nature cell biology

ARTICLES

https://doi.org/10.1038/s41556-021-00728-4

Check for updates

LKB1 inactivation modulates chromatin accessibility to drive metastatic progression

Sarah E. Pierce[1,8] ✉, Jeffrey M. Granja[1,2,8], M. Ryan Corces[2], Jennifer J. Brady[1], Min K. Tsai[1], Aubrey B. Pierce[1], Rui Tang[1], Pauline Chu[3], David M. Feldser[4], Howard Y. Chang[1,2,5], Michael C. Bassik[1,6], William J. Greenleaf[1,2] ✉ and Monte M. Winslow[1,6,7] ✉

Metastasis is the leading cause of cancer-related deaths and enables cancer cells to compromise organ function by expanding in secondary sites. Since primary tumours and metastases often share the same constellation of driver mutations, the mechanisms that drive their distinct phenotypes are unclear. Here we show that inactivation of the frequently mutated tumour suppressor gene *LKB1* (encoding liver kinase B1) has evolving effects throughout the progression of lung cancer, which leads to the differential epigenetic re-programming of early-stage primary tumours compared with 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 tumours. Using an in vivo model of metastatic progression, we further show that loss of LKB1 activates the early endoderm transcription factor SOX17 in metastases and a metastatic-like sub-population of cancer cells within primary tumours. The expression of SOX17 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 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[1–3]. Loss of LKB1 leads to both increased primary tumour growth and the acquisition of metastatic ability in lung adenocarcinoma—the most common subtype of lung cancer[4–8]. However, beyond its well-established role as an activator of AMPK-related kinases[6,9–12], 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 contain mutations in chromatin-modifying genes, such as *SETD2*, *ARID1A* and *SMARCA4*[1], which suggests that genetic alterations can drive tumour progression by influencing epigenetic state. This interaction between genetic and epigenetic mechanisms is starting to be characterized in other cancer types[13]; for example, H3K27M mutations in diffuse midline gliomas suppress epigenetic repressive capacity and differentiation[14] and SDH deficiency in gastrointestinal stromal tumours initiates global DNA hyper-methylation and unique oncogenic programs[15]. The chromatin accessibility profiling of mouse lung adenocarcinoma tumours has also started to unveil the epigenomic state transitions that occur during the development of lung cancer[16]. Genetic profiling of primary tumours and matched metastases has revealed a lack of metastasis-specific driver mutations, opening up the possibility that epigenetic mechanisms drive metastasis[17–19]. 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, p53-deficient (*Kras^{G12D};Trp53$^{-/-}$;KP*) mouse lung tumours that contain homozygous restorable alleles of *Lkb1* (*Lkb1$^{TR/TR}$*; *Lkb1* is also known as *Stk11*) and a tamoxifen-inducible FlpOER allele (*Rosa26^{FlpOER}*)[20] (Extended Data Fig. 1a,b and Methods). 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 that encode functional domains of LKB1. Treatment with 4-hydroxytamoxifen (4-OHT) results in nuclear translocation of FlpOER and excision of the FRT-flanked gene trap cassette, thereby restoring full-length expression of *Lkb1* (Extended Data Fig. 1c–e). Restoring LKB1 decreased proliferation in cells in culture and decreased tumour growth after transplantation into mice, whereas treating LKB1-unrestorable, FlpOER-negative cell lines (LU1 and LU2) with 4-OHT had no effect (Extended Data Fig. 1f–i).

A genetic link between LKB1–SIK and chromatin regulation

To identify genes and pathways that contribute to LKB1-mediated tumour suppression, we performed a proliferation-based genome-scale CRISPR–Cas9 knockout screen in both LKB1-deficient and LKB1-restored cells (Fig. 1a). We first transduced a Cas9-expressing

[1]Department of Genetics, Stanford University School of Medicine, Stanford, CA, USA. [2]Center for Personal and Dynamic Regulomes, Stanford University School of Medicine, Stanford, CA, USA. [3]Department of Comparative Medicine, Stanford University School of Medicine, Stanford, CA, USA. [4]Department of Cancer Biology and Abramson Family Cancer Research Institute, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. [5]HHMI, Stanford University School of Medicine, Stanford, CA, USA. [6]Chemistry, Engineering, and Medicine for Human Health (ChEM-H), Stanford University, Stanford, CA, USA. [7]Department of Pathology, Stanford University School of Medicine, Stanford, CA, USA. [8]These authors contributed equally: Sarah E. Pierce, Jeffrey M. Granja. ✉e-mail: pierces@stanford.edu; wjg@stanford.edu; mwinslow@stanford.edu

