, design = design)”. Lastly, we used the glmQLFTest to compute log<sub>2</sub> fold changes and adjusted p-values. We chose the indicated significance cutoffs based on the thresholds set by our control LKB1-unrestorable cell lines (LU1 and LU2) treated with 4-OHT.

## Immunoblot analysis

Adherent cells were rinsed with ice-cold PBS, lysed in RIPA buffer, scraped from plates, and spun at 13,000g for 30 minutes at 4°C. The concentration of protein-containing supernatant was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher). 10ug of each sample was loaded onto NuPage 4–12% Bis-Tris protein gels (Thermo Fisher) and transferred to polyvinylidene fluoride (PVDF) membranes (Bio-Rad) at 10V overnight. Blocking, primary, and secondary incubations were performed in Tris-buffered saline (TBS) with 0.1% Tween-20. Blocking was performed in 5% dry milk and primary antibody incubation was performed in 5% bovine serum albumin (BSA) (Cell Signaling). Secondary antibody incubation was performed in 5% dry milk with anti-rabbit (Cell Signaling, 7074S, 1:2000 dilution) or anti-mouse (Santa Cruz Biotechnology, sc-2005, 1:2000 dilution) antibodies. LKB1 (Cell signaling, 13031S, 1:1000 dilution) and SOX17 (Abcam ab224637, 1:500 dilution) protein expression was assessed by Western blotting. HSP90 (BD Biosciences, 610418, 1:2000 dilution) was used as a sample processing control on a separate blot that was processed in parallel with the same input master mix.

## Allograft studies in immunocompromised mice

The use of mice for the current study has been approved by and was compliant with the guidelines set by the Institutional Animal Care and Use Committee at Stanford University, protocol number 26696. All transplant studies were performed in ten- to twelve-week-old immunocompromised NSG mice. For intravenous transplants, cells were treated with either a vehicle control or 4-OHT for six days and  $5 \times 10^4$  cells were injected into one of the lateral tail veins. Mice were sacrificed 21 days post-injection (n=4 male NSG mice per condition for Extended Data Fig. 1g, or n = 24 mice total; n=3 male NSG mice for sgSafe + vehicle and n=4 male NSG mice for all other conditions for Fig. 6a–c, or n=23 mice total). For intrasplenic transplants, cells were treated with 4-OHT for six days and  $5 \times 10^4$  cells were injected via intrasplenic injection. To perform intrasplenic injections, the left flank of each mouse was shaved and disinfected with 70% ethanol. A small incision was made to expose the spleen and a ligation on the splenic branch of the lienopancreatic artery was performed. Following injection of cells, a surgical knot was made in the upper part of the spleen and the lower part of the spleen was removed prior to sewing the body wall back with surgical knots. The skin incision was closed with staples and antiseptic solution was applied to clean the wound. Mice were sacrificed three weeks post-injection (n=7 female and 2 male NSG mice for BFP and n=5 female and 3 male NSG mice for Sox17 for Extended Data Fig.

10j-1, or n=17 mice total). For the initial subcutaneous transplants,  $2 \times 10^5$  untreated cells were re-suspended in 200uL PBS and injected into two sites per mouse (n=2-5 female NSG mice per condition for Extended Data Fig. 1f and Source Data Extended Data Fig. 1f, n=25 mice total). Once tumors were readily palpable, mice were randomized and treated via oral gavage with either a vehicle control (200uL 10% ethanol 90% corn oil) or tamoxifen (200uL of 20 mg ml<sup>-1</sup> tamoxifen dissolved in 10% ethanol 90% corn oil) (Sigma Aldrich) for three consecutive days. The height, width, and length, of each tumor was measured using calipers every two days for 14 days (LU1, LR2) or every four days for 16 days (LU2, LR1). Tumor volume was roughly calculated by multiplying height  $\times$  width  $\times$  length of each tumor. During the experimental time-course, tumor burden never exceeded 1.7cm<sup>3</sup> per mouse, which is the maximal tumor burden allowed by our ethics committee. For subcutaneous transplants to model metastatic spread to the lung,  $5 \times 10^4$  cells of the indicated genotypes pre-treated with six days of 4-OHT or vehicle were re-suspended in 200uL PBS and injected into two sites per mouse. Mice were sacrificed five weeks post-injection (n=4 female NSG mice per condition for Fig. 6d-f, n=12 mice total).

### Immunohistochemistry and histological quantification

Lung samples were fixed in 4% formalin and paraffin embedded. Hematoxylin and eosin staining was performed using standard methods and percent tumor area was calculated using ImageJ. For IHC, we used an antibody to SOX17 (Abcam, ab224637) at a 1:1000 dilution. Heat-mediated antigen retrieval was performed in Tris/EDTA buffer with pH 9.0. To evaluate SOX17 expression, we quantified the number of tumors with tumor area composed of 0% SOX17+ cells (none), <25% SOX17+ cells (low), 25-50% SOX17+ cells (medium), and >50% SOX17+ cells (high) using ImageJ.

### Lentiviral production

Lentiviruses were produced by co-transfecting lentiviral backbones with packaging vectors (delta8.2 and VSV-G) into 293T cells using PEI (Polysciences). The viral-containing supernatant was collected at 48- and 72-hours post-transfection, filtered through a 0.45uM filter, and combined with fresh media to transduce cells for up to two days. Human cell lines were incubated with 8ug/mL polybrene (Sigma) to enhance transduction efficiency.

### CRISPR/Cas9 screen and sample processing

The genome-scale CRISPR/Cas9 knock-out library was synthesized by Agilent and designed and cloned as previously described<sup>20</sup>. The genome-scale library was designed to have ~200,000 sgRNAs targeting ~20,000 coding genes (10 sgRNAs per gene), with >13,000 negative control sgRNAs that are either non-targeting (sgNT) or safe-targeting (sgSafe) (Supplementary Table 1). This library is composed of ten sub-library pools roughly divided according to gene function (<https://www.addgene.org/pooled-library/bassik-mouse-crispr-knockout/>). The entire genome-scale screen was performed in two halves, each composed of five sub-library pools. In addition, the second half of the screen included a repeat of the sub-library containing sgRNAs targeting *Lkb1* as a positive control. The two screens were performed sequentially.

For both halves of the screen, the combined sub-library plasmid pools were transfected into 293T cells to produce lentiviral pools, which were transduced into LR1;Cas9 cells. Cells were transduced at a multiplicity of infection of 0.3, and after 48 hours were selected with puromycin (8 ug/mL) for 3 days until the library-transduced population was >90% mCherry+ (a marker for lentivirus transduction). Cells were expanded for another 2 days and aliquots were saved as day 0 stocks. Remaining cells were plated and treated in duplicate with either vehicle or 4-OHT. The screens were performed at 200x cell number coverage per sgRNA. Due to the fast doubling time of this cell line, each half of the screen required passaging >165 15cm dishes every two days. 12 days later, cells were collected and stored in cryovials in liquid nitrogen for further processing. Genomic DNA was extracted from each sample in technical duplicate with the Qiagen Blood Maxi Kit (Qiagen). sgRNA cassettes were PCR-amplified from genomic DNA and sample indices, sequencing adapters, and flow-cell adapters were added in two sequential rounds of PCR as previously described<sup>20</sup>.

### CRISPR/Cas9 screen data alignment and analysis

We aligned each half of the genome-scale CRISPR/Cas9 screen individually using casTLE<sup>38</sup>, which uses bowtie alignment. This alignment returned a counts matrix for each sgRNA per sample. We then identified the sgRNAs that were overlapping in each half of the CRISPR screen and then computed the mean reads in these sgRNAs. We then scaled each half of the screen such that the mean reads in overlapping sgRNA was identical. We then used the values from each half for sgRNAs specific to that half and overlapping sgRNAs the mean reads across both screen halves was used. We depth-normalized across all samples. We computed the log<sub>2</sub> correlations and plotted the Pearson correlation matrix in R. To quantify the sgRNAs that were enriched/disenriched in the screen we used MAGeCK<sup>39</sup>. Briefly, we used mageck test with parameters “-k counts.tsv -t day12\_LKB1\_Restored -c day12\_LKB1\_Unrestored”. We then accessed the MAGeCK RRA scores from gene\_summary.txt file and filtered targets with less than 5 sgRNAs assigned to each target. We then took the top 50 sgRNA and used them as input to PANTHER GO term enrichment. The aligned CRISPR screen matrix is available in Supplementary Table 1.

### ATAC-sequencing library preparation for cell lines

Cell lines were treated with 4-OHT or vehicle for the indicated time-points prior to transposition. For the ATAC-seq time-course (Extended Data Fig. 2), samples were treated in a reverse time-course such that transposition for all time-points occurred at the same time. The media for all cells was changed at each time-point to control for fluctuations in growth factors or other media contents between samples. For all experiments, adherent cells were rinsed with PBS, trypsinized for 5 minutes at 37°C, spun down, and re-suspended in PBS. 50,000 cells in technical duplicate were resuspended in 250uL PBS and centrifuged at 500rcf for 5 minutes at 4°C in a fixed-angle centrifuge. Pelleted cells were re-suspended in 50uL ATAC-seq resuspension buffer (RSB; 10mM Tris-HCl pH 7.4, 10mM NaCl, and 3mM MgCl<sub>2</sub> in ddH<sub>2</sub>O made fresh) containing 0.1% NP40, 0.1% Tween-20, and 0.01% digitonin according to the omni-ATAC-seq protocol<sup>23</sup>. After incubating on ice for 3 minutes, 1mL of ATAC-seq RSB containing 0.1% Tween-20 was added. Nuclei were centrifuged at 500rcf for 5 minutes at 4°C in a fixed-angle centrifuge, 900uL of the supernatant was taken off,

and then the nuclei were centrifuged for an additional 5 minutes under the same conditions. The remaining 200uL of supernatant was aspirated and nuclei were resuspended in 50uL of transposition mix (25uL 2X TD buffer (2mL 1M Tris-HCl pH 7.6, 1mL 1M MgCl<sub>2</sub>, 20mL DMF, and 77mL ddH<sub>2</sub>O aliquoted and stored at -20°C), 2.5uL transposase (100nM final), 16.5uL PBS, 0.5uL 1% digitonin, 0.5uL 10% Tween-20, and 5uL ddH<sub>2</sub>O). Transposition reactions were incubated at 37°C for 30 minutes with 1,000 r.p.m. shaking in a thermomixer and cleaned up using MinElute PCR purification columns (Qiagen). The transposed samples were then amplified to add sample indices and sequencing flow cell adapters and cleaned up with MinElute PCR purification columns (Qiagen), with a target concentration of 20uL at 4nM. Paired-end sequencing was performed on an Illumina NextSeq using 75-cycle kits.

### ATAC-seq data processing and alignment

Adaptor sequence trimming, mapping to the mouse (mm10) or human (hg38) reference genome using Bowtie2 and PCR duplicate removal using Picard Tools were performed. Aligned reads (BAM) mapping to “chrM” were also removed from downstream analysis. BAM files were subsequently corrected for the Tn5 offset (“+” stranded +4 bp, “-” stranded -5 bp) using Rsamtools “scanbam” and Genomic Ranges. These ATAC-seq fragments were then saved as R binarized object files (.rds) for further downstream analysis. Additional information on detailed ATAC-seq data analysis can be found in Supplemental Information. We have made available all raw sequencing, aligned fragments, and bigwigs for all ATAC-seq samples in Supplementary Table 2 through AWS. Additionally, all matrices (peak matrix and chromVAR) are available in Supplementary Table 2 through AWS.

### ATAC-seq data Analysis – TCGA LUAD Tumors

We downloaded the TCGA-ATAC-seq data from <https://gdc.cancer.gov/about-data/publications/ATACseq-AWG> for tumors with matched RNA. We then scored each tumor for being high (top 10%), medium (middle 80%) and low (bottom 10%) in LKB1 (STK11) expression (TPM). We additionally identified which tumors had a medium-high predicted mutation (VarScan2). We then unbiasedly identified the top 10,000 variable peaks and grouped them into 5 k-means clusters. We then plotted a heatmap of the scaled log<sub>2</sub> accessibility as described above. To test the enrichment of specific gene mutations in each chromatin sub-type (See Extended Data Fig. 4a), we computed the proportion of medium-high predicted mutation burden of the gene (VarScan2) and computed a binomial enrichment vs the mutation frequency of all TCGA LUAD tumors (n = 230). We then computed the FDR for the binomial enrichments with p.adjust in R.

### scRNA-seq Analysis – LUAD Metastases Laughney et al. 2020

We downloaded the raw data from Laughney et al. 2020 from [https://s3.amazonaws.com/dp-lab-data-public/lung-development-cancer-progression/PATIENT\\_LUNG\\_ADENOCARCINOMA\\_ANNOTATED.h5](https://s3.amazonaws.com/dp-lab-data-public/lung-development-cancer-progression/PATIENT_LUNG_ADENOCARCINOMA_ANNOTATED.h5). We then read the subgroup (hdf5 formatted file) “INDF\_EPITHELIAL\_NOR\_TUMOR\_MET” for the normalized scRNA-seq matrix. We then computed z-scores for all genes. We then averaged the scaled expression for all cells from each donor that belonged in cluster “H0” and “H3” (to increase number of donors) which represent the most undifferentiated metastatic cells. We then computed the standard error of the mean (SEM) for all cells from each donor in these

clusters. We then plotted the average and SEM SOX17 expression vs STK11 (LKB1) expression for each of the donors.

### Cloning and generating knock-out and overexpression cell lines

To generate individual knock-out cell lines, we first cloned individual sgRNAs into the pMJ114 backbone (Addgene #85995) using Q5 site-directed mutagenesis (NEB). A list of all sgRNA sequences used in this study is located in Supplementary Table 4. sgRNA sequences were chosen based on the most highly enriched sgRNAs in the genome-scale screen (sgLkb1) or by choosing the top two sgRNAs with the highest predicted cutting activity from the Brie library on Addgene (#73633). After making lentivirus and transducing cells with the lentiviral supernatant, we waited two days and then selected cells with 8ug/mL puromycin for at least three days to enrich for cells transduced with the lentivirus, before initiating treatment with vehicle or 4-OHT.

To generate double and triple knock-out cell lines, we used Gibson assembly to create lentiviral vectors with sgRNAs transcribed in series by the bovine U6 promoter, human U6 promoter, and murine U6 promoter, as previously described<sup>40</sup>. In brief, we first cloned individual sgRNAs into the pMJ114 (Addgene #85995), pMJ117 (Addgene #85997), and pMJ179 (Addgene #85996) backbones, then stitched them together using Gibson assembly (NEB). For LKB1 downstream effector families with only two gene paralogs, we still included the third murine U6 promoter driving expression of sgSafe-1 to control for the effects of three cutting events occurring simultaneously in the same cell. Similarly, for the sgLkb1 control experiments, a bovine U6 promoter driving expression of sgLkb1-1 was combined with a human and mouse U6 driving expression of sgSafe-1 and sgSafe-2. After transducing cells with the lentiviral supernatant, we waited two days and then selected cells with 8ug/mL puromycin for at least three days to enrich for cells transduced with the lentivirus, before initiating treatment with vehicle or 4-OHT.