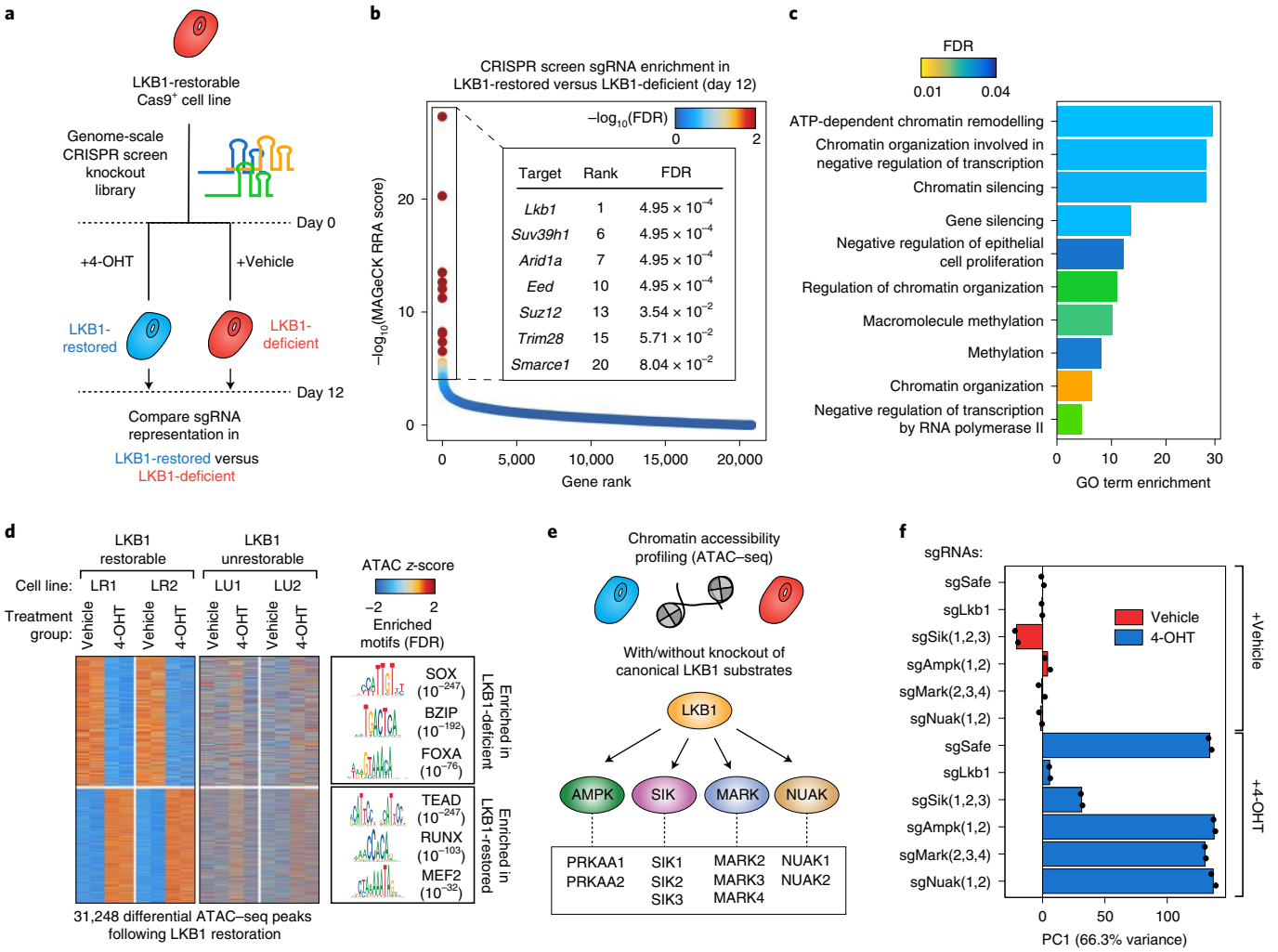


Fig. 1 | An LKB1–SIK axis regulates chromatin accessibility in lung adenocarcinoma. a, Schematic of a genome-scale screen in an LKB1-restorable, Cas9$^+$ lung adenocarcinoma cell line (LR1;Cas9) treated with 4-OHT to restore LKB1 or treated with a vehicle to remain LKB1-deficient (see also Extended Data Fig. 1a). **b**, sgRNA targets (genes) enriched in LKB1-restored versus LKB1-deficient cells are ranked by $\log_{10}$(MAGeCK RRA score) (see Methods) and coloured by $-\log_{10}$(FDR) values, in which FDR denotes the false discovery rate. *LKB1* and six chromatin-related genes with an FDR < 0.01 are shown alongside their individual rank and FDR values. See Supplementary Table 1 for a full list. **c**, PANTHER GO term enrichment of the top 50 sgRNA targets (genes) enriched in LKB1-restored cells compared with LKB1-deficient cells. Bars are coloured by enrichment FDR values. **d**, Left: heat map of chromatin peak accessibility for each cell line after treatment with 4-OHT or vehicle for 6 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 values indicated in parentheses. **e**, Schematic of knocking out canonical LKB1 substrate families with arrays of sgRNAs in LR1;Cas9 cells, treatment with 4-OHT or vehicle for 6 days, and ATAC–seq analysis. **f**, Principle component analysis (PCA) of the top 10,000 variable ATAC–seq peaks across the 12 indicated sgRNA populations that were treated with either 4-OHT or vehicle for 6 days. Individual principle components besides PC1 (66.3%) account for <4% of the variance in the dataset. $n = 2$ technical replicates per sgRNA population.

LKB1-restorable cell line (LR1;Cas9) with a lentiviral library containing ~10 single-guide RNAs (sgRNAs) per gene in the genome as well as ~13,000 inert controls[21]. After selecting for transduced cells, we treated the cells with 4-OHT or vehicle control 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 with LKB1-deficient cells was *Lkb1* itself (Fig. 1b and Extended Data Fig. 1k). Gene Ontology (GO) term enrichment analysis[22] of the remaining top targets surprisingly revealed a strong enrichment of chromatin-related processes (Fig. 1c). In particular, 6 of the top 20 targets were chromatin modifiers (*Suv39h1, Arid1a, Eed, Suz12, Trim28* and *Smarce1*) (Fig. 1b and Extended Data Fig. 1l), which suggests 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)[23,24] (Fig. 1d, Extended Data Fig. 2a–c and Supplementary Table 2). Notably, 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[25,26] (Extended Data Fig. 2e). In addition, most LKB1-induced chromatin changes

occurred within 24–48 h of LKB1 restoration (Extended Data Fig. 2f–j), which suggests 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). Inactivating the top chromatin modifier hits from the screen (*Eed*, *Suz12*, *Trim28* and *Suv39h1*) in the LR1;Cas9 cell line seemed to delay, but not prevent, LKB1-induced chromatin accessibility changes (Extended Data Fig. 3), which suggests compensation between chromatin regulatory pathways.

The canonical tumour suppressive role for LKB1 involves the phosphorylation and activation of AMPK-related kinases, including the AMPK, SIK, NUAK and MARK families[9]. To evaluate whether the downstream substrates of LKB1 contribute to LKB1-induced chromatin changes, we knocked out each family with 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 Fig. 4b–g), whereas inactivation of the *Ampk*, *Nuak* or *Mark* families or the individual *Sik* paralogues (*Sik1*, *Sik2* or *Sik3* independently) had no effect (Fig. 1f and Extended Data Fig. 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 mouse restoration model, we next evaluated whether LKB1 mutation status correlates with chromatin accessibility differences across human lung adenocarcinoma primary tumours. De novo hierarchical clustering of the 21 lung adenocarcinoma samples from the The Cancer Genome Atlas (TCGA) ATAC–seq dataset[26] revealed two chromatin sub-types of lung cancer (annotated as Chromatin Types 1 and 2) (Fig. 2a). Of the top ~200 mutated genes in lung adenocarcinoma, *LKB1* was the most significantly enriched mutated gene in Chromatin Type 2 tumours compared with Chromatin Type 1 tumours (FDR = 0.088) (Extended Data Fig. 5a).

We next evaluated how the distinct chromatin accessibility states of Chromatin Types 1 and 2 human primary tumours compared with the acute chromatin accessibility changes induced by LKB1 restoration in mouse cells. We first calculated the differential accessibility of transcription factor binding motifs between Chromatin Types 1 and 2 human tumours and between LKB1-restored and LKB1-deficient mouse cells using chromVAR[27]. For motifs that were conserved across mouse and human datasets, we then compared their differential motif deviation scores (Fig. 2b). Overall, the differences between Chromatin Types 1 and 2 primary tumours were highly concordant with the differences between LKB1-restored and LKB1-deficient mouse lung cancer cells. In particular, genomic regions that contained TEAD and RUNX motifs were more accessible in Chromatin Type 1 tumours and LKB1-restored mouse cells, whereas genomic regions that contained SOX and FOXA motifs were more accessible in Chromatin Type 2 tumours and LKB1-deficient mouse cells (Fig. 2b,c and Extended Data Fig. 5b). These results suggest that not only are LKB1 mutations enriched in Chromatin Type 2 tumours, but also that inactivation of LKB1 is most 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 and H1355). PCA and hierarchical clustering unbiasedly stratified wild-type and mutant LKB1 cell lines on the basis of their chromatin profiles (Fig. 2d and Extended Data Fig. 6a). Genomic regions that contained RUNX and TEAD motifs were more accessible in wild-type LKB1 cell lines, whereas genomic regions that contained SOX and FOXA motifs were more accessible in LKB1-mutant cell lines (Extended Data Fig. 6b,c). Furthermore, similar to the mouse restoration model, the expression of wild-type LKB1 in LKB1-mutant human lung cancer cells markedly 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 ($R = 0.96$)[28] (Extended Data Fig. 6e). Expression of an orthogonal tumour suppressor KEAP1 in KEAP1-mutant cell lines (A549, H460 and H1355) induced very minor chromatin changes (Extended Data Fig. 6f–h), which emphasizes the specificity of the LKB1 tumour suppressor pathway in regulating chromatin accessibility states in lung cancer.

LKB1-driven chromatin accessibility states in mouse primary tumours and metastases

LKB1 deficiency cooperates with oncogenic KRAS in mouse models of lung adenocarcinoma to promote both early-stage tumour growth and late-stage metastasis[4]. To determine whether LKB1 loss has stage-specific effects on tumour progression, we used an in vivo model system to directly compare LKB1-proficient and LKB1-deficient primary tumours and metastases. We incorporated homozygous *Lkb1*-floxed alleles into the metastatic, $Kras^{LSL-G12D/+}$;$Trp53^{flox/flox}$;$Rosa26^{LSL-tdTomato}$ (*KPT*) mouse model to maintain a common genetic background between LKB1-proficient and LKB1-deficient tumours. The administration of lentiviral Cre recombinase into the lungs of *KPT* and *KPT;*$Lkb1^{flox/flox}$ mice led to the development of aggressive primary tumours capable of seeding spontaneous metastases within 4–7 months. Overall, LKB1 deficiency increased the rate of metastatic progression ($P = 0.00016$) (Supplementary Table 5 and Extended Data Fig. 7a–d). We isolated cancer cells from individual primary tumours and metastases by fluorescence activated cell sorting (FACS) and performed ATAC–seq ($n = 12$ *KPT* primary tumours, 13 *KPT;*$Lkb1^{-/-}$ primary tumours, 4 *KPT* metastases, and 5 *KPT;*$Lkb1^{-/-}$ metastases) (Fig. 3a). PCA of the 25 primary tumours stratified samples on the basis of their LKB1 status, similar to the stratification of Chromatin Type 1 and 2 human primary tumours (Fig. 3b). In addition, the motif accessibility differences between LKB1-proficient and LKB1-deficient mouse samples were consistent in directionality with our previous datasets (Extended Data Fig. 7e,f), with SOX motifs more accessible in LKB1-deficient samples and the binding sites for TEAD, RUNX and MEF2 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 tumour chromatin accessibility profiles. Downregulation of the transcription factor *Nkx2-1* has previously been shown to increase metastatic ability in lung adenocarcinoma[29]; similarly, all metastases had decreased local accessibility at the *Nkx2-1* locus, decreased expression of *Nkx2-1* mRNA and decreased accessibility of genomic regions that contain NKX2 motifs compared with primary tumours (Fig. 3c–e and Extended Data Fig. 7g). By contrast, the most prominent genotype-specific difference was that LKB1-deficient metastases had high accessibility of genomic regions

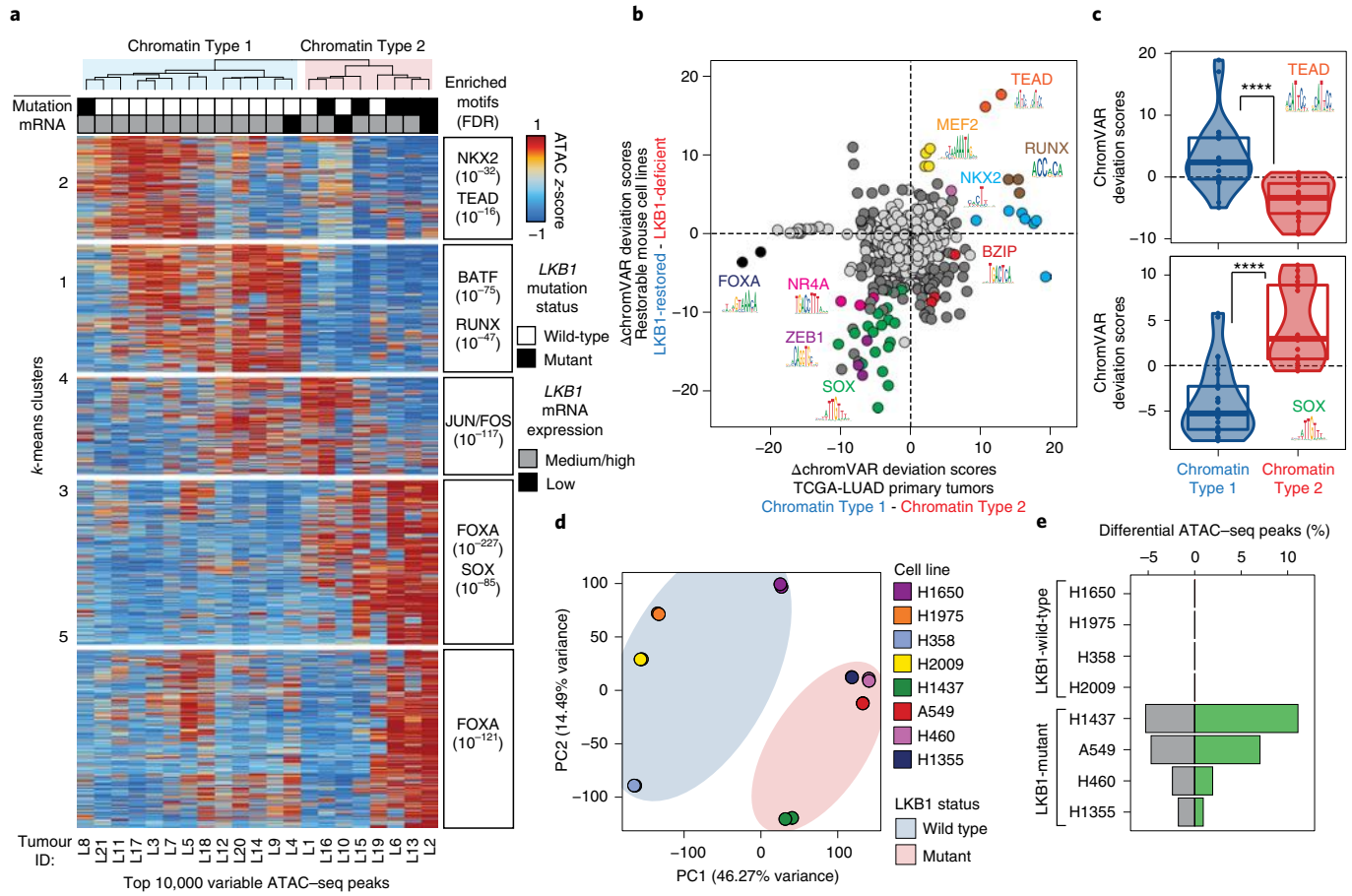


Fig. 2 | LKB1 mutation status distinguishes the two main chromatin sub-types of human lung adenocarcinoma. a, Left: unsupervised hierarchical clustering of 21 human lung adenocarcinoma (LUAD) samples from the TCGA ATAC–seq dataset using the top 10,000 variable peaks across all samples, visualized as a heat map of peak accessibility. Each row represents a *z*-score of $\log_2$-transformed normalized accessibility. Right: transcription factor hypergeometric motif enrichment in each *k*-means cluster, with FDR values indicated in parentheses. **b**, Comparison of the changes in motif accessibility (ΔchromVAR deviation scores) between Chromatin Type 1 and Chromatin Type 2 human primary tumours (*x* axis) and between LKB1-restored and LKB1-deficient mouse cells (*y* axis). Dark grey or coloured points denote significant differences ($q < 0.05$; Methods) across both comparisons. Light grey points are not significant. Only motifs for transcription factors that are shared across human and mouse CIS-BP 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. $^{****}P < 10^{-6}$, two-sided Student's *t*-test. $P = 9 \times 10^{-6}$ for TEAD, and $P = 6 \times 10^{-7}$ for SOX. $n = 13$ and 8 biologically independent samples from Chromatin Types 1 and 2, respectively. 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. **d**, PCA of the top 10,000 variable ATAC–seq peaks across eight human lung cancer cell lines. LKB1 mutant status is indicated. **e**, Percentage of differential ATAC–seq peaks ($|\log_2$-transformed fold change$| > 0.5$, FDR < 0.05) in cells transduced to express LKB1 compared with a GFP control.

that contain 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 tumours expressed low levels of *Sox17* and LKB1-proficient primary tumours 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 tumours, and inaccessible in LKB1-proficient samples (Fig. 3e). Thus, high SOX17 expression and increased accessibility of genomic regions that contain SOX motifs correlate with metastatic progression of LKB1-deficient lung adenocarcinoma. Although 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[30], which suggests that SOX17 can have strong effects on cell state and behaviour.

LKB1-deficient primary tumours contain 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 tumours and metastases. LKB1-proficient primary tumours and metastases were universally SOX17-negative, whereas all LKB1-deficient metastases contained SOX17$^+$ cancer cells (Fig. 4a and Extended Data Fig. 8a,b). In addition, a fraction of LKB1-deficient primary tumours (63 out of 203 tumours) contained sub-populations of SOX17$^+$ cells, primarily located within invasive acinar structured areas (Extended Data Fig. 8a,b). In support of the hypothesis that LKB1 signalling 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 tumours[31] ($R = -0.81$) (Extended Data Fig. 8c).