To generate cell lines with overexpression of Sox17, we codon optimized murine Sox17 cDNA to simultaneously mutate the sgSox17-1 and sgSox17-2 cut sites and ordered this sequence as a gBlock (IDT). We used Gibson assembly to replace BFP in pMJ114 with this modified Sox17 sequence. After making lentivirus and transducing cells with the lentiviral supernatant, we waited two days and then selected murine cell lines with 8ug/mL puromycin for at least three days to enrich for cells transduced with the lentivirus before initiating treatment with vehicle or 4-OHT. To increase the *Sox17* knock-out efficiency of sgSox17-1 and sgSox17-2 cell lines, these cell lines were additionally FACS sorted to enrich for cells with the highest expression of BFP (the fluorescent reporter on the sgRNA backbone).

To generate human cell lines with overexpression of KEAP1 or LKB1, we amplified human KEAP1 and LKB1 off of human cDNA and used Gibson assembly to replace GFP in pMCB306 (Addgene #89360) with these sequences. After making lentivirus and transducing human cells with the lentiviral supernatant, we waited 2 days, selected human cell lines with 2ug/mL puromycin for at least 4 days to enrich for cells transduced with lentivirus, then let the cells recover in fresh media for 2 days before collecting for ATAC-seq library preparation.

## Human cell lines

All human non-small cell lung cancer cell lines (NCI-H1437 (ATCC CRL-5872), A549 (ATCC CCL-185), NCI-H460 (ATCC HTB-177), NCI-H1355 (ATCC CRL-5865), NCI-H1650 (ATCC CRL-5883), NCI-H1975 (ATCC CRL-5908), NCI-H358 (ATCC CRL-5807), NCI-H2009 (ATCC CRL-5911)) were either purchased directly from ATCC or were a gift from Dr. Michael Bassik's laboratory, who previously purchased them from ATCC. Human cell lines were cultured in RPMI media supplemented with 10% FBS, 1% penicillin-streptomycin-glutamate, and 0.1% amphotericin. All human cell lines tested negative for mycoplasma using the MycoAlert Mycoplasma Detection Kit (Lonza).

## Autochthonous mouse models

The use of mice for the current study has been approved by and was compliant with the guidelines set by the Institutional Animal Care and Use Committee at Stanford University, protocol number 26696. Homozygous floxed *Lkb1* alleles (*Lkb1<sup>fl/fl</sup>*) were bred into the metastatic *KPT* (*Kras<sup>LSL-G12D</sup>;p53<sup>fl/fl</sup>;Rosa26<sup>LSL-tdTomato</sup>*) model to generate *LKB1*-proficient and *LKB1*-deficient models of lung adenocarcinoma metastasis. Lentiviral Cre recombinase was co-transfected with packaging vectors (delta8.2 and VSV-G) into 293T cells using PEI, the supernatant was collected at 48 and 72 hours post-transfection, ultracentrifuged at 25,000g for 90 minutes, and resuspended in PBS. Tumors were initiated by intratracheal transduction of 10- to 12-week old mice with lentiviral vectors expressing Cre recombinase<sup>41</sup>. For ATAC-seq, tumors were collected and processed at staggered time-points where tumor burden was similar between *KPT* and *KPT;Lkb1<sup>fl/fl</sup>* cohorts of mice (n=11 female *KPT* mice, 3 male *KPT* mice, 2 female *KPT;Lkb1<sup>fl/fl</sup>*, and 6 male *KPT;Lkb1<sup>fl/fl</sup>* mice). For the survival curve, mice were sacrificed immediately after exhibiting physical symptoms of distress from lung tumor burden.

## Tumor dissociation, cell sorting, and freezing

Primary tumors and metastases were individually microdissected and dissociated using collagenase IV (Thermo Fisher), dispase (Corning), and trypsin (Invitrogen) at 37°C for 30 minutes. After dissociation, the samples remained on ice and in the presence of 2mM EDTA (Promega) and 1U/mL DNase (Sigma-Aldrich) to prevent aggregation. Cells were stained with antibodies to CD45 (30-F11), CD31 (390), F4/80 (BM8), and Ter119 (all from Biolegend) to exclude hematopoietic and endothelial lineages (lineage-positive (Lin+) cells). DAPI was used to exclude dead cells. BD FACSAria sorters (BD Biosciences) were used for cell sorting. tdTomato+, Lin-, DAPI- cells were FACS sorted into microcentrifuge tubes, spun down at 500rcf for 5 minutes at 4°C in a fixed-angle centrifuge, re-suspended in 250uL freezing media (Bambanker; Wako Chemicals USA), and left at -80°C overnight before being transferred to liquid nitrogen storage until bulk ATAC-seq and scATAC-seq library preparation.

## ATAC-sequencing library preparation for primary tumors and metastases

FACS-isolated samples were taken out of storage in liquid nitrogen, quickly thawed at 37°C, diluted with 1mL PBS, and centrifuged at 300rcf for 5 minutes at 4°C in a fixed-angle centrifuge. Primary tumors and metastases were then processed for ATAC-seq library



preparation using the same protocol used for cell lines, except the amount of transposase was decreased proportionally for samples with less than 50K cells. For example, for a sample with 10K cells, 1/5<sup>th</sup> the normal amount of transposase was used in the 50uL transposition reaction. The remaining volume was replaced with ddH<sub>2</sub>O.

### **RNA-sequencing library preparation for primary tumors and metastases**

RNA was extracted from sorted cancer cells using the AllPrep DNA/RNA Micro Kit (Qiagen). RNA quality of each tumor sample was assessed using the RNA6000 PicoAssay for the Bioanalyzer 2100 (Agilent) as per the manufacturer's instructions. All of the RNA used for RNA-seq had an RNA integrity number (RIN) > 8.0. RNA-sequencing libraries were generated as previously described<sup>41</sup> and sequenced using 200-cycle kits on an Illumina HiSeq 2000.

### **scATAC-sequencing library preparation for primary tumors and metastases**

FACS-isolated samples were taken out of storage in liquid nitrogen, quickly thawed at 37°C, diluted with 1mL PBS, and centrifuged at 300rcf for 5 minutes at 4°C in a fixed-angle centrifuge. Cells were re-suspended in PBS + 0.04% BSA, passed through a 40uM Flowmi cell strainer (Sigma) to minimize cell debris, and cell concentration was determined. Primary tumors and metastases were then processed for scATAC-seq library preparation according to standard droplet-based protocols (10x Genomics; Chromium Single Cell ATAC Library and Gel Bead Kit v1.0).

### **scATAC-seq data processing and alignment**

Raw sequencing data was converted to fastq format using cellranger atac mkfastq (10x Genomics, version 1.2.0). Single-cell ATAC-seq reads were aligned to the mm10 reference genome and quantified using cellranger count (10x Genomics, version 1.2.0). The current version of Cell Ranger can be accessed here: <https://support.10xgenomics.com/single-cell-atac/software/downloads/latest>. The 10x cell ranger atac output files and all scATAC-seq matrices used in this study are available in Supplementary Table 2 through AWS.

### **scATAC-seq data analysis – ArchR**

We used ArchR<sup>42</sup> for all downstream scATAC-seq analysis ([https://greenleaflab.github.io/ArchR\\_Website/](https://greenleaflab.github.io/ArchR_Website/)). We used the fragments files for each sample with their corresponding csv file with cell information. We then created Arrow files using “createArrowFiles” with using the barcodes from the sample 10x CSV file with “getValidBarcodes”. This step adds the accessible fragments a genome-wide 500-bp tile matrix and a gene-score matrix. We then added doublet scores for each single cell with “addDoubletScores” and then filtered with “filterDoublets”. Additionally, we then filtered cells that had a TSS enrichment below 6, less than 1,000 fragments or more than 50,000 fragments. We then reduced dimensionality with “addIterativeLSI” excluding chrX and chrY from this analysis. We then added clusters with “addClusters” with a resolution of 0.4. We then added a UMAP with “addUMAP” and minDist of 0.6. We identified 12 scATAC-seq clusters with this analysis. We then created a reproducible non-overlapping peak matrix with “addGroupCoverages” and “addReproduciblePeakSet”. We then quantified the number of



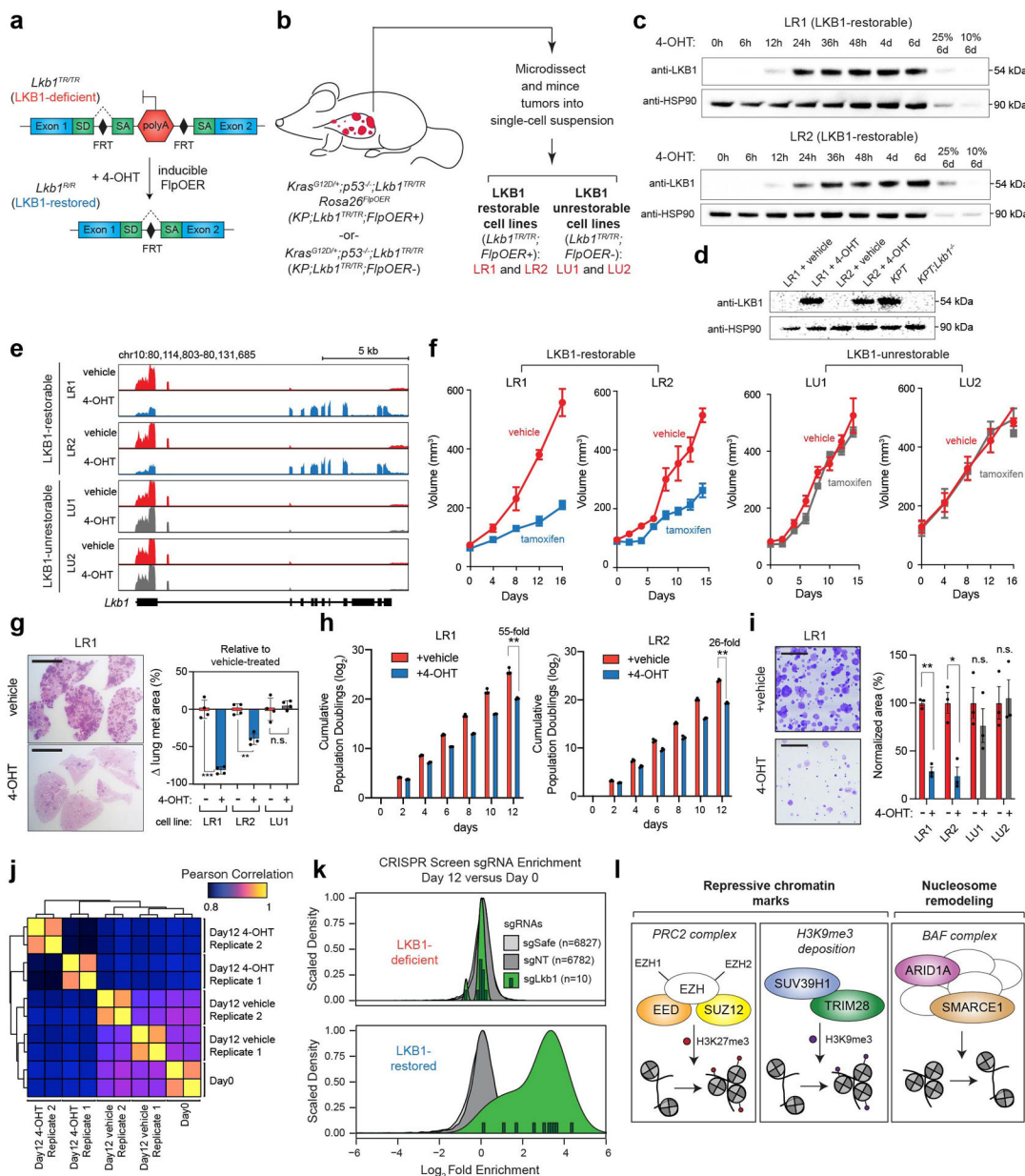
Tn5 insertions per peak per cell using “addPeakMatrix”. We then added motif annotations using “addMotifAnnotations” with chromVAR mouse motifs version 1 “mouse\_pwm\_v1”. We then computed chromVAR deviations for each single cell with “addDeviationsMatrix”. For TF footprinting of NKX2-1 and SOX17 we used “plotFootprints” with normalization method “subtract” which subtracts the Tn5 bias from the ATAC footprint.

To further characterize the 12 scATAC-seq clusters based on their metastatic state, we computed differential peaks from the LKB1-deficient bulk primary tumors and metastases. We took significantly differential peaks ( $|\log_2 \text{Fold Change}| > 3$  and  $\text{FDR} < 0.01$ ) and overlapped these peaks with the scATAC-seq peaks. The average accessibility and SEM across these overlapping regions was then plotted for peaks specific to primary tumors and peaks specific to metastases independently.

### Statistics and Reproducibility

Experimental data were plotted and analyzed using GraphPad Prism 9.0.1 (GraphPad Software) and R (3.6). Significance, where indicated, was calculated using an unpaired Student’s t-test. No statistical method was used to predetermine sample size. No data were excluded from the analyses. The experiments were not randomized. The Investigators were not formally blinded to allocation during experiments and outcome assessment.