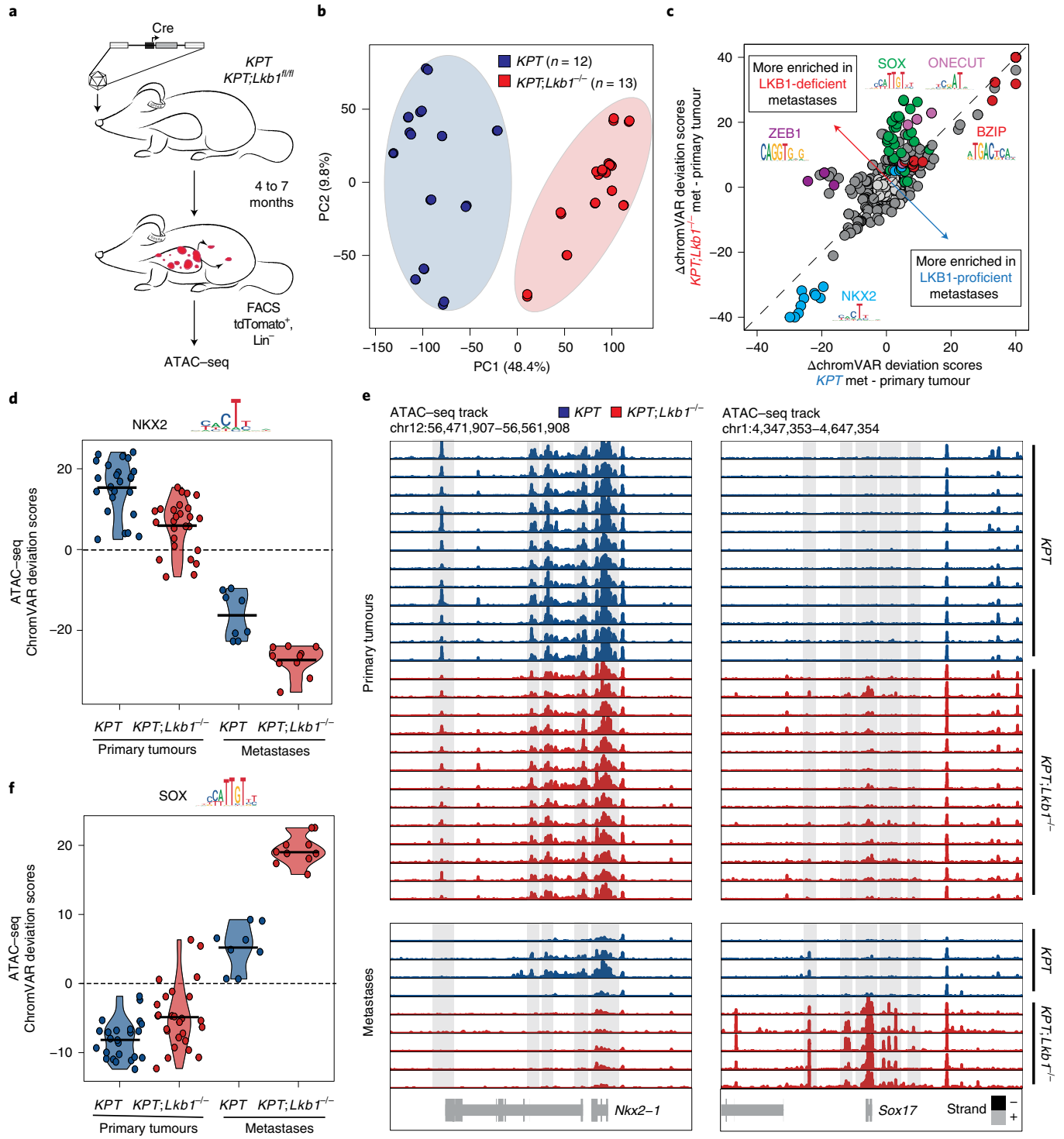


Fig. 3 | Genotype-specific activation of SOX17 expression in metastatic, LKB1-deficient cells. a, Schematic of tumour initiation, sample processing and genomics profiling. The administration of lentiviral Cre initiates tumours in *KPT* mice with and without homozygous *Lkb1*$^{flox/flox}$ alleles. tdTomato^{+} 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 analysis. **b**, PCA of the top 10,000 variable ATAC-seq peaks across 25 primary tumour samples. Technical replicates are averaged. **c**, Comparison of the changes in motif accessibility (ΔchromVAR deviation scores) across LKB1-proficient (*x* axis) and LKB1-deficient (*y* axis) metastases compared with primary tumours of the same genotype. Dark grey and coloured points denote significant differences ($q < 0.05$; Methods) across both comparisons. Light grey points are not significant. **d**, Violin plot of chromVAR deviation scores for NKX2 motifs across LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1*$^{-/-}$) primary tumour and metastasis samples. Black bars represent the mean of all samples. **e**, *Nkx2-1* (left) and *Sox17* (right) genome accessibility tracks for each primary tumour (top) and metastasis (bottom) sample. **f**, Violin plot of chromVAR deviation scores for SOX motifs across LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1*$^{-/-}$) primary tumour and metastasis samples. Black bars represent the mean of all samples.