Extended Data



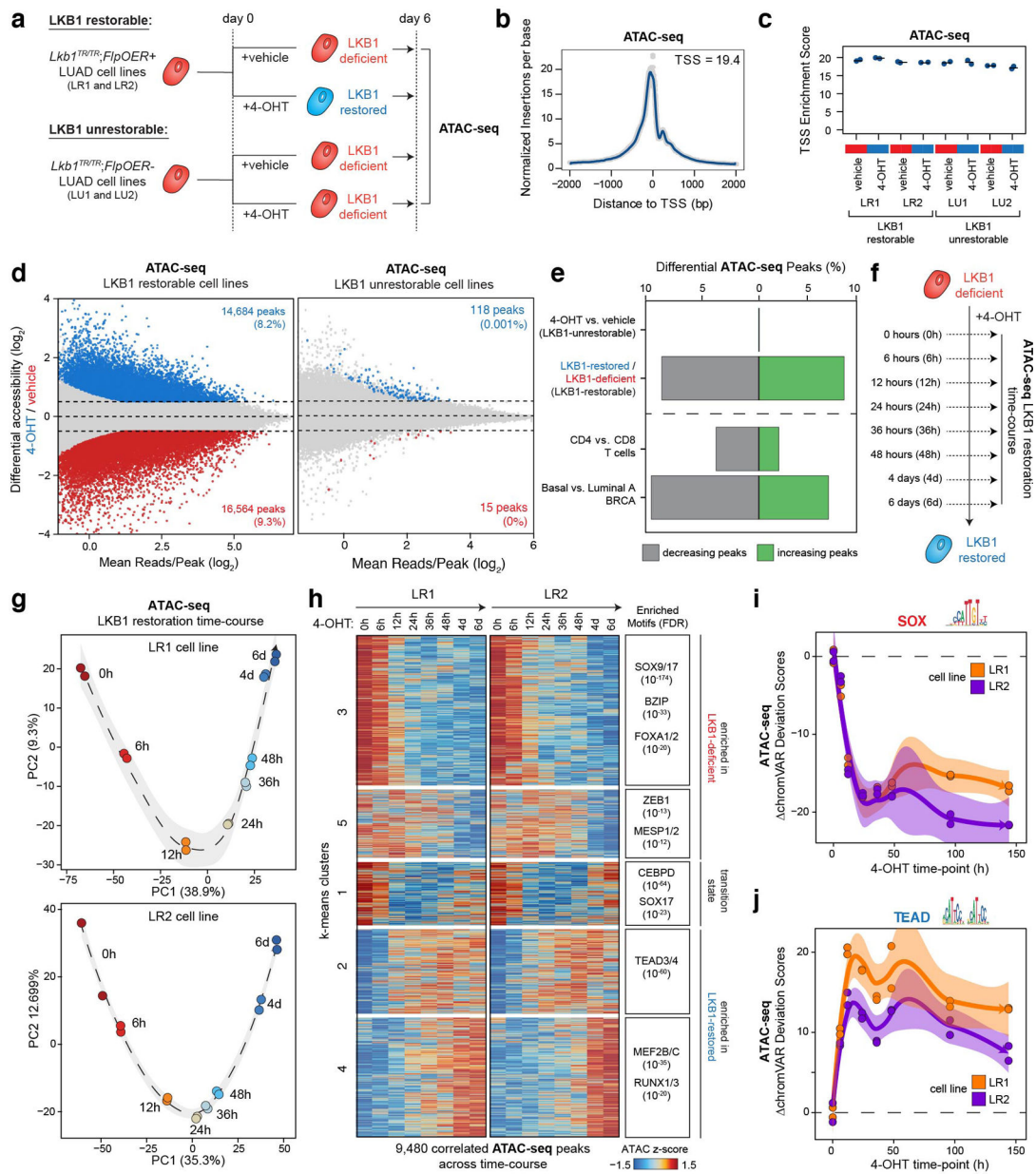
**Extended Data Fig. 1. Validation and quality control of inducible LKB1 restoration model and genome-scale CRISPR/Cas9 screen.**

a. Schematic of restorable *Lkb1*<sup>TR/TR</sup> alleles. SA = splice acceptor, SD = splice donor, FRT = flippase recognition target.

b. Schematic of the derivation of LKB1-restorable cell lines.

c. Expression of LKB1 by immunoblot over a time-course of 4-OHT treatment, represented in hours (h) and days (d). HSP90 is a sample processing control. 25% and 10% of input after six days of 4-OHT treatment is shown for a visual comparison.

- d. Expression of LKB1 by immunoblot in LR1 and LR2 cells treated with vehicle or 4-OHT compared to a KPT cell line and a *KPT;Lkb1*<sup>-/-</sup> cell line. HSP90 is a sample processing control.
- e. RNA-sequencing reads mapping to the *Lkb1* locus following six days of 4-OHT or vehicle treatment.
- f. Subcutaneous growth assay following injection of cell lines into recipient NSG mice. Tamoxifen or vehicle treatment was initiated on day 0. Mean tumor volume as measured by calipers of six tumors per condition  $\pm$  SEM is shown.
- g. Intravenous (i.v.) transplant assays. Left: Representative lung histology. Right: Change in % tumor area in LKB1-restored cells. Mean area of four mice per condition  $\pm$  SEM is shown. \*\*p = 0.001, \*\*\*p = 0.0001, n.s. = not significant with a two-sided t-test. Scale bars represent 5mm.
- h. Cumulative population doublings recorded over 12 days of 4-OHT treatment. Each cell line and condition was cultured and analyzed in triplicate. Mean  $\pm$  SEM is shown. \*\*p = 0.0002 for LR1, \*\*\*p = 0.0001 for LR2.
- i. Left: Representative image of clonogenic growth in LR1 cells. Right: % normalized area of cell growth. Each treatment group was cultured and analyzed in triplicate. Mean  $\pm$  SEM is shown. \*p < 0.01, \*\*p < 0.001, n.s. = not significant with a two-sided t-test. Scale bars represent 10mm. p = 0.0001 for LR1 and p = 0.0059 for LR2.
- j. Heatmap of Pearson correlation matrix of log-normalized counts across all samples in the genome-scale CRISPR/Cas9 screen.
- k. Log<sub>2</sub> fold enrichment of negative control sgRNAs and sgRNAs targeting *Lkb1* at day 12 versus day 0.



**Extended Data Fig. 2. LKB1 restoration drives widespread changes in chromatin accessibility in lung adenocarcinoma cells.**

- Schematic of preparing LKB1-deficient and LKB1-restored samples prior to ATAC-seq library preparation. Cell lines were treated with 4-OHT or vehicle for six days.
- Representative plot of aggregate signal around the transcription start site (TSS) for all ATAC-seq peaks in one vehicle-treated, LR1 replicate. This plot represents the signal-to-noise quantification of our ATAC-seq data. TSS enrichment scores greater than 10 indicate high quality ATAC-seq data.
- TSS enrichment scores for 16 ATAC-seq libraries with technical replicates.
- Differential accessibility across 178,783 ATAC-seq peaks following 4-OHT treatment in the LKB1-restorable (LR1 and LR2) and LKB1-unrestorable (LU1 and LU2) cell lines. The x-axis represents the  $\log_2$  mean accessibility per peak and the y-axis represents the  $\log_2$  fold

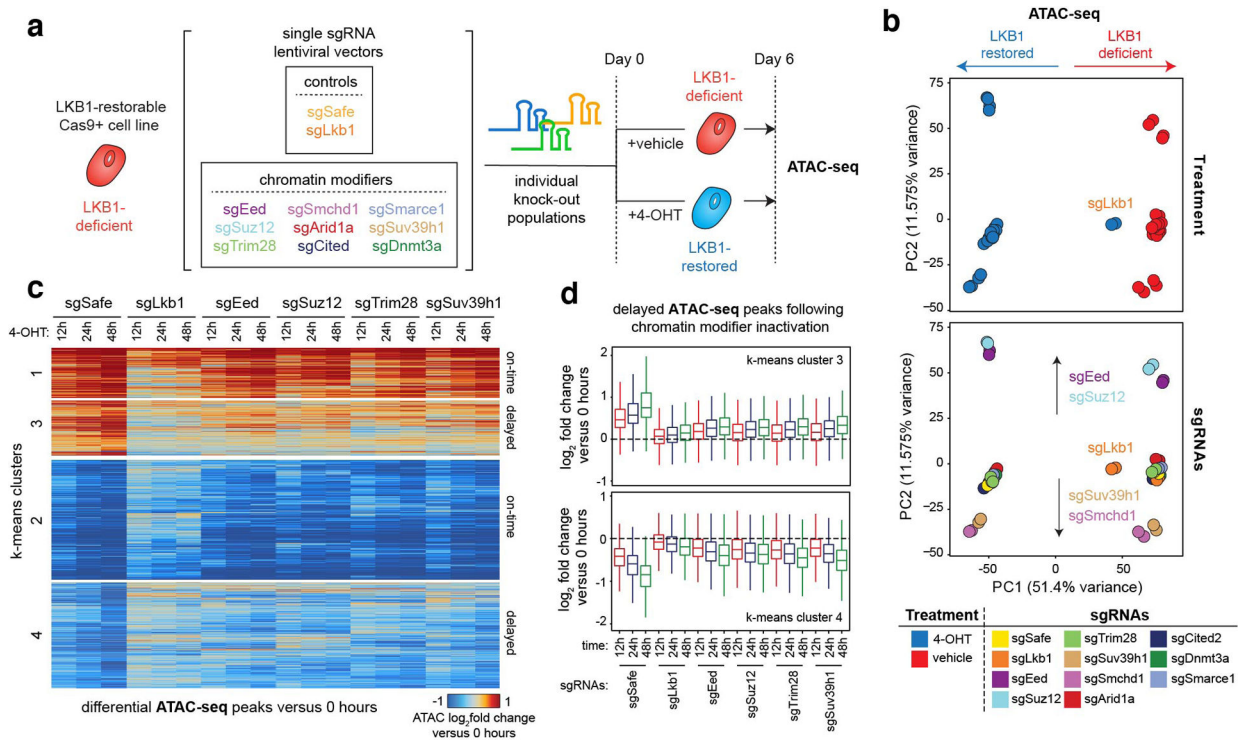
change in accessibility following 4-OHT treatment. Colored points are significant ( $|\log_2$  fold change $>0.5$ , FDR  $<0.05$ ).

e. Percentage of differential peaks ( $|\log_2$  fold change $>0.5$ , FDR  $<0.05$ ) across multiple ATAC-seq comparisons.

f. Schematic of preparing samples for LKB1-restoration time-course. Cell lines were treated with 4-OHT for eight different time-points (0 hours, 6 hours, 12 hours, 24 hours, 36 hours, 48 hours, 4 days, and 6 days) in two cell lines (LR1 and LR2).

g and h. PCA (g) and k-means clustering (h) of 9,480 correlated, variable ATAC-seq peaks across the LKB1 restoration time-course in two cell lines (LR1 and LR2). Each row of the heatmap represents a z-score of  $\log_2$  normalized accessibility across all samples within each cell line.

i and j. SOX (i) and TEAD (j) motif accessibility changes ( chromVAR deviation scores) across time in two cell lines (LR1 and LR2) treated with 4-OHT for the indicated time-points. Shaded area represents 95th percent confidence interval.



**Extended Data Fig. 3. Inactivating chromatin modifiers only delays LKB1-induced chromatin changes.**

a. Schematic of generating single knock-out populations of chromatin modifiers identified in the CRISPR screen, treating with 4-OHT or vehicle for six days, and processing for ATAC-seq.

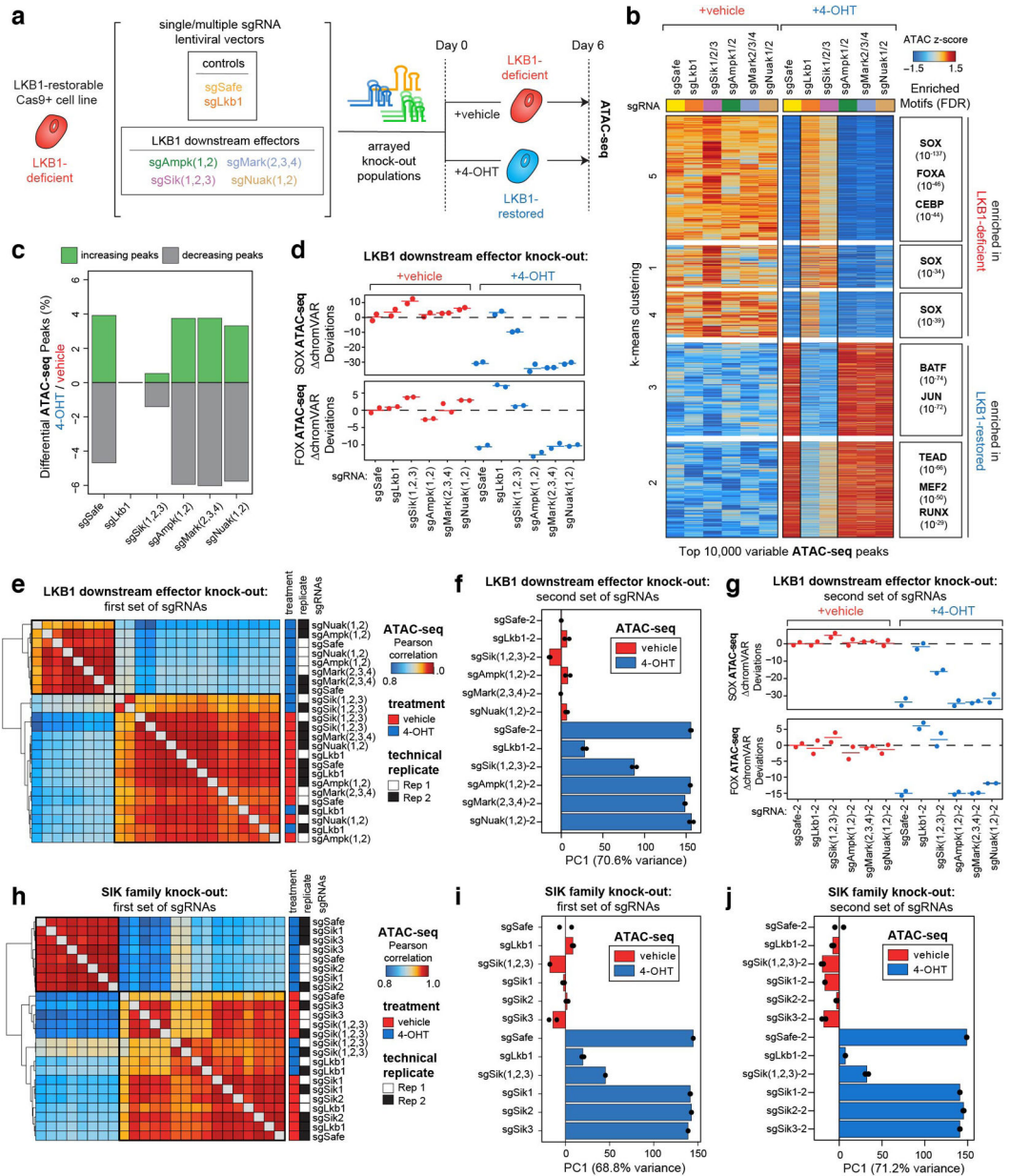
b. Principle component analysis (PCA) of the top 10,000 variable ATAC-seq peaks across the indicated LR1;Cas9 knock-out populations treated with 4-OHT or vehicle.

c. K-means clustered heatmap of differential peak accessibility ( $\log_2$  fold change) for each genotype of LR1;Cas9 cells treated with 4-OHT for up to 48 hours compared to 0 hours. All peaks differential between sgSafe (0 hours 4-OHT) and sgSafe (48 hours 4-OHT) are shown.



Each row represents the log<sub>2</sub> fold change of each genotype and time-point versus the same genotype's initial time-point (day 0).

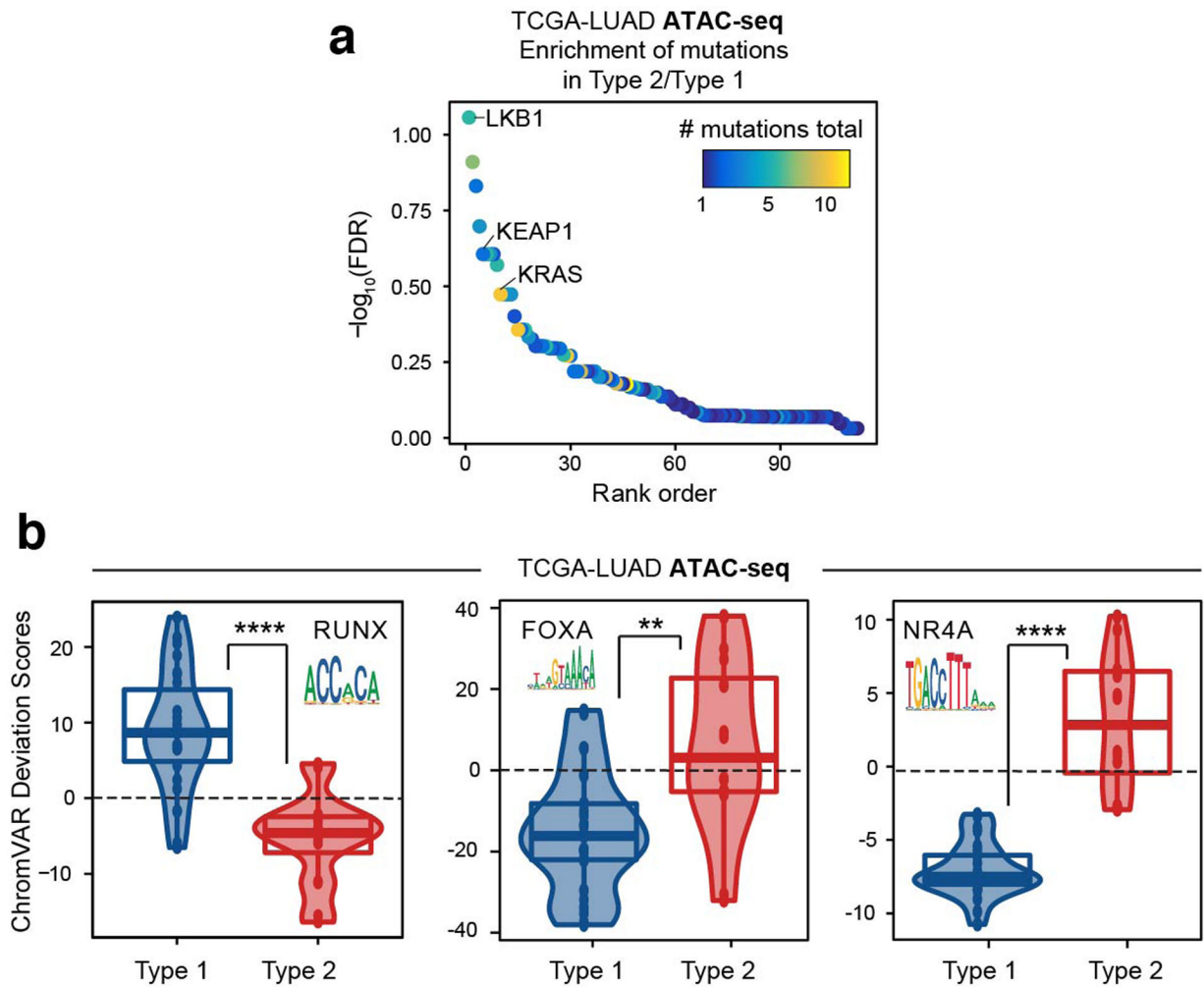
d. Log<sub>2</sub> fold change in mean peak accessibility for all peaks in k-means cluster 3 (top) and cluster 4 (bottom) from (c) for the indicated genotype and 4-OHT time-points compared to 0 hours 4-OHT. N=2 technical replicates per sgRNA population and time-point. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.



Extended Data Fig. 4. SIK family members mediate LKB1-induced chromatin changes.

- a. Schematic of generating single and multiple sgRNA knock-out cell lines and processing for ATAC-seq. LR1;Cas9 cells were treated with 4-OHT or vehicle for six days.
- b. Left: Heatmap of peak accessibility between each knock-out population treated with 4-OHT or vehicle. Each row represents a z-score of  $\log_2$  normalized accessibility across all samples. Right: Transcription factor hypergeometric motif enrichment in each k-means cluster.
- c. Percent of differential ATAC-seq peaks ( $|\log_2$  fold change $>0.5$ , FDR  $<0.05$ ) across LKB1-restorable cells treated with 4-OHT or vehicle.
- d. SOX (top) and FOXA (bottom) motif accessibility changes ( chromVAR deviation scores normalized to vehicle-treated sgSafe) across LKB1-restorable knock-out populations treated with 4-OHT or vehicle.
- e. Heatmap of Pearson correlation matrix of  $\log_2$ -normalized accessibility (in counts per million (CPM)) across LKB1 downstream effector knock-out genotypes with and without LKB1 restoration in LR1;Cas9 cells.
- f. PCA of the top 10,000 variable ATAC-seq peaks across LR1;Cas9 knock-out populations treated with 4-OHT or vehicle. Principle components besides PC1 (70.6%) account for  $<4\%$  of the variance in the dataset. N=2 technical replicates per sgRNA population.
- g. SOX (top) and FOXA (bottom) motif accessibility changes ( chromVAR deviation scores normalized to vehicle-treated sgSafe) across LKB1-restorable knock-out populations treated with 4-OHT or vehicle. Line represents average between two technical replicates.
- h. Heatmap of Pearson correlation matrix of  $\log_2$ -normalized accessibility (in counts per million (CPM)) across LKB1 downstream effector knock-out genotypes with and without LKB1 restoration in LR1;Cas9 cells.
- i and j. PCA of the top 10,000 variable ATAC-seq peaks across LR1;Cas9 knock-out populations treated with 4-OHT or vehicle. Principle components besides PC1 account for  $<4\%$  of the variance in the dataset. N=2 technical replicates per sgRNA population.

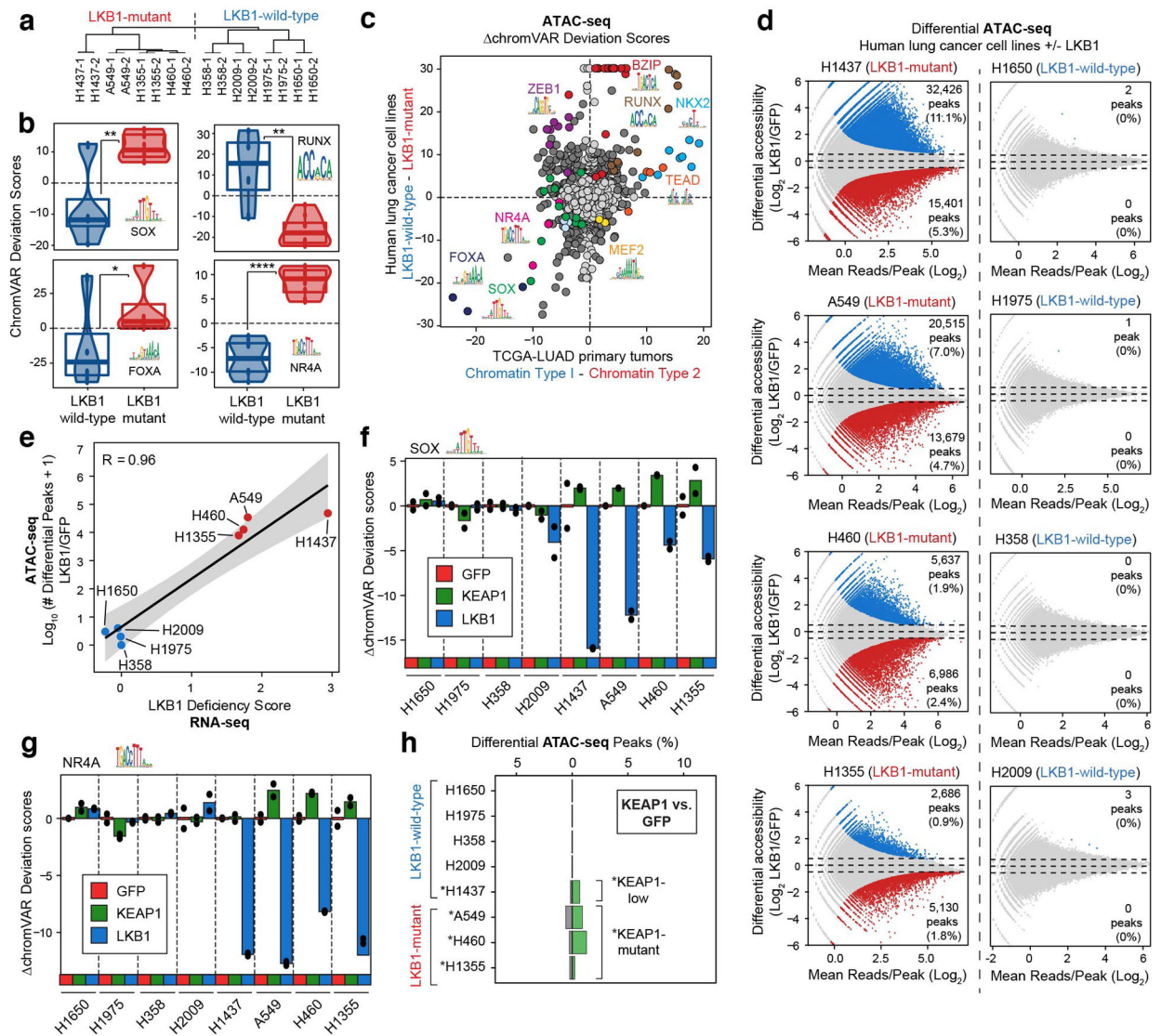




**Extended Data Fig. 5. Loss of LKB1 partitions human lung adenocarcinoma primary tumors into two chromatin accessibility sub-types.**

a. Enrichment of mutations in Chromatin Type 2 tumors compared to Chromatin Type 1 tumors. Genes are ranked according to  $-\log_{10}(\text{FDR})$ , with Rank 1 (LKB1) being the most significant (see Methods), as indicated on the y-axis. Points are colored by the number of mutations in the TCGA-LUAD ATAC-seq dataset (out of 21 samples).

b. ChromVAR deviation scores for the indicated transcription factor motifs for samples in the TCGA-LUAD ATAC-seq dataset. \* $p < 0.1$ , \*\* $p < 0.005$ , \*\*\*\* $p < 10^{-6}$  using a two-sided t-test.  $p=1 \times 10^{-7}$  for RUNX,  $p=0.002$  for FOXA, and  $p=1 \times 10^{-7}$ .  $N=13$  biologically independent samples for Chromatin Type 1 and 8 biologically independent samples for Chromatin Type 2. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.



**Extended Data Fig. 6. Loss of LKB1 drives a unique chromatin accessibility state in human lung adenocarcinoma cell lines.**

a. Hierarchical clustering of human lung cancer cell lines using the Euclidian distance within the first three principle components from Fig. 2d.

b. ChromVAR deviation scores for the indicated transcription factor motifs in eight human lung cancer cell lines at baseline. \* $p < 0.1$ , \*\* $p < 0.005$ , \*\*\*\* $p < 10^{-6}$  using a two-sided t-test.  $p = 0.066$  for FOXA,  $p = 0.003$  for SOX,  $p = 3.1 \times 10^{-7}$  for NR4A, and  $p = 0.001$  for RUNX.  $N = 4$  biologically independent samples for each group. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.

c. Comparison of the changes in motif accessibility (ChromVAR deviation scores) across LKB1-wild-type and LKB1-mutant human lung cancer cell lines (y-axis) and Chromatin Type 1 and Type 2 tumors (x-axis). Dark grey or colored points are called significantly

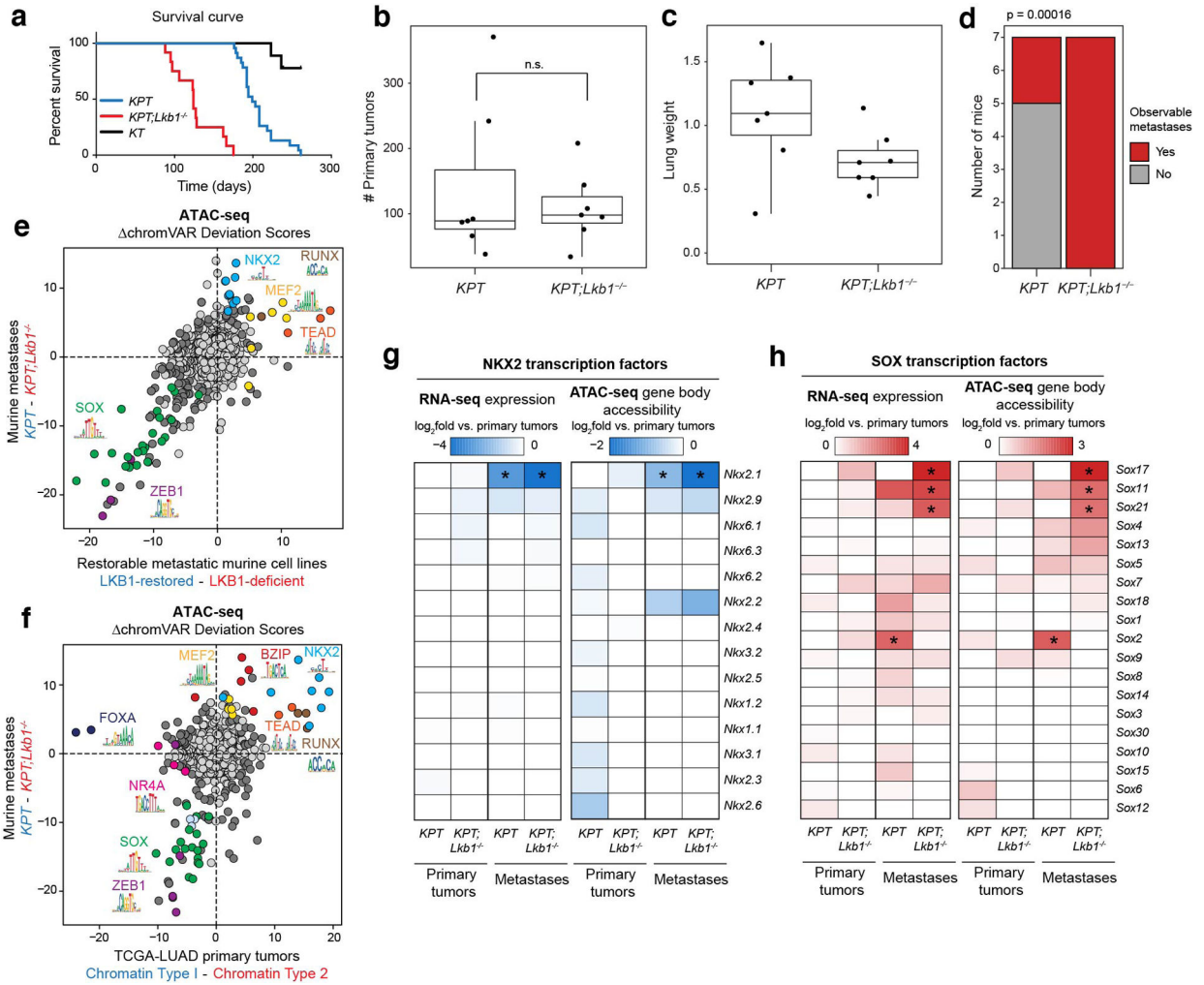
different ( $q < 0.05$ ) across both comparisons. Light grey points are not significant. A selection of motif families and their associated motif logos are indicated.

d. Differential accessibility across ATAC-seq peaks following LKB1 wild-type expression in eight human lung cancer cell lines. The x-axis represents the  $\log_2$  fold change in accessibility following LKB1 restoration. LKB1-mutant and LKB1-wild-type status at baseline is indicated. Colored points are significant ( $|\log_2 \text{fold change}| > 0.5$ ,  $\text{FDR} < 0.05$ ).

e. LKB1-deficiency score by RNA-seq (using 16-gene signature from Kaufmann et al., 2017) compared to  $\log_{10}(\text{number of differential ATAC-seq peaks} + 1)$  following LKB1 expression in each indicated cell line. Pearson correlation indicated in top left. Shaded area represents 95th percent confidence interval.

f and g. Relative chromVAR deviation scores for SOX (f) and NR4A (g) motifs in the indicated cell lines transduced with GFP, LKB1, or KEAP1. Scores are normalized based on the GFP control for each cell line.  $N=2$  technical replicates per cell line and overexpression condition.

h. Percent of differential ATAC-seq peaks ( $|\log_2 \text{fold change}| > 0.5$ ,  $\text{FDR} < 0.05$ ) in cells transduced to express KEAP1 compared to GFP.