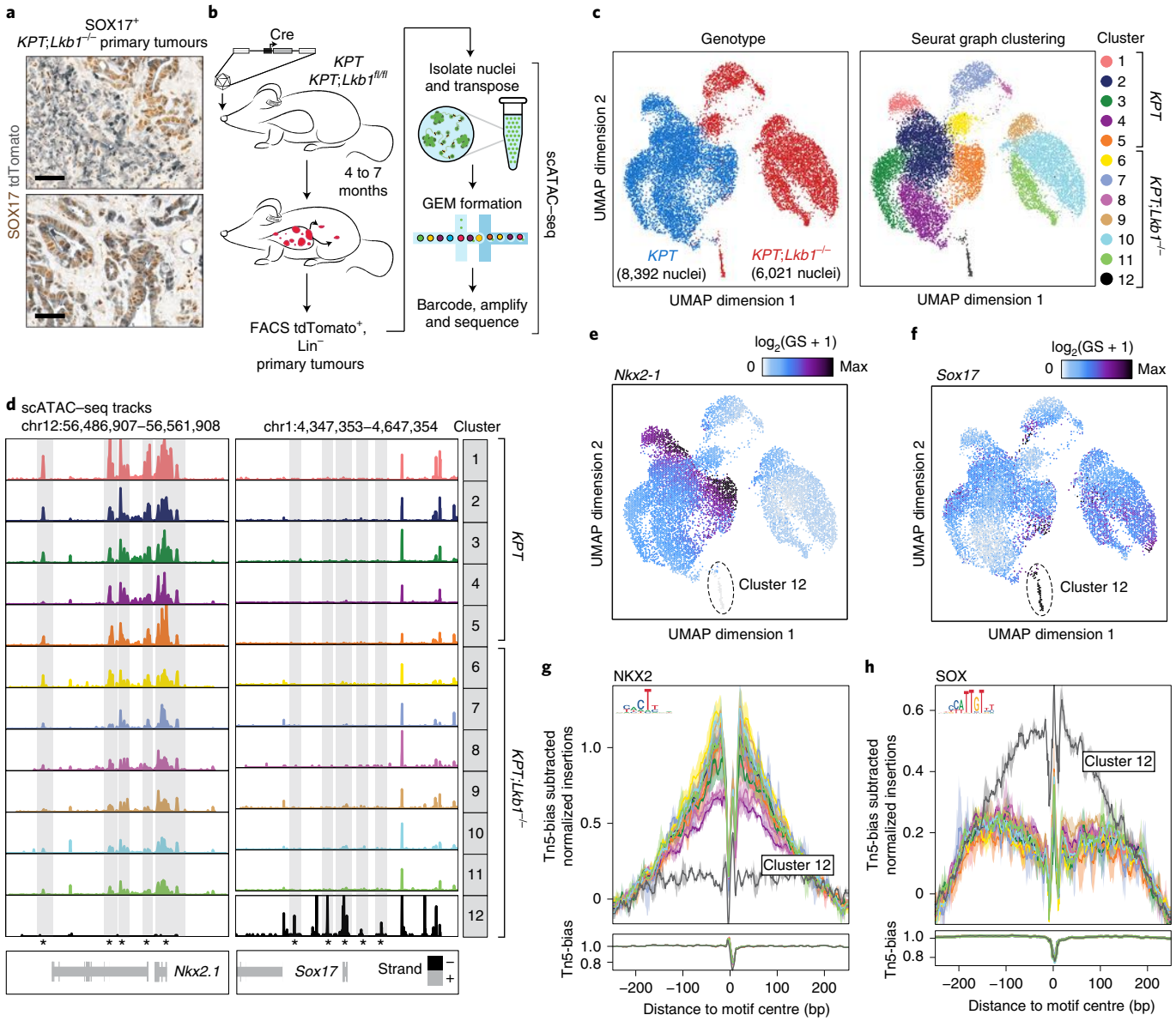


Fig. 4 | LKB1-deficient primary tumours contain sub-populations of SOX17⁺ cells. a, Representative immunohistochemistry for SOX17 (brown) and tdTomato (grey) in two LKB1-deficient lung adenocarcinoma primary tumours. Scale bars, 50 µm. Images are representative of 117 *KPT* primary tumours, 203 *KPT;Lkb1*$^{-/-}$ primary tumours, 14 *KPT* metastases and 8 *KPT;Lkb1*$^{-/-}$ metastases, as quantified in Extended Data Fig. 8b. **b**, Schematic of tumour initiation and processing for scATAC-seq analysis. tdTomato⁺DAPI⁻ cancer cells that were negative for the Lin markers CD45, CD31, F4/F80 and Ter119 were sorted by FACS before scATAC-seq library preparation. GEM, gel-bead emulsion. **c**, Uniform manifold approximation and projection (UMAP) analysis of 8,392 nuclei from 4 *KPT* primary tumours and 6,021 nuclei from 3 *KPT;Lkb1*$^{-/-}$ primary tumours, coloured by genotype (left) or cluster according to Seurat graph clustering (right). Asterisks denote consistency with bulk ATAC-seq differential peaks. **d**, *Nkx2-1* (left) and *Sox17* (right) genome accessibility tracks for each cluster indicated in **c**. Significant ATAC-seq peaks from bulk chromatin accessibility profiling (Fig. 3e) are highlighted in grey and indicated by an asterisk. **e**,**f**, UMAP coloured by the average gene body accessibility for *Nkx2-1* (**e**) or *Sox17* (**f**) in each cell. GS, gene scores. **g**,**h**, Top: footprint of accessibility for each scATAC-seq cluster for genomic regions that contain NKX2 (**g**) and SOX (**h**) motifs. Bottom: modelled hexamer insertion bias of Tn5 around sites that contain each motif.

Furthermore, LKB1-deficient Chromatin Type 2 human primary tumours had higher accessibility at the *SOX17* locus than Chromatin Type 1 human primary tumours (Extended Data Fig. 8d).

To evaluate the epigenetic profiles of SOX17⁺ primary tumour cells, we performed droplet-based single-cell ATAC–seq (scATAC-–seq)[32,33] on cancer cells from LKB1-proficient ($n=4$) and LKB1-deficient ($n=3$) primary tumours (Fig. 4b and Extended Data Fig. 9a,b). We identified 12 distinct clusters of cells (Fig. 4c, Extended Data Fig. 9c and Methods)[34], with clusters 1–5 primarily composed of LKB1-proficient cells and clusters 6–12 mainly consisting 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 (Fig. 4d–f). Cluster 12 mainly consists of cells from two LKB1-deficient primary tumours derived from mouse 13 (13A and 13B). Motif enrichment and transcription factor footprinting[26] 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 tumours compared to LKB1-deficient metastases had the lowest average accessibility in cells in cluster 12 compared with clusters 1–11 (Extended Data Fig. 9e). Genomic regions with the highest accessibility in LKB1-deficient metastases compared to LKB1-deficient primary tumours had the highest average accessibility in cells in cluster 12 compared with clusters 1–11 (Extended Data Fig. 9f). Thus, sub-populations of cancer cells within LKB1-deficient primary tumours 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 markedly 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 that contained SOX-binding sites in human and mouse cell lines after LKB1 restoration (Fig. 1d and Extended Data Figs. 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 was sufficient to maintain high *Sox17* mRNA expression and local chromatin accessibility at the *Sox17* locus after LKB1 restoration (Extended Data Fig. 10c,d). These results suggest that not only do LKB1-deficient metastases express higher levels of SOX17 than 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. 5f). In LKB1-restored cells, *Sox17* expression maintained accessibility at genomic regions containing SOX-binding sites (Extended Data Fig. 10f; cluster 4). We confirmed these results in a second independent cell line (LR1;Cas9) (Extended Data Fig. 5g). Furthermore, expression of a sgRNA-resistant *Sox17* cDNA abrogated the effects of knocking out endogenous *Sox17* (Fig. 5g). Therefore, SOX17 is necessary and sufficient to maintain accessibility at genomic regions that contain SOX-binding sites in LKB1-deficient, metastatic cells.

SOX17 drives tumour 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 3 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 tumour burden relative to an sgSafe control (Fig. 6b,c and Extended Data Fig. 10g). By contrast, overexpressing *Sox17* in LKB1-restored cells increased tumour burden (Fig. 6b,c and Extended Data Fig. 10h). In addition, to evaluate the ability of *Sox17*-overexpressing cells to both leave the primary tumour and establish metastases, we injected LKB1-restored cells with and without overexpressed *Sox17* cDNA subcutaneously into recipient mice (Fig. 6d). After 5 weeks of growth, we evaluated cells that had left the subcutaneous 'primary' tumour and colonized the lung (Fig. 6e). Although overexpression of *Sox17* did not change subcutaneous tumour growth (Fig. 6f, top), *Sox17*-overexpressing cells had a significantly greater ability to colonize the lung (Fig. 6f, bottom). Furthermore, 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 3 weeks of growth, *Sox17*-overexpressing cells were more capable of colonizing the liver than wild-type cells ($P=0.055$) (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 tumour-suppressive kinase, drives widespread chromatin accessibility changes in lung adenocarcinoma that evolve throughout cancer progression. Although LKB1 has been well-studied for its roles in cancer metabolism, LKB1-induced chromatin changes are surprisingly AMPK-independent and depend almost exclusively on the SIK family of kinases. Recent studies have also shown that deleting AMPK hinders rather than helps lung cancer growth, and AMPK1 is preferentially amplified in lung adenocarcinoma, which suggests that AMPK is not a classic tumour suppressor in this cancer type[5,34]. Therefore, SIKs are emerging as the main drivers of LKB1-mediated tumour suppression and epigenetic regulation in lung cancer. Notably, the SIK family of kinases has a known role in the inhibition of class IIa histone deacetylases (HDACs)[35]. 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[36]. Thus, the relationship between SIK and HDACs might be relevant for 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 and SIK signalling drives two separate waves of epigenetic remodelling, with the first set of changes occurring within lung primary tumours 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 tumour development and subsequently enhance metastatic ability. Although the LKB1 pathway probably 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 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 tumours[4,37], 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 tumours at single-cell resolution, we further discovered sub-populations of cancer cells in primary tumours that share a common epigenetic state with the cancer cells in metastases. This result suggests that primary tumours contain 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