**Extended Data Fig. 7. Genotype-specific activation of SOX17 in LKB1-deficient metastatic cells.**

a. Percent survival of *KPT* and *KPT;Lkb1<sup>-/-</sup>* mice compared to *KT* mice.

b and c. Number of primary tumors observed in *KPT* and *KPT;Lkb1<sup>-/-</sup>* mice (b). Lung weights of *KPT* and *KPT;Lkb1<sup>-/-</sup>* mice (c). N=7 biologically independent mice for each genotype. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR. n.s. = non-significant with a two-sided t-test.

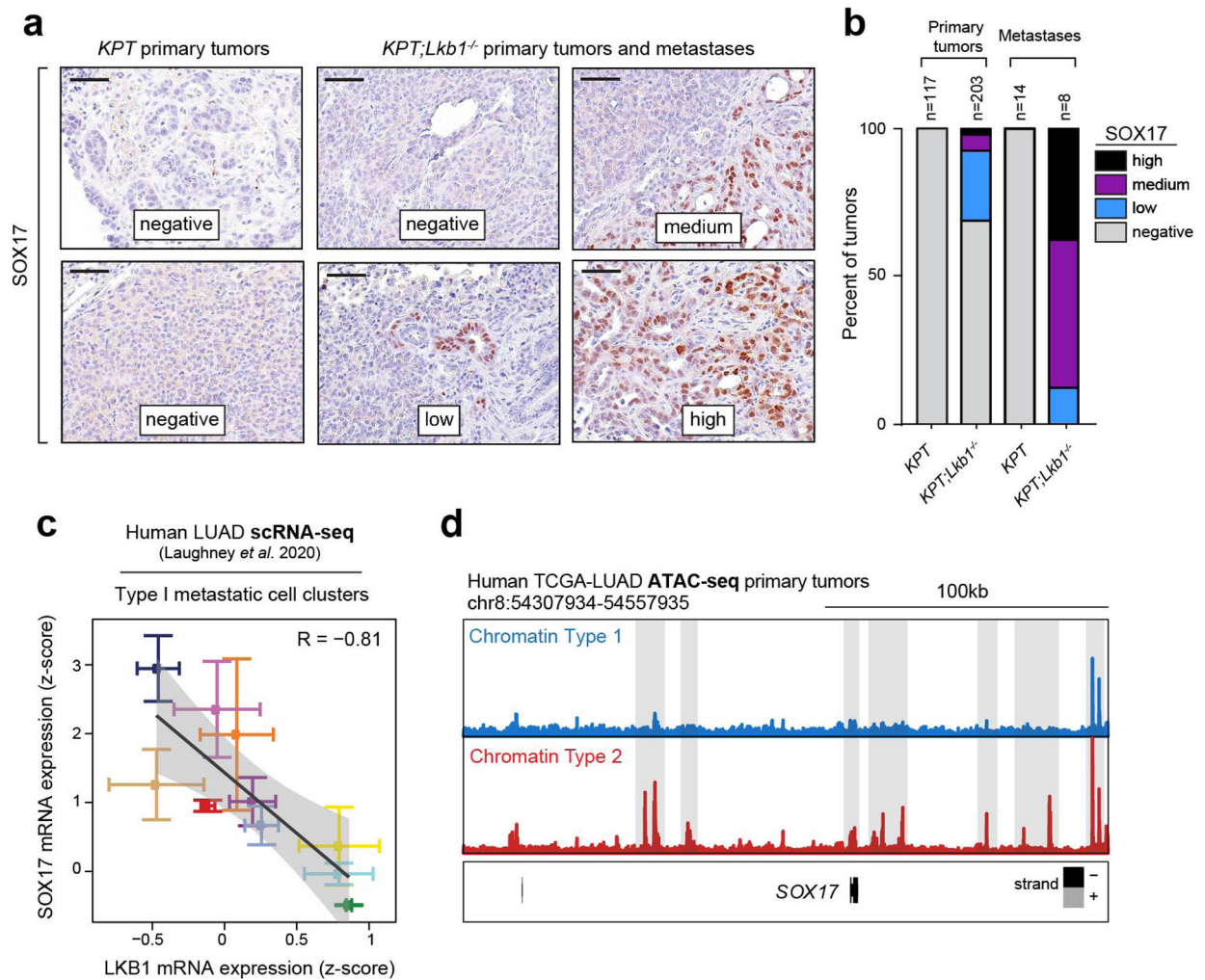
d. Metastatic rates of *KPT* (2/7 mice with metastases) and *KPT;Lkb1<sup>-/-</sup>* (7/7 mice with metastases). p-value = 0.00016 with a one-sided binomial test.

e and f. Comparison of the changes in motif accessibility (chromVAR deviation scores) between murine LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1<sup>-/-</sup>*) metastases (y-axis) and between murine LKB1-restored and LKB1-deficient cells (x-axis; e) or Chromatin Type 1 tumors and Chromatin Type 2 tumors (x-axis; f). Dark grey or colored points are called significantly different ( $q < 0.05$ ) across both comparisons. Light grey points are not significant. A selection of motif families and their associated motif logos are indicated.

g.  $\log_2$  fold change in mRNA expression (left) and accessibility within the gene body (right) of each NKX2 transcription factor compared to the average expression and accessibility in primary tumor samples. Asterisks indicate transcription factors with greater than  $\log_2$  fold change of -1 in both RNA and ATAC measurements.

h.  $\log_2$  fold change in mRNA expression (left) and accessibility within the gene body (right) of each SOX transcription factor compared to the average expression and accessibility in primary tumor samples. Asterisks indicate transcription factors with greater than  $\log_2$  fold change of 2 in both RNA and ATAC measurements.





**Extended Data Fig. 8. LKB1-deficient primary tumors harbor sub-populations of SOX17+ cells.**

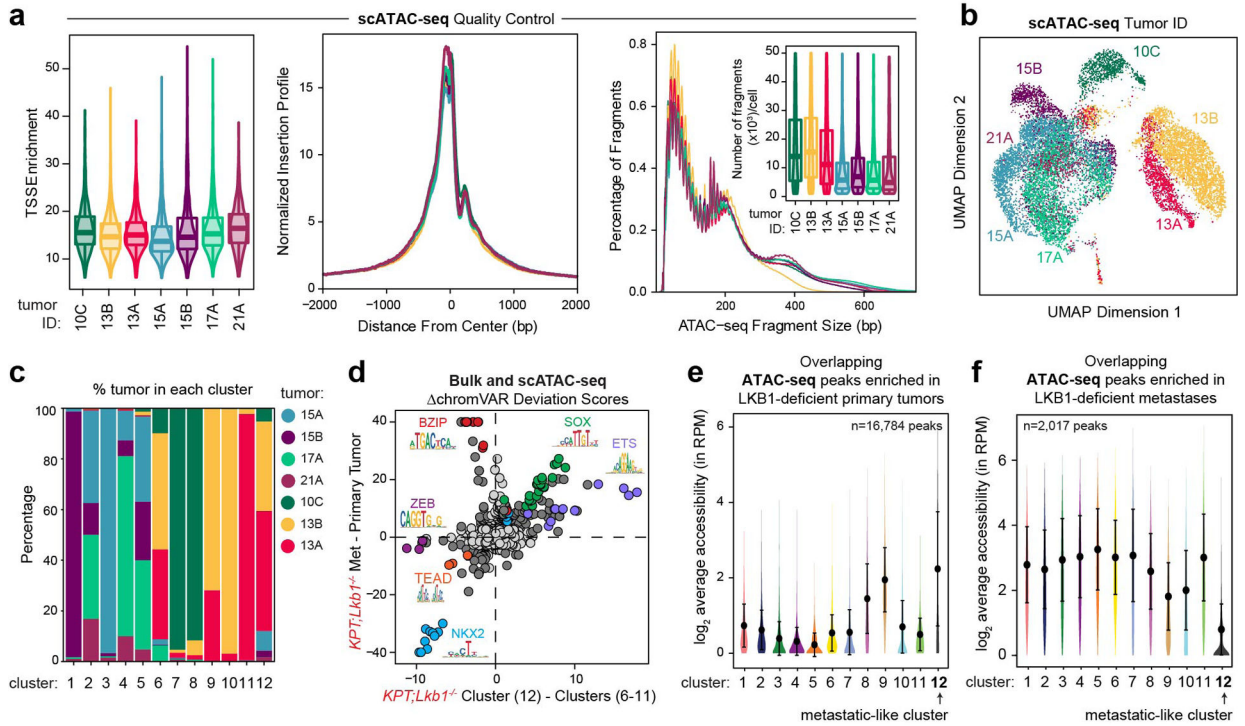
a. Representative immunohistochemistry (IHC) against SOX17 and grading of SOX17 expression for LKB1-proficient *KPT* and LKB1-deficient *KPT;Lkb1<sup>-/-</sup>* samples. Images are annotated according to percent area of the tumor composed of SOX17+ cells. Negative (0%), low (<25%), medium (25–50%), and high (>50%). Scale bars represent 50uM. Images are representative of 117 *KPT* primary tumors, 203 *KPT;Lkb1<sup>-/-</sup>* primary tumors, 14 *KPT* metastases, and 8 *KPT;Lkb1<sup>-/-</sup>* metastases, as quantified in (b).

b. Quantitation of SOX17 protein expression in LKB1-proficient *KPT* and LKB1-deficient *KPT;Lkb1<sup>-/-</sup>* primary tumors and metastases, graded according to (a). The number of samples analyzed for histology for each genotype and tumor type is indicated at the top. Overall 0% of LKB1-proficient primary tumors or metastases had SOX17+ cells, 31% of LKB1-deficient primary tumors had SOX17+ cells, and 100% of LKB1-deficient metastases had SOX17+ cells.

c. Correlation of SOX17 mRNA expression (y-axis) and LKB1 mRNA expression (x-axis) in ten human lung adenocarcinoma samples that contain Type 1 metastatic cell clusters (H0 and H3; Laughney et al. 2020). Each point indicates the mean value of SOX17 or LKB1

expression for each sample  $\pm$  SEM for all single cells evaluated by scRNA-seq. Shaded area represents 95th percent confidence interval.

d. *SOX17* genome accessibility track of the average ATAC-seq signal from Chromatin Type 1 and Chromatin Type 2 tumors.



**Extended Data Fig. 9. A subset of LKB1-deficient primary tumors harbor metastatic-like, SOX17+ sub-populations.**

a. scATAC-seq quality control metrics. TSS enrichment (left, middle), insertion profiles (right), and number of fragments per cell (right inset) in seven primary tumors. N=998 cells for 10C, 3556 cells for 13B, 1467 cells for 13A, 3373 cells for 15A, 1310 cells for 15B, 2858 cells for 17A, and 851 cells for 21A. Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.

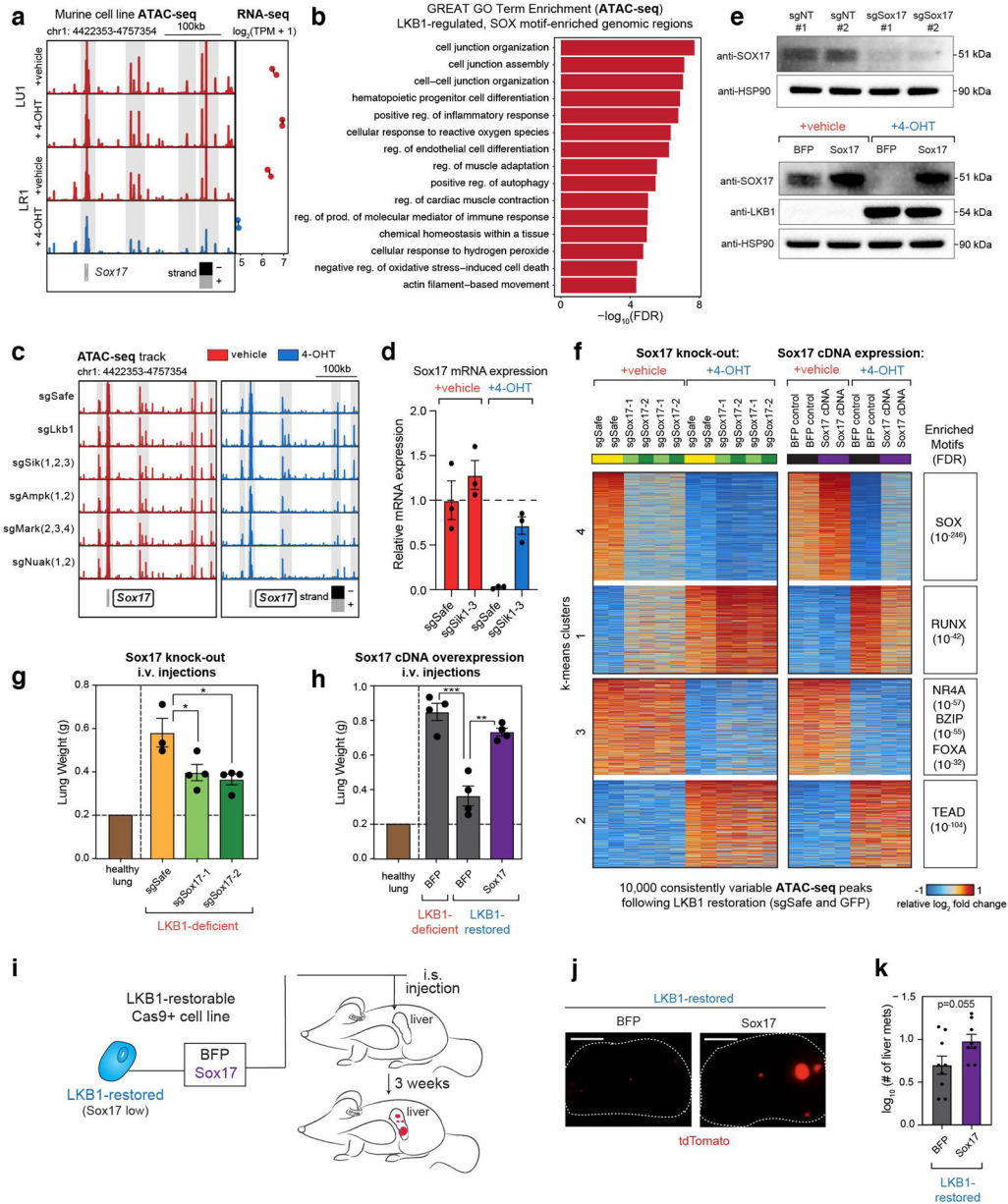
b. UMAP of cells from seven primary tumors.

c. Percent of cells from each cluster in each primary tumor.

d. Comparison of the changes in motif accessibility ( chromVAR deviation scores) between LKB1-deficient metastases and primary tumors (y-axis) versus the average difference between cluster 12 cells and cells in clusters 1–11 (x-axis). Dark grey or colored points are called significantly different ( $q < 0.05$ ) across both comparisons. Light grey points are not significant.

e and f. Average accessibility of peaks in each scATAC-seq cluster that are enriched in LKB1-deficient primary tumors compared to LKB1-deficient metastases (e) or enriched in LKB1-deficient metastases compared to LKB1-deficient primary tumors (f) and are overlapping with the scATAC-seq peakset. Error bars indicate  $\pm$  SEM. N= 2993 cells

(Cluster 1), N= 1011 cells (Cluster 2), N= 508 cells (Cluster 3), N= 856 cells (Cluster 4), N= 408 cells (Cluster 5), N= 3435 cells (Cluster 6), N= 468 cells (Cluster 7), N= 1517 cells (Cluster 8), N= 1733 cells (Cluster 9), N= 119 cells (Cluster 11), N= 116 cells (Cluster 12). Box-whisker plot; lower whisker is the lowest value greater than the 25% quantile minus 1.5 times the interquartile range (IQR), the lower hinge is the 25% quantile, the middle is the median, the upper hinge is the 75% quantile and the upper whisker is the largest value less than the 75% quantile plus 1.5 times the IQR.