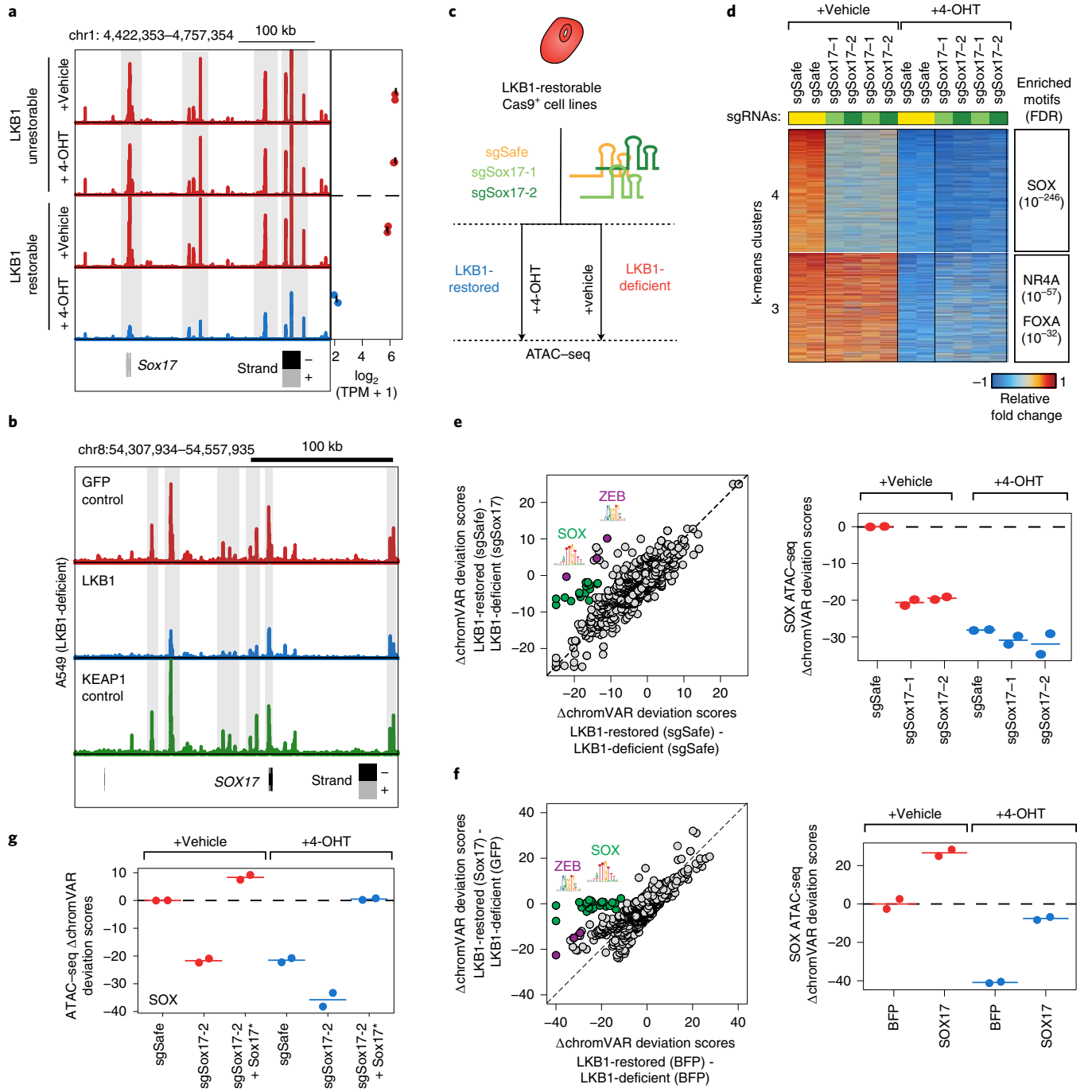


Fig. 5 | SOX17 regulates the chromatin accessibility state of metastatic, LKB1-deficient cells. a, *Sox17* genome accessibility track (left) and mean *Sox17* mRNA expression across technical replicates (right) of an LKB1-unrestorable cell line (LU2) and a metastatic LKB1-restorable cell line (LR2) treated with 4-OHT or vehicle for 6 days. Significantly differential ATAC-seq peaks (log_2-transformed fold change < −0.5, FDR < 0.05) after LKB1 restoration are highlighted in grey. *Sox17* mRNA expression was also significantly decreased after LKB1 restoration (log_2-transformed fold change < −1, FDR < 0.05). TPM, transcripts per million. **b**, *SOX17* genome accessibility track of an LKB1-deficient cell line (A549) transduced with a GFP- (top), LKB1- (middle) or KEAP1-expressing lentiviral vector (bottom). **c**, Schematic of knocking out SOX17 with and without LKB1 restoration followed by ATAC-seq. **d**, Heat map of the relative log_2-transformed fold changes in *k*-means clusters 3 and 4 of the indicated genotypes of cells with and without LKB1 restoration compared with the average log_2-transformed fold changes in sgSafe control cells. A subset (5,379 peaks; all decreasing peaks) of the top 10,000 consistent, variable ATAC-seq peaks after LKB1 restoration in cells transduced with either sgSafe or blue fluorescent protein (BFP) controls are shown. See Extended Data Fig. 10f for the full heat map. **e**,**f**, Left: comparison of the changes in motif accessibility (ΔchromVAR deviation scores) between the indicated perturbated populations. Dark grey or coloured points denote significant differences ($q < 0.05$) 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, bars represent 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, bars represent 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 (Methods). $n = 2$ technical replicates for each condition.