**Extended Data Fig. 10. SOX17 regulates chromatin accessibility state and growth in metastatic, LKB1-deficient cells.**

a. *Sox17* genome accessibility track (left) and mean mRNA expression (right) following 4-OHT or vehicle. Significantly, LKB1-deficient cells. differential ATAC-seq peaks in grey



- (log<sub>2</sub> fold change < -0.5, FDR < 0.05). *Sox17* also has significantly decreased mRNA expression (log<sub>2</sub> fold change < -1, FDR < 0.05).
- b. GREAT GO term enrichment of genes nearby the differential peaks that contain SOX binding motifs.
  - c. *Sox17* genome accessibility track of an LKB1-restorable cell line (LR1;Cas9) transduced with the indicated sgRNAs and treated with 4-OHT or vehicle.
  - d. Relative *Sox17* mRNA expression in LR1;Cas9 cells transduced with sgSafe or sgSik1-3 and treated with either vehicle or 4-OHT. Mean values +/- SEM. N=3 biologically independent samples examined over 2 experiments.
  - e. Expression of SOX17 and/or LKB1 by immunoblot in LR2;Cas9 cells transduced with non-targeting (sgNT#1 and sgNT#2) or Sox17-targeting sgRNAs (sgSox17#1 and sgSox17#2) (top) or LR2;Cas9 cells transduced with BFP-overexpressing (control) or Sox17-overexpressing constructs and treated with vehicle or 4-OHT. HSP90 is a sample processing control.
  - f. Heatmap of relative log<sub>2</sub>fold changes of the indicated genotypes of LR2;Cas9 cells. The top 10,000 consistent, variable ATAC-seq peaks following LKB1 restoration in both sgSafe and BFP transduced cells are shown. Clusters 3 and 4 from the Sox17 knock-out experiment are shown independently for emphasis in Fig. 5d.
  - g and h. Lung weight following injection of LR2;Cas9 cells treated with vehicle or 4-OHT after Sox17 knock-out (g) or Sox17 overexpression (h). \*p<0.05, \*\*p<0.005, \*\*\*p<0.0005 with a two-sided t-test. N=3 biologically independent mice evaluated for LKB1-deficient (sgSafe) and 4 biologically independent mice for all other conditions. p=0.0481 for sgSafe vs. sgSox17-1 (LKB1-deficient), p = 0.0184 for sgSafe vs. sgSox17-2 (LKB1-deficient), p=0.0008 for BFP-vehicle vs. BFP-4-OHT, and p = 0.001 for BFP-4-OHT vs. Sox17-4-OHT.
  - i. Schematic of intrasplenic (i.s.) injections into immunocompromised NSG mice.
  - j. Representative fluorescent tdTomato+ images of the left lateral lobe of the liver. Scale bars represent 5mm.
  - k. Log<sub>10</sub> (number of liver metastases) following intrasplenic injection of cells. Condition +/- SEM is shown. p=0.055 with a two-sided t-test. N=9 mice BFP, N=8 mice for Sox17.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments:

We thank J. Sage, A. Trevino, and members of the Greenleaf and Winslow laboratories for helpful comments. We thank the Stanford Shared FACS facility, Veterinary Service Center, and P. Chu for technical support. We thank A. Orantes for administrative support. S.E.P was supported by an NSF Graduate Research Fellowship Award and the Tobacco-Related Diseases Research Program Predoctoral Fellowship Award (T31DT1900). This work was supported by NIH R01-CA204620 and NIH R01-CA230919 (to M.M.W), NIH RM1-HG007735 and UM1-HG009442 (to H.Y.C. and W.J.G.), R35-CA209919 (to H.Y.C.), UM1-HG009436 and U19-AI057266 (to W.J.G.), and in part by the Stanford Cancer Institute support grant (NIH P30-CA124435).

## Data availability:

RNA-seq, scATAC-seq, and ATAC-seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE167381. The human lung adenocarcinoma data were derived from the TCGA Research Network: <http://cancergenome.nih.gov/>. The data-set derived from this resource that supports the findings of this study is publicly available: <https://gdc.cancer.gov/about-data/publications/ATACseq-AWG>. All other data supporting the findings of this study are available from the corresponding author on reasonable request. Transcription factor binding motifs were derived from CIS-BP: <http://cisbp.cbr.utoronto.ca/index.php>.

## References:

1. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543–550 (2014). [PubMed: 25079552]
2. Waddell Net al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* 518, 495–501 (2015). [PubMed: 25719666]
3. Sanchez-Cespedes MA role for LKB1 gene in human cancer beyond the Peutz-Jeghers syndrome. *Oncogene* 26, 7825–7832 (2007). [PubMed: 17599048]
4. Ji Het al. LKB1 modulates lung cancer differentiation and metastasis. *Nature* 448, 807–810 (2007). [PubMed: 17676035]
5. Carretero Jet al. Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. *Cancer Cell* 17, 547–559 (2010). [PubMed: 20541700]
6. Shackelford DB & Shaw RJ The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. *Nat. Rev. Cancer* 9, 563–575 (2009). [PubMed: 19629071]
7. Jin Let al. The PLAG1-GDH1 Axis Promotes Anoikis Resistance and Tumor Metastasis through CamKK2-AMPK Signaling in LKB1-Deficient Lung Cancer. *Mol. Cell* 69, 87–99.e7 (2018). [PubMed: 29249655]
8. Calles Aet al. Immunohistochemical Loss of LKB1 Is a Biomarker for More Aggressive Biology in KRAS-Mutant Lung Adenocarcinoma. *Clin. Cancer Res* 21, 2851–2860 (2015). [PubMed: 25737507]
9. Lizcano JMet al. LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* 23, 833–843 (2004). [PubMed: 14976552]
10. Kottakis Fet al. LKB1 loss links serine metabolism to DNA methylation and tumorigenesis. *Nature* 539, 390–395 (2016). [PubMed: 27799657]
11. Hollstein PEet al. The AMPK-Related Kinases SIK1 and SIK3 Mediate Key Tumor-Suppressive Effects of LKB1 in NSCLC. *Cancer Discov.* 9, 1606–1627 (2019). [PubMed: 31350328]
12. Murray CWet al. An LKB1-SIK Axis Suppresses Lung Tumor Growth and Controls Differentiation. *Cancer Discov.* 9, 1590–1605 (2019). [PubMed: 31350327]
13. Filbin MGet al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* 360, 331–335 (2018). [PubMed: 29674595]
14. Flavahan WAet al. Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* 575, 229–233 (2019). [PubMed: 31666694]
15. LaFave LMet al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* 38, 212–228.e13 (2020). [PubMed: 32707078]
16. Reiter JGet al. Minimal functional driver gene heterogeneity among untreated metastases. *Science* 361, 1033–1037 (2018). [PubMed: 30190408]
17. Hu Z, Li Z, Ma Z & Curtis C Multi-cancer analysis of clonality and the timing of systemic spread in paired primary tumors and metastases. *Nat. Genet* 52, 701–708 (2020). [PubMed: 32424352]
18. Turajlic S & Swanton C Metastasis as an evolutionary process. *Science* 352, 169–175 (2016). [PubMed: 27124450]

19. Robles-Oteiza C et al. Recombinase-based conditional and reversible gene regulation via XTR alleles. *Nat. Commun*6, 8783 (2015). [PubMed: 26537451]
20. Morgens DW et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun*8, 15178 (2017). [PubMed: 28474669]
21. Mi H, Muruganujan A, Ebert D, Huang X & Thomas PD PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res* 47, D419–D426 (2019). [PubMed: 30407594]
22. Buenrostro JD, Giresi PG, Zaba LC, Chang HY & Greenleaf WJ Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* 10, 1213–1218 (2013). [PubMed: 24097267]
23. Corces MR et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* 14, 959–962 (2017). [PubMed: 28846090]
24. Corces MR et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet* 48, 1193–1203 (2016). [PubMed: 27526324]
25. Corces MR et al. The chromatin accessibility landscape of primary human cancers. *Science* 362, (2018).
26. Schep AN, Wu B, Buenrostro JD & Greenleaf WJ chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* 14, 975–978 (2017). [PubMed: 28825706]
27. Kaufman JM et al. A transcriptional signature identifies LKB1 functional status as a novel determinant of MEK sensitivity in lung adenocarcinoma. *Cancer Res.* 77, 153–163 (2017). [PubMed: 27821489]
28. Winslow MM et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* 473, 101–104 (2011). [PubMed: 21471965]
29. Park K-S, Wells JM, Zorn AM, Wert SE & Whitsett JA Sox17 influences the differentiation of respiratory epithelial cells. *Dev. Biol* 294, 192–202 (2006). [PubMed: 16574095]
30. Laughney A et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med* 26, 259–269 (2020). [PubMed: 32042191]
31. Satpathy A et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol* 37, 925–936 (2019). [PubMed: 31375813]
32. Granja JM et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol* 37, 1458–1465 (2019). [PubMed: 31792411]
33. Stuart T et al. Comprehensive Integration of Single-Cell Data. *Cell* 177, 1888–1902.e21 (2019). [PubMed: 31178118]
34. Walkinshaw D et al. The tumor suppressor kinase LKB1 activates the downstream kinases SIK2 and SIK3 to stimulate nuclear export of class IIa histone deacetylases. *J. Biol. Chem* 288, 9345–9362 (2013). [PubMed: 23393134]
35. Parra M Class IIa HDACs - new insights into their functions in physiology and pathology. *FEBS J* 282, 1736–1744 (2015). [PubMed: 25244360]
36. Zhang H et al. Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nat. Commun* 8, 14922 (2017). [PubMed: 28387316]
37. Bray NL, Pimentel H, Melsted P & Pachter L Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol* 34, 525–527 (2016). [PubMed: 27043002]
38. Morgens DW, Deans RM, Li A & Bassik MC Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol* 34, 634–636 (2016). [PubMed: 27159373]
39. Li W et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* 15, 554 (2014). [PubMed: 25476604]
40. Adamson B et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167, 1867–1882.e21 (2016). [PubMed: 27984733]
41. Chuang C-H et al. Molecular definition of a metastatic lung cancer state reveals a targetable CD109-Janus kinase-Stat axis. *Nat. Med* 23, 291–300 (2017). [PubMed: 28191885]

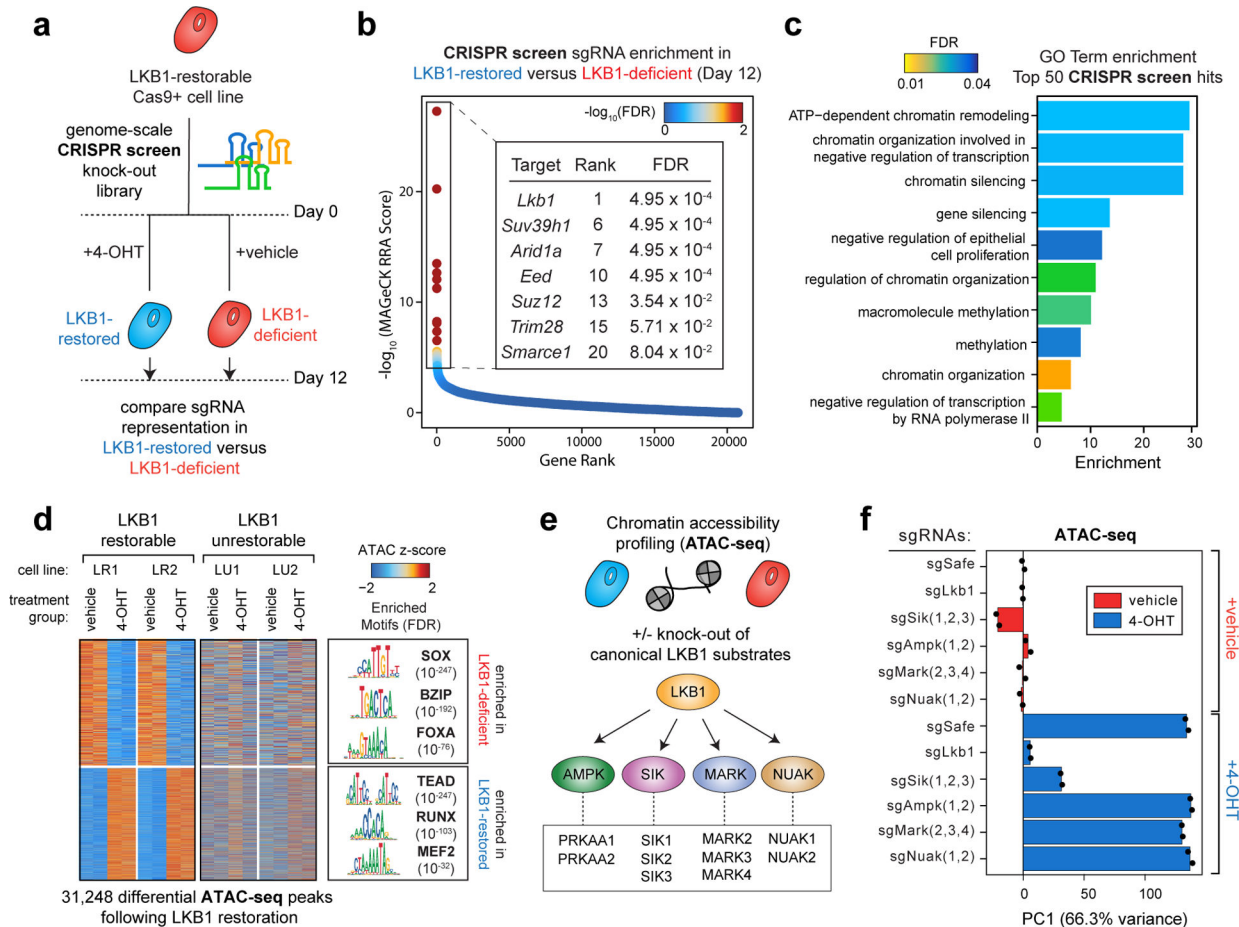
42. Granja JMet al.ArchR: An integrative and scalable software package for single-cell chromatin accessibility analysis. *BioRxiv* (2020). doi:10.1101/2020.04.28.066498
43. Jeffmgranja. GreenleafLab/LKB1\_2021: Release\_1.0.1. *Zenodo* (2021). doi:10.5281/zenodo.5035694