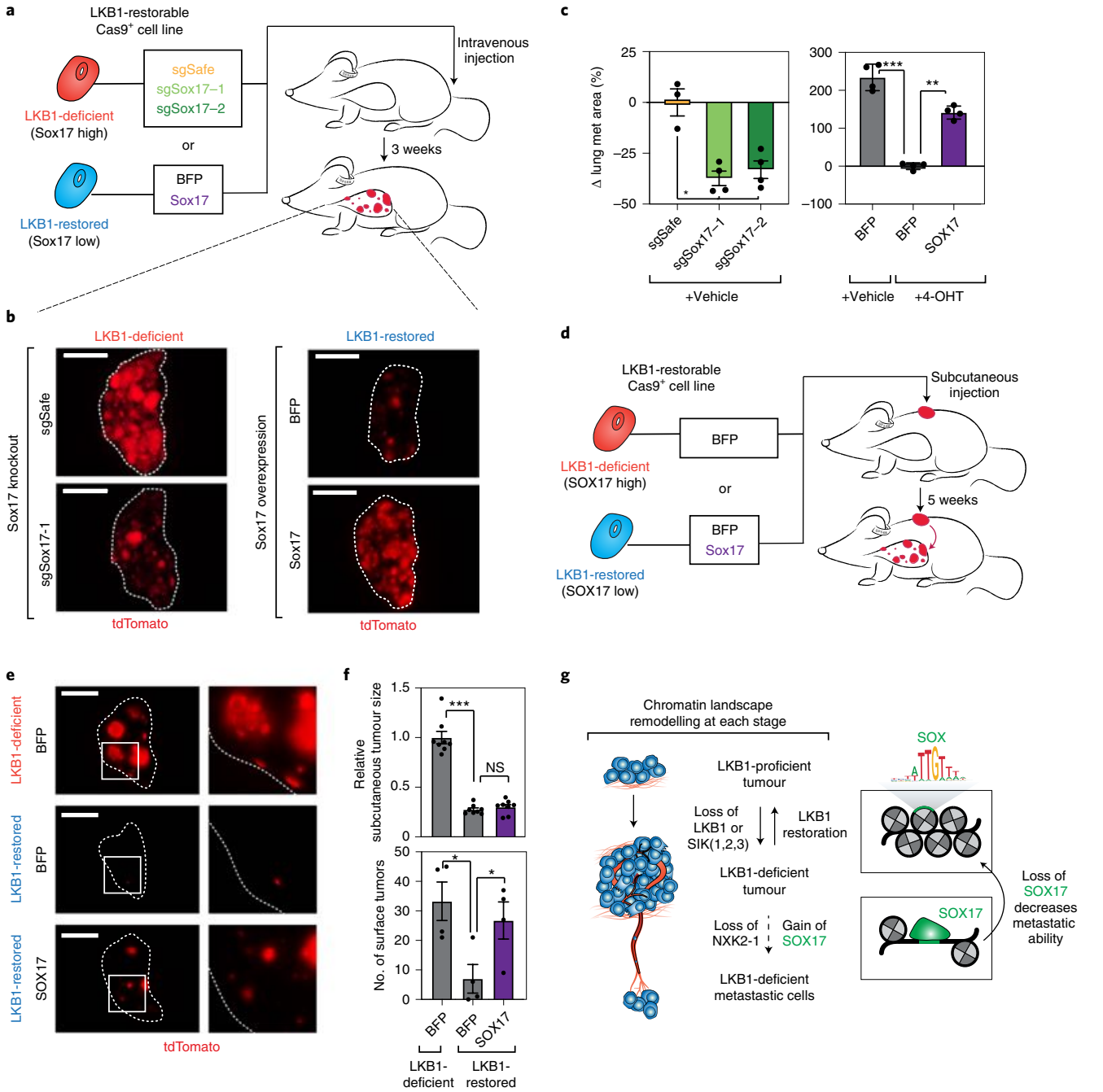


Fig. 6 | SOX17 regulates the metastatic ability of LKB1-deficient cells. a, Schematic of injecting LKB1-deficient cells expressing sgRNAs that target *Sox17* (sgSox17-1 and sgSox17-2) or injecting LKB1-restored cells expressing *Sox17* cDNA intravenously into immunocompromised non-obese diabetic/severe combined immunodeficiency (NOD/SCID) gamma (NSG) mice. Tumour burden was analysed 3 weeks after injection. **b**, Representative fluorescent tdTomato$^+$ images of single lung lobes after intravenous injection. Similar results were observed from three additional mice (four mice in total) per condition, except for sgSafe + vehicle, in which similar results were observed from two additional mice (three mice in total). Scale bars, 5 mm. **c**, Change in the percentage tumour area compared with LKB1-deficient cells (sgSafe + vehicle, left) or with LKB1-restored cells (BFP + 4-OHT, right). Each point represents an individual mouse. Data are mean ± s.e.m. $^*P<0.01$, $^{**}P<0.001$, $^{***}P<0.0001$, two-sided *t*-test. $n=3$ (sgSafe + vehicle) and $n=4$ (all other conditions) biologically independent samples per condition. $P=0.0031$ for sgSox17-1 versus sgSafe; $P=0.0072$ for sgSox17-2 versus sgSafe; $P=0.0031$ for BFP + 4-OHT versus SOX17 + 4-OHT; $P<0.0001$ for BFP + 4-OHT versus BFP + vehicle. **d**, Schematic of injecting LKB1-deficient cells (LR2) expressing BFP or injecting LKB1-restored cells (LR2) expressing *Sox17* cDNA or BFP subcutaneously into immunocompromised NSG mice. Metastatic tumour burden to the lung was analysed 5 weeks after injection. **e**, Representative fluorescent tdTomato$^+$ images of single lung lobes after subcutaneous injection as outlined in **d**. Similar results were observed from three additional mice (four mice in total) per condition. Scale bars, 5 mm. **f**, Top: relative tumour size after subcutaneous injection of the indicated cells. Bottom: number of surface tumours observed in the five lung lobes after subcutaneous injection of the indicated cells. Data are mean ± s.e.m. $^*P<0.05$, $^{***}P=0.0001$, two-sided *t*-test. $P=0.0178$ for BFP–4-OHT versus BFP–vehicle; $P=0.0471$ for BFP–4-OHT versus SOX17–4-OHT. NS, not significant. $n=4$ biologically independent mice evaluated per condition with 2 subcutaneous tumours each. **g**, Summary of LKB1-induced chromatin accessibility changes in primary tumours and metastases.

possibility of identifying biomarkers to predict which tumours have already seeded micrometastases before detection is possible. In addition, we anticipate that genotype-driven epigenetic differences between primary tumours and metastases may inform how patients respond to therapies. Overall, these findings help to explain the paradox in which primary tumours and metastases share the same genetic mutations yet exhibit extremely different behaviours, and we anticipate that an evolving mechanism of tumour suppression is more broadly applicable to other commonly mutated driver genes and cancer types.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-021-00728-4.

Received: 20 July 2020; Accepted: 5 July 2021;
Published online: 2 August 2021

References

1. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
2. Waddell, N. et al. Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
3. Sanchez-Cespedes, M. A role for LKB1 gene in human cancer beyond the Peutz-Jeghers syndrome. *Oncogene* **26**, 7825–7832 (2007).
4. Ji, H. et al. LKB1 modulates lung cancer differentiation and metastasis. *Nature* **448**, 807–810 (2007).
5. Carretero, J. et al. Integrative genomic and proteomic analyses identify targets for Lkb1-deficient metastatic lung tumors. *Cancer Cell* **17**, 547–559 (2010).
6. Shackelford, D. B. & Shaw, R. J. The LKB1-AMPK pathway: metabolism and growth control in tumour suppression. *Nat. Rev. Cancer* **9**, 563–575 (2009).
7. Jin, L. et 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 (2018).
8. Calles, A. et al. Immunohistochemical loss of LKB1 is a biomarker for more aggressive biology in KRAS-mutant lung adenocarcinoma. *Clin. Cancer Res.* **21**, 2851–2860 (2015).
9. Lizcano, J. M. et al. LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. *EMBO J.* **23**, 833–843 (2004).
10. Kottakis, F. et al. LKB1 loss links serine metabolism to DNA methylation and tumorigenesis. *Nature* **539**, 390–395 (2016).
11. Hollstein, P. E. et al. The AMPK-related kinases SIK1 and SIK3 mediate key tumor-suppressive effects of LKB1 in NSCLC. *Cancer Discov.* **9**, 1606–1627 (2019).
12. Murray, C. W. et al. An LKB1-SIK axis suppresses lung tumor growth and controls differentiation. *Cancer Discov.* **9**, 1590–1605 (2019).
13. Pierce, S. E., Granja, J. M. & Greenleaf, W. J. High-throughput single-cell chromatin accessibility CRISPR screens enable unbiased identification of regulatory networks in cancer. *Nat. Commun.* **12**, 2969 (2021).
14. Filbin, M. G. et al. Developmental and oncogenic programs in H3K27M gliomas dissected by single-cell RNA-seq. *Science* **360**, 331–335 (2018).
15. Flavahan, W. A. et al. Altered chromosomal topology drives oncogenic programs in SDH-deficient GISTs. *Nature* **575**, 229–233 (2019).
16. LaFave, L. M. et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell* **38**, 212–228 (2020).
17. Reiter, J. G. et al. Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).
18. 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).
19. Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169–175 (2016).
20. Robles-Oteiza, C. et al. Recombinase-based conditional and reversible gene regulation via XTR alleles. *Nat. Commun.* **6**, 8783 (2015).
21. Morgens, D. W. et al. Genome-scale measurement of off-target activity using Cas9 toxicity in high-throughput screens. *Nat. Commun.* **8**, 15178 (2017).
22. Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. *Nucleic Acids Res.* **47**, D419–D426 (2019).
23. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. 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).
24. Corces, M. R. et al. An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* **14**, 959–962 (2017).
25. Corces, M. R. et al. Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. *Nat. Genet.* **48**, 1193–1203 (2016).
26. Corces, M. R. et al. The chromatin accessibility landscape of primary human cancers. *Science* **362**, eaav1898 (2018).
27. Schep, A. N., Wu, B., Buenrostro, J. D. & Greenleaf, W. J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat. Methods* **14**, 975–978 (2017).
28. Kaufman, J. M. 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).
29. Winslow, M. M. et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature* **473**, 101–104 (2011).
30. Park, K.-S., Wells, J. M., Zorn, A. M., Wert, S. E. & Whitsett, J. A. Sox17 influences the differentiation of respiratory epithelial cells. *Dev. Biol.* **294**, 192–202 (2006).
31. Laughney, A. M. et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat. Med.* **26**, 259–269 (2020).
32. Satpathy, A. T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat. Biotechnol.* **37**, 925–936 (2019).
33. Granja, J. M. et al. Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat. Biotechnol.* **37**, 1458–1465 (2019).
34. Stuart, T. et al. Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902 (2019).
35. Walkinshaw, D. R. 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).
36. Parra, M. Class IIa HDACs - new insights into their functions in physiology and pathology. *FEBS J.* **282**, 1736–1744 (2015).
37. Zhang, H. et al. Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. *Nat. Commun.* **8**, 14922 (2017).

Methods

Mouse cell lines. Mouse cell lines were generated from individual primary tumours and metastases from *Kras*$^{LSL-G12D}$;*Trp53*$^{flox/flox}$;*Lkb1*$^{XTR/XTR}$;*Rosa26*$^{LSL-tdTomato}$ (cell lines LU1 and LU2), *Kras*$^{LSL-G12D}$;*Trp53*$^{flox/flox}$;*Lkb1*$^{XTR/XTR}$;*Rosa26*$^{FlpOER/LSL-tdTomato}$ (cell line LR2) and *Kras*$^{LSL-G12D}$;*Trp53*$^{flox/flox}$;*Lkb1*$^{XTR/XTR}$;*Rosa26*$^{FlpOER/+}$ (cell line LR1) mice previously transduced with lentiviral Cre. The *Lkb1*$^{XTR/XTR}$ mouse allele has been deposited at The Jackson Laboratory (034052) and was generated using the same design and methods as outlined for the *Tp53*$^{XTR/XTR}$ allele[20]. All cell lines have gene expression patterns consistent with being in a metastatic state (*Nkx2-1*^{low};*Hmga2*^{high}) (Supplementary Table 3). To derive cell lines, tumours were excised from the lungs or lymph nodes of mice, minced into pieces using scissors, and directly cultured in DMEM medium supplemented with 10% FBS, 1% penicillin-streptomycin-glutamate and 0.1% amphotericin at 37 °C with 5% CO_2 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 mouse cell lines (LR1, LR2, LU1 and LU2) were grown in DMEM medium 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 (10 μg ml^{-1}) 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), because these cell lines were previously transduced with Cre recombinase to initiate tumour 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). To induce LKB1 restoration, cells were treated with either 1 μM 4-OHT (Sigma Aldrich) dissolved in 100% ethanol or a vehicle (1:2,000 100% ethanol) for the indicated time points.