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript



**Figure 1. An LKB1-SIK axis regulates chromatin accessibility in lung adenocarcinoma.**

a. Schematic of a genome-scale screen in an LKB1-restorable, Cas9+ lung adenocarcinoma cell line (LR1;Cas9) treated with 4-OHT to restore LKB1 or treated with vehicle to remain LKB1-deficient (see also Extended Data Fig. 1a).

b. sgRNA targets (genes) enriched in LKB1-restored versus LKB1-deficient cells are ranked by  $\log_{10}(\text{MAGeCK RRA score})$  and colored by  $-\log_{10}(\text{FDR})$  values. LKB1 and six chromatin-related genes with an  $\text{FDR} < 0.01$  are shown alongside their individual rank and FDR values. Full list available in Supplementary Table 1.

c. PANTHER GO term enrichment of the top 50 sgRNA targets (genes) enriched in LKB1-restored cells compared to LKB1-deficient cells. Bars are colored by enrichment FDR values.

d. Left: Heatmap of chromatin peak accessibility for each cell line after treatment with 4-OHT or vehicle for six days. Each row represents a z-score of  $\log_2$  normalized accessibility within each cell line using ATAC-seq. Right: Transcription factor hypergeometric motif enrichment with FDR indicated in parentheses.

e. Schematic of knocking out canonical LKB1 substrate families with arrays of sgRNAs in an LKB1-restorable cell line (LR1;Cas9), treating with 4-OHT or vehicle for six days, and performing ATAC-seq.

f. Principle component analysis (PCA) of the top 10,000 variable ATAC-seq peaks across the twelve indicated sgRNA populations that were treated with either 4-OHT or vehicle for six



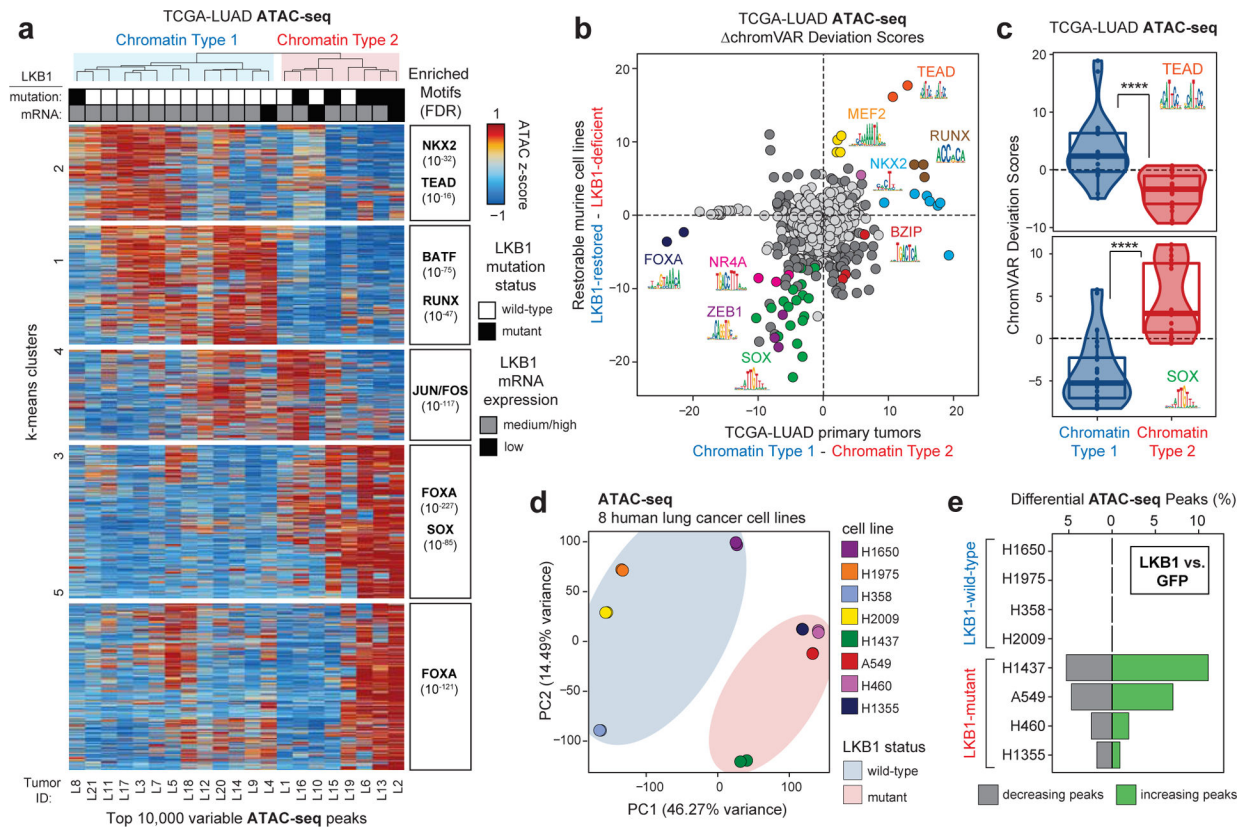
days. Individual principle components besides PC1 (66.3%) account for <4% of the variance in the dataset. N=2 technical replicates per sgRNA population.

Author Manuscript

Author Manuscript

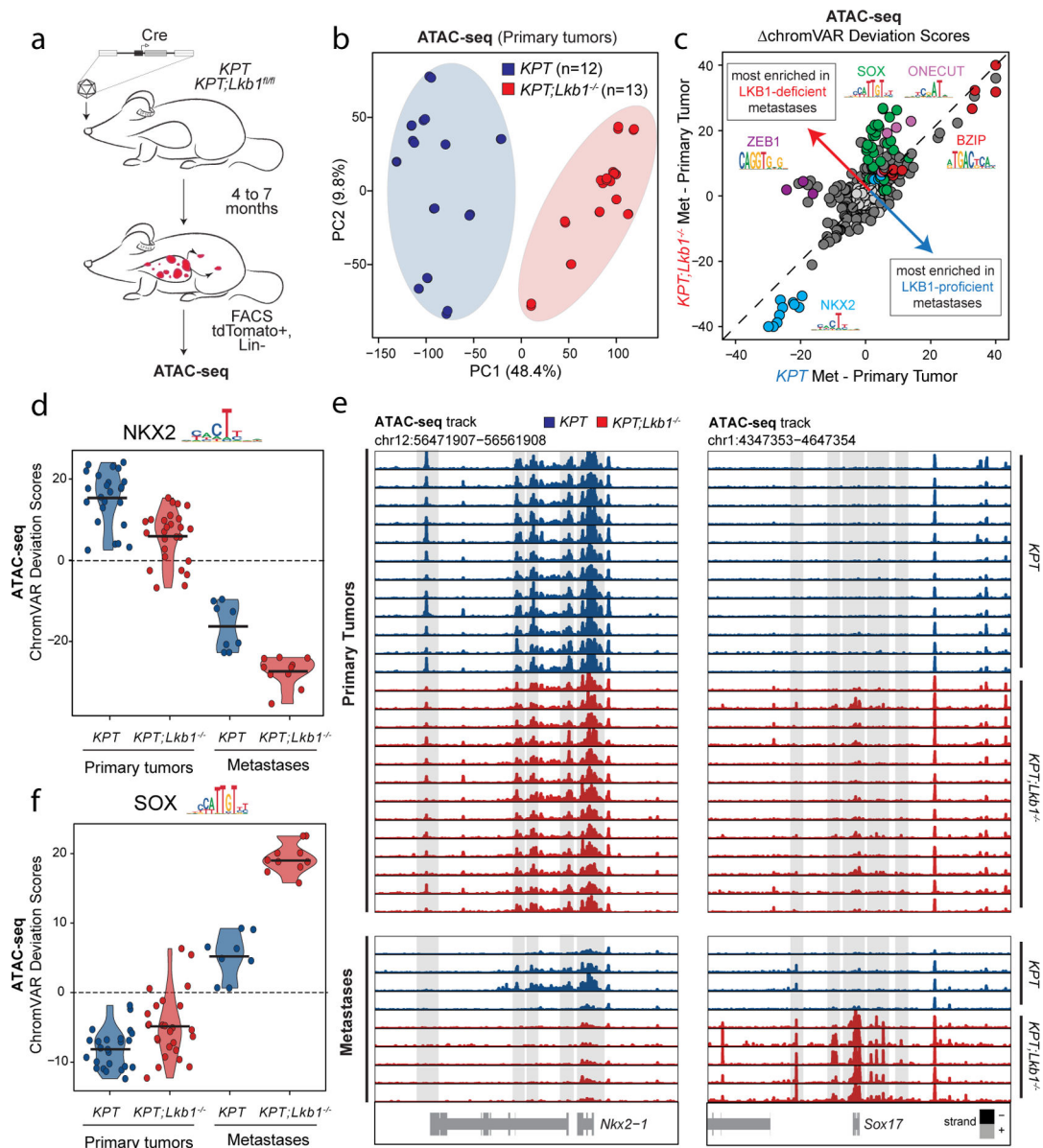
Author Manuscript

Author Manuscript



**Figure 2. LKB1 mutation status distinguishes the two main chromatin sub-types of human lung adenocarcinoma.**

- a. Left: Unsupervised hierarchical clustering of 21 human lung adenocarcinoma samples from the TCGA-ATAC dataset using the top 10,000 variable peaks across all samples, visualized as a heatmap of peak accessibility. Each row represents a z-score of  $\log_2$  normalized accessibility. Right: Transcription factor hypergeometric motif enrichment in each k-means cluster with FDR indicated in parentheses.
- b. Comparison of the changes in motif accessibility ( $\Delta$  chromVAR Deviation Scores) across Chromatin Type 1 and Chromatin Type 2 human primary tumors (x-axis) and across LKB1-restored and LKB1-deficient murine cells (x-axis). Dark grey or colored points are significantly different ( $q < 0.05$ , see Methods) across both comparisons. Light grey points are not significant. Only motifs for transcription factors shared across human and murine CisBP databases are shown.
- c. ChromVAR deviation scores for TEAD (top) and SOX (bottom) transcription factor motifs for samples in the TCGA-LUAD ATAC-seq dataset. \*\*\*\* $p < 10^{-6}$  using a two-sided t-test.  $p = 9 \times 10^{-6}$  for TEAD and  $p = 6 \times 10^{-7}$  for SOX.  $N = 13$  biologically independent samples from Chromatin Type 1 and 8 biologically independent samples from Chromatin Type 2.
- d. PCA of the top 10,000 variable ATAC-seq peaks across eight human lung cancer cell lines. LKB1 mutant status is indicated.
- e. Percent of differential ATAC-seq peaks ( $|\log_2 \text{fold change}| > 0.5$ ,  $\text{FDR} < 0.05$ ) in cells transduced to express LKB1 compared to a GFP control.



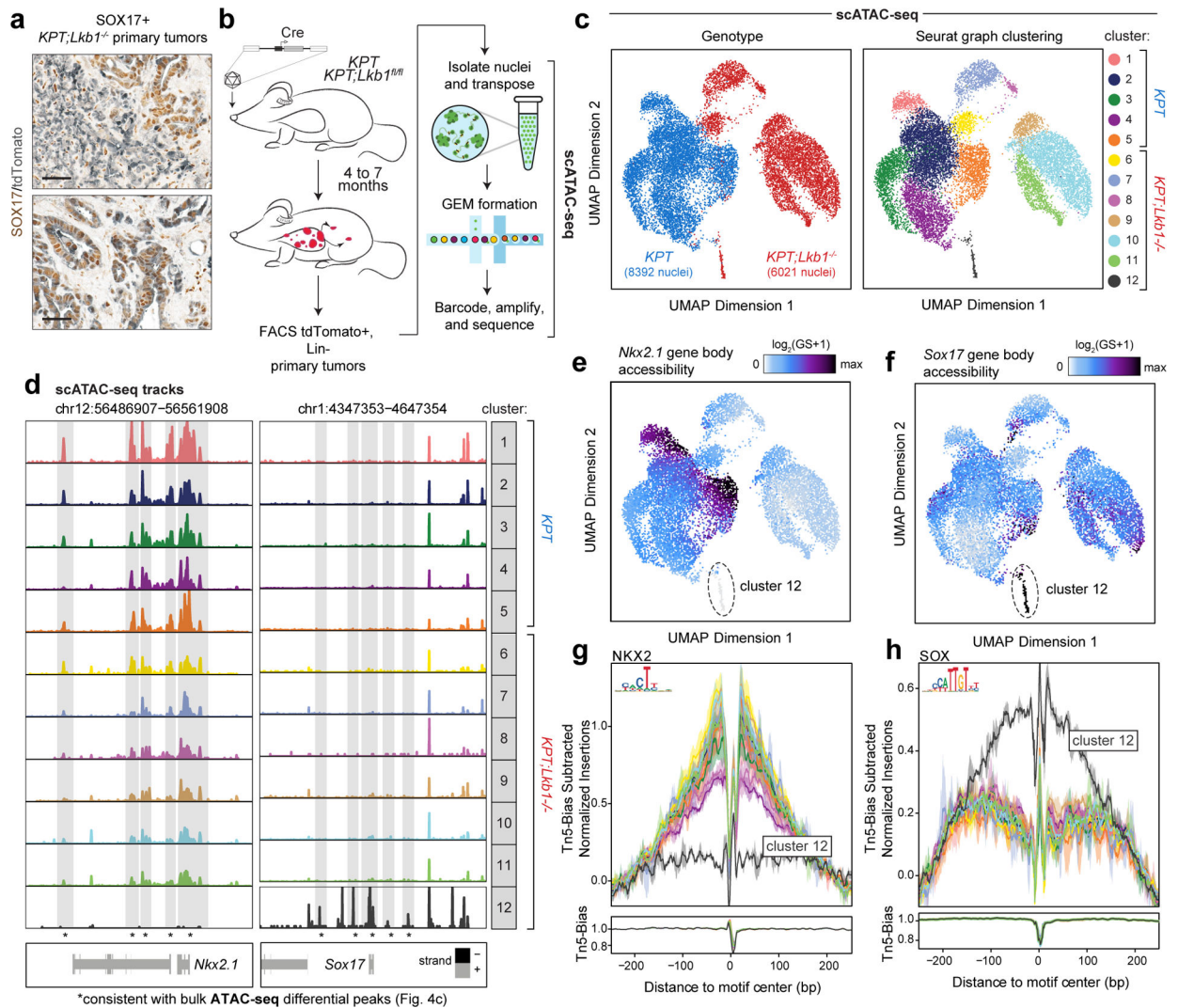
**Figure 3. Genotype-specific activation of SOX17 expression in metastatic, LKB1-deficient cells.**

a. Schematic of tumor initiation, sample processing, and multi-omic profiling. Lentiviral Cre initiates tumors in *Kras*<sup>LSL-G12D</sup>;*p53*<sup>flx/flx</sup>;*Rosa26*<sup>LSL-tdTomato</sup> (*KPT*) mice with and without homozygous *Lkb1*<sup>flx/flx</sup> alleles. tdTomato<sup>+</sup> cancer cells negative for the lineage markers CD45, CD31, F4/F80, and Ter119 were sorted by FACS before library preparation for ATAC-seq, scATAC-seq, and RNA-seq.

b. PCA of the top 10,000 variable ATAC-seq peaks across 25 primary tumor samples. Technical replicates are averaged.

c. Comparison of the changes in motif accessibility (chromVAR Deviation Scores) across LKB1-proficient (x-axis) and LKB1-deficient (y-axis) metastases compared to primary tumors of the same genotype. Dark grey or colored points are called significantly different ( $q < 0.05$ , see Methods) across both comparisons. Light grey points are not significant.

- d. chromVAR deviation scores for NKX2 motifs across LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1<sup>-/-</sup>*) primary tumor and metastasis samples.
- e. Nkx2.1 and Sox17 genome accessibility tracks for each primary tumor (top) and metastasis (bottom) sample.
- f. chromVAR deviation scores for SOX motifs across LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1<sup>-/-</sup>*) primary tumor and metastasis samples.



**Figure 4. LKB1-deficient primary tumors harbor sub-populations of SOX17+ cells.**

a. Representative immunohistochemistry for SOX17 (in brown) and tdTomato (in grey) on two LKB1-deficient lung adenocarcinoma primary tumors. Scale bars represent 50 μm. Images are representative of 117 *KPT* primary tumors, 203 *KPT;Lkb1<sup>-/-</sup>* primary tumors, 14 *KPT* metastases, and 8 *KPT;Lkb1<sup>-/-</sup>* metastases, as quantified in Extended Data Fig. 8b.

b. Schematic of tumor initiation and processing for scATAC-seq. tdTomato+, DAPI- cancer cells that were negative for the lineage (Lin) markers CD45, CD31, F4/F80, and Ter119 were sorted by FACS before scATAC-seq library preparation.

c. Uniform Manifold Approximation and Projection (UMAP) of 8392 nuclei from 4 *KPT* primary tumors and 6021 nuclei from 3 *KPT;Lkb1<sup>-/-</sup>* primary tumors, colored by genotype (left) or cluster according to Seurat graph clustering (right).

d. *Nkx2.1* and *Sox17* genome accessibility tracks for each cluster indicated in Fig. 4c. Significant ATAC-seq peaks from bulk chromatin accessibility profiling (Fig. 3e) are highlighted in grey and indicated with an asterisk (\*).

e and f. UMAP colored by the average gene body accessibility for *Nkx2.1* (e) or *Sox17* (f) in each cell.



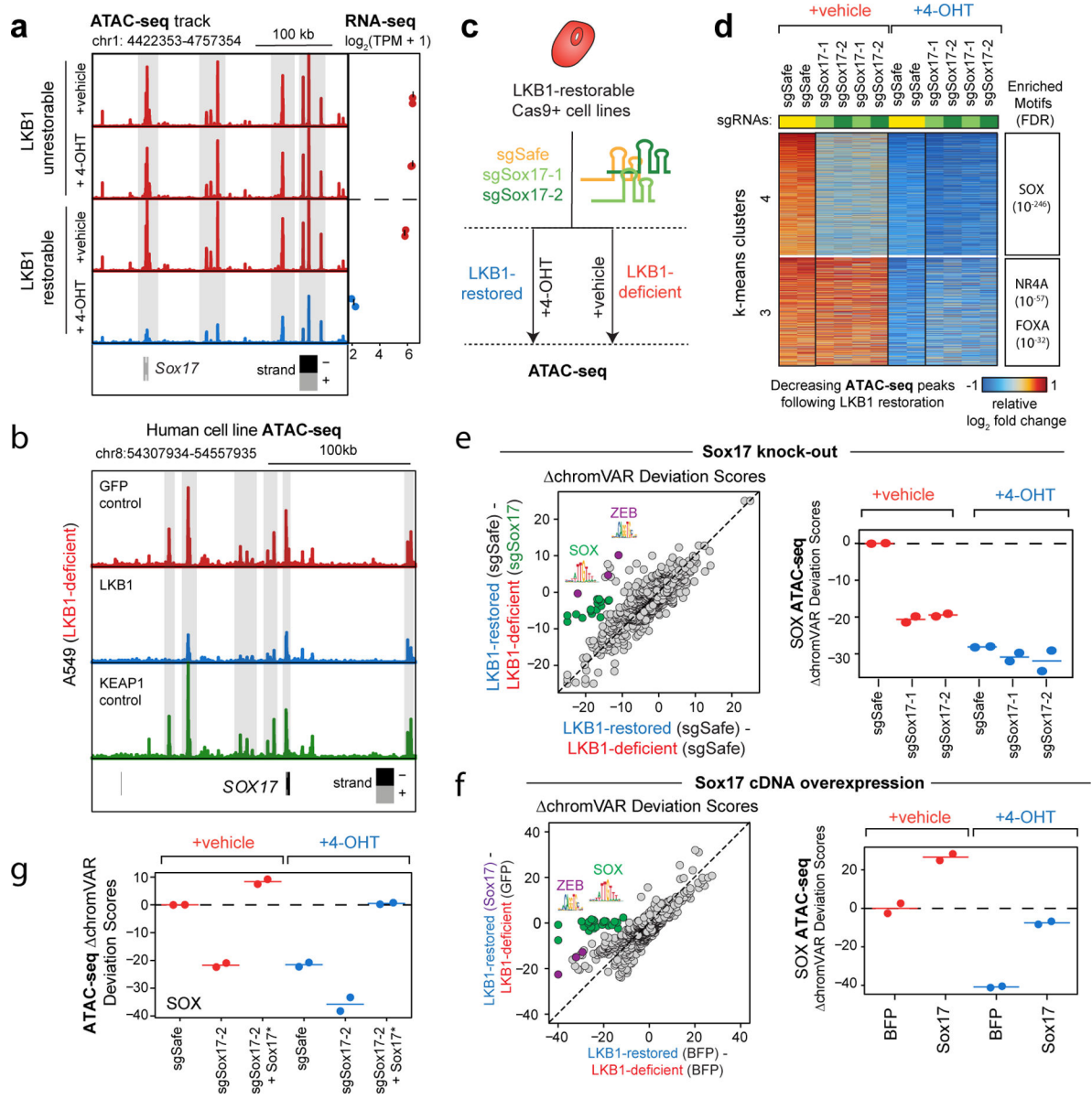
g and h. Top: Footprint of accessibility for each scATAC-seq cluster for genomic regions containing NKX2 (g) and SOX (h) motifs. Bottom: Modeled hexamer insertion bias of Tn5 around sites containing each motif.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

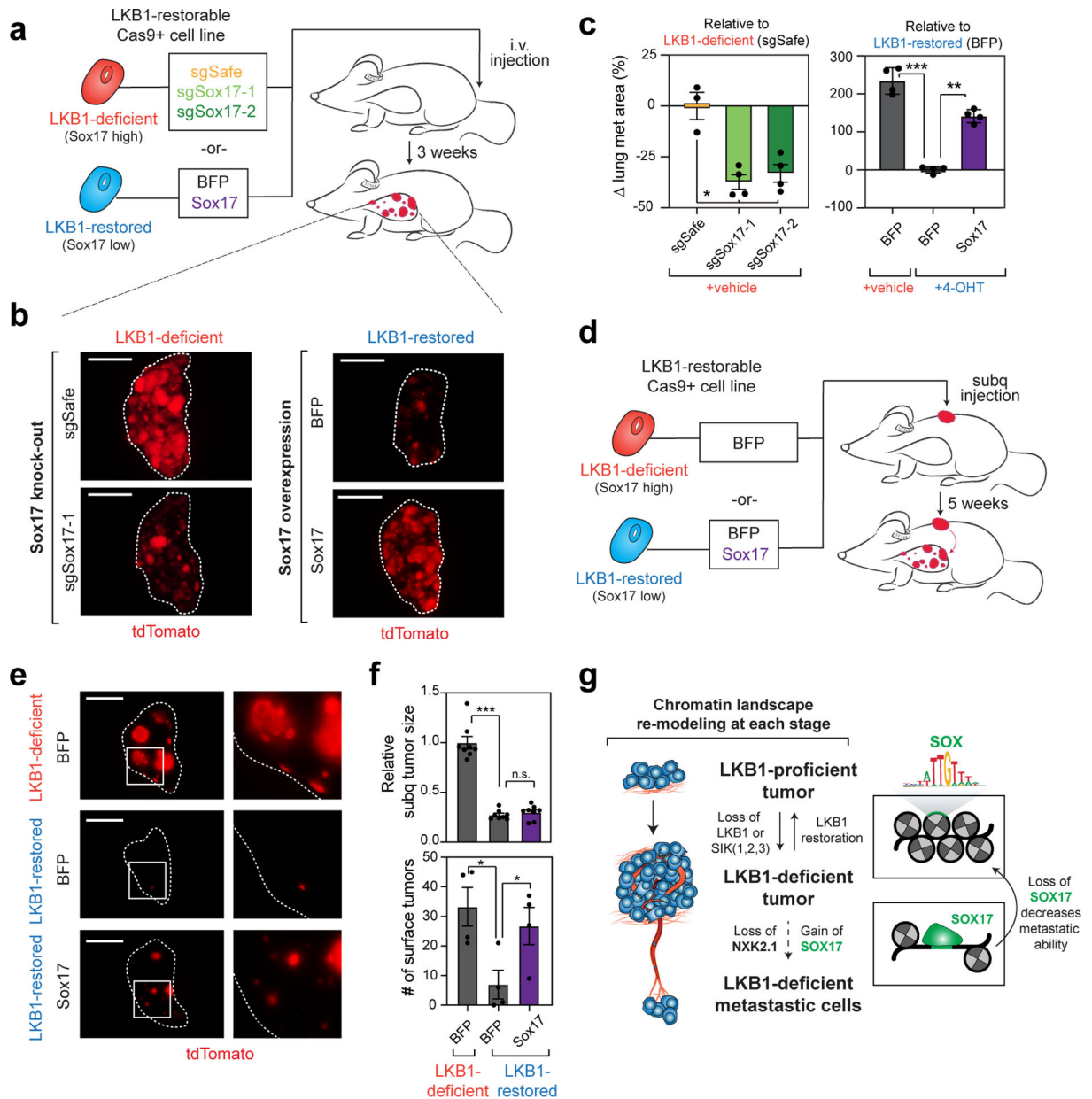


**Figure 5. SOX17 regulates the chromatin accessibility state of metastatic, LKB1-deficient cells.**  
 a. *Sox17* genome accessibility track (left) and mean *Sox17* mRNA expression across technical replicates (right) of an LKB1-unrestorable cell line (LU2) and a metastatic LKB1-restorable cell line (LR2) treated with 4-OHT or vehicle for six days. Highlighted in grey are significantly differential ATAC-seq peaks ( $\log_2$  fold change  $< -0.5$ , FDR  $< 0.05$ ) following LKB1 restoration. *Sox17* also has significantly decreased mRNA expression ( $\log_2$  fold change  $< -1$ , FDR  $< 0.05$ ) following LKB1 restoration.  
 b. *SOX17* genome accessibility track of an LKB1-deficient cell line (A549) transduced with GFP, LKB1, or KEAP1.  
 c. Schematic of knocking out *Sox17* with and without LKB1 restoration and performing ATAC-seq.

d. Heatmap of the relative  $\log_2$  fold changes in k-means clusters 3 and 4 of the indicated genotypes of cells with and without LKB1 restoration compared to the average  $\log_2$  fold changes in sgSafe control cells. A subset (5,379 peaks; all decreasing peaks) of the top 10,000 consistent, variable ATAC-seq peaks following LKB1 restoration in cells transduced with either sgSafe or blue fluorescent protein (BFP) controls are shown. Full heatmap is shown in Extended Data Fig. 7f.

e and f. Left: Comparison of the changes in motif accessibility ( chromVAR deviation scores) between the indicated perturbation populations. Dark grey or colored points are called significantly different ( $q < 0.05$ ) across both comparisons. Right: ChromVAR deviation scores for SOX motifs in each group (normalized to vehicle-treated sgSafe (e) or vehicle-treated BFP (f)). Each point represents an ATAC-seq technical replicate, bar represents the mean.

g. ChromVAR deviation scores for SOX motifs in each group in another cell line (LR1;Cas9). Each individual point represents an ATAC-seq technical replicate, bar represents the mean. sgSox17-2 + Sox17\* indicates that the cells were transduced with a construct containing a sgRNA targeting *Sox17* as well as a Sox17 cDNA that is resistant to sgRNA cutting (see Methods). N=2 technical replicates for each condition.



**Figure 6. SOX17 regulates the metastatic ability of LKB1-deficient cells.**

a. Schematic of injecting LKB1-deficient cells expressing sgRNAs targeting Sox17 (sgSox17-1 and sgSox17-2) or injecting LKB1-restored cells expressing Sox17 cDNA intravenously (i.v.) into immunocompromised NSG mice. Tumor burden was analyzed three weeks post-injection.

b. Representative fluorescent tdTomato+ images of single lung lobes following i.v. injection. Similar results were observed from three additional mice (four mice total) per condition, except for sgSafe + vehicle in which similar results were observed from two additional mice (three mice total). Scale bars represent 5mm.

c. Change in % tumor area compared to sgSafe + vehicle (left) or compared to BFP + 4-OHT (right). Each point represents an individual mouse, mean  $\pm$  SEM is shown. \* $p <$

0.01, \*\*  $p < 0.001$ , \*\*\*  $p < 0.0001$ . N=3 (sgSafe + vehicle) and N=4 (all other conditions) biologically independent samples per condition. P = 0.0031 for sgSox17-1 vs. sgSafe, P = 0.0072 for sgSox17-2 vs. sgSafe, P = 0.0031 for BFP + 4-OHT vs. Sox17 + 4-OHT, P < 0.0001 for BFP + 4-OHT vs. BFP + vehicle.

d. Schematic of injecting LKB1-deficient cells (LR2) expressing BFP or injecting LKB1-restored cells (LR2) expressing Sox17 cDNA or BFP subcutaneously (subq) into immunocompromised NSG mice. Metastatic tumor burden to the lung was analyzed five weeks post-injection.

e. Representative fluorescent tdTomato+ images of single lung lobes following subq injection as outlined in (d). Similar results were observed from three additional mice (four mice total) per condition. Scale bars represent 5mm.

f. Top: Relative tumor size following subcutaneous injection of the indicated cells. Condition +/- SEM is shown. \*\*\*  $p = 0.0001$ , n.s. = not significant with a two-sided t-test. Bottom: Number of surface tumors observed in the five lung lobes following subcutaneous injection of the indicated cells. Condition +/- SEM is shown. \* $p < 0.05$  with a two-sided t-test. P = 0.0178 for BFP-4-OHT vs. BFP-vehicle. P = 0.0471 for BFP-4-OHT vs. Sox17-4-OHT. N=4 biologically independent mice evaluated per condition with 2 subcutaneous tumors each.

g. Summary of LKB1-induced chromatin accessibility changes in primary tumors and metastases.