Proliferation doubling assays. For population doubling assays, cell lines were treated with 4-OHT or vehicle for 12 days. Every other day, cells were trypsinized, counted and re-plated with 50,000 cells per well of a 6-well plate 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 that is $\log_2(n/50{,}000)$. 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 6 days. Cells were trypsinized, counted and re-plated at 500 cells per well of a 6-well plate in triplicate. Plates were incubated at 37 °C with 5% CO_2 for 6 days. For analysis, cells were rinsed with PBS at room temperature, fixed with ethanol for 5 min at room temperature and stained with 1% crystal violet solution in water (Millipore-Sigma) for a further 5 min. Plates were rinsed with water, scanned into the computer and analysed using ImageJ (1.52q with Java 1.8.0_172, available at https://imagej.nih.gov/ij/). The percentage area of the plate covered by cells was normalized to the average percentage area of the plate covered by cells treated with a vehicle control. Two-tailed t-tests were performed to determine statistical significance.

RNA-seq library preparation for cell lines. Cell lines were treated with 4-OHT or vehicle for 6 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 of 10.0. Total RNA (500 ng) for each sample was processed into libraries using the TruSeq RNA Library Prep Kit v2 (Illumina) and sequenced according to standard protocols.

RNA-seq data processing and alignment. RNA-seq data were trimmed with CutAdapt and aligned with kallisto[38]. 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 transcripts per million) are available in Supplementary Tables 3 and 6.

RNA-seq data analysis, differential expression. To compute differential gene expression, we used the edgeR 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 $\log_2$-transformed fold changes and adjusted P values. We chose the indicated significance cut-off values 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 13,000*g* for 30 min at 4 °C. The concentration of protein-containing supernatant was quantified using the Pierce BCA Protein Assay Kit (Thermo Fisher). Ten micrograms 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 10 V overnight. Blocking, primary and secondary antibody 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, 1:2,000 dilution) or anti-mouse (Santa Cruz Biotechnology, sc-2005, 1:2,000 dilution) antibodies. LKB1 (Cell Signaling, 13031S, 1:1,000 dilution) and SOX17 (Abcam, ab224637, 1:500 dilution) protein expression was assessed by western blotting. HSP90 (BD Biosciences, 610418, 1:2,000 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 10- to 12-week-old immunocompromised NSG mice. For intravenous transplants, cells were treated with either a vehicle control or 4-OHT for 6 days and 5×10^4 cells were injected into one of the lateral tail veins. Mice were euthanized 21 days after 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 6 days and 5×10^4 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. After injection of cells, a surgical knot was made in the top part of the spleen and the bottom part was removed before the body wall back was sewn back with surgical knots. The skin incision was closed with staples and antiseptic solution was applied to clean the wound. Mice were euthanized 3 weeks after 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–k, or $n=17$ mice total). For the initial subcutaneous transplants, 2×10^5 untreated cells were resuspended in 200 μl PBS and injected into two sites per mouse ($n=2$–5 female NSG mice per condition for Extended Data Fig. 1f, $n=25$ mice in total). Once tumours were readily palpable, mice were randomized and treated via oral gavage with either a vehicle control (200 μl 10% ethanol 90% corn oil) or tamoxifen (200 μl of 20 mg ml^{-1} tamoxifen dissolved in 10% ethanol 90% corn oil) (Sigma Aldrich) for 3 consecutive days. The height, width and length of each tumour was measured using calipers every 2 days for 14 days (LU1 and LR2) or every 4 days for 16 days (LU2 and LR1). Tumour volume was roughly calculated by multiplying height × width × length of each tumour. During the experimental time course, tumour burden never exceeded 1.7 cm^3 per mouse, which is the maximal tumour burden allowed by our ethics committee. For subcutaneous transplants to model metastatic spread to the lung, 5×10^4 cells of the indicated genotypes pre-treated with 6 days of 4-OHT or vehicle were resuspended in 200 μl PBS and injected into two sites per mouse. Mice were euthanized 5 weeks after 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. Haematoxylin and eosin staining was performed using standard methods and the percentage tumour area was calculated using ImageJ. For immunohistochemistry, we used an antibody to SOX17 (Abcam, ab224637) at a 1:1,000 dilution. Heat-mediated antigen retrieval was performed in Tris-EDTA buffer at pH 9.0. To evaluate SOX17 expression, we quantified the number of tumours with a tumour 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 polyethylenimine (Polysciences). The viral-containing supernatant was collected at 48 and 72 h after transfection, filtered through a 0.45-μm filter, and combined with fresh medium to transduce cells for up to 2 days. Human cell lines were incubated with 8 μg ml^{-1} polybrene (Sigma) to enhance transduction efficiency.

CRISPR–Cas9 screen and sample processing. The genome-scale CRISPR–Cas9 knockout library was synthesized by Agilent and designed and cloned as previously described[21]. 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 h were selected with puromycin (8 μg ml^{-1}) 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 200× cell number coverage per sgRNA. Owing to the fast doubling time of this cell line, each half of the screen required passaging >165 15 cm dishes every 2 days. Then, 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 duplicates 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[21].

CRISPR–Cas9 screen data alignment and analysis. We aligned each half of the genome-scale CRISPR–Cas9 screen individually using casTLE[39], 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 computed the mean reads of these sgRNAs. We scaled each half of the screen such that the mean reads in overlapping sgRNAs were identical. For sgRNAs that were specific to one half of the screen, we used the normalized values from only that half of the screen. For sgRNAs that were present in both halves of the screen, we used the average normalized reads across both halves. We further depth-normalized across all samples. We computed the $\log_2$-transformed correlations and plotted the Pearson correlation matrix in R. To quantify the sgRNAs that were enriched/deficient in the screen we used model-based analysis of genome-wide CRISPR–Cas9 knockout (MAGeCK)[40]. In brief, we used MAGeCK test with parameters '-k counts.tsv -t day12_LKB1_Restored -c day12_LKB1_Unrestored'. We then accessed the MAGeCK robust ranking aggregation (RRA) scores from gene_summary.txt file and filtered targets with fewer than five 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 shown in Supplementary Table 1.

ATAC–seq library preparation for cell lines. Cell lines were treated with 4-OHT or vehicle for the indicated time points before 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 medium for all cells was changed at each time point to control for fluctuations in growth factors or other medium contents between samples. For all experiments, adherent cells were rinsed with PBS, trypsinized for 5 min at 37 °C, spun down and resuspended in PBS. Approximately 50,000 cells in technical duplicate were resuspended in 250 μl PBS and centrifuged at 500*g* for 5 min at 4 °C in a fixed-angle centrifuge. Pelleted cells were resuspended in 50 μl ATAC–seq resuspension buffer (RSB; 10 mM Tris-HCl pH 7.4, 10 mM NaCl and 3 mM $MgCl_2$ in ddH_2O made fresh) containing 0.1% NP40, 0.1% Tween-20 and 0.01% digitonin according to the omni-ATAC–seq protocol[24]. After incubating on ice for 3 min, 1 ml of ATAC–seq RSB containing 0.1% Tween-20 was added. Nuclei were centrifuged at 500*g* for 5 min at 4 °C in a fixed-angle centrifuge, 900 μl of the supernatant was taken off, and the nuclei were centrifuged for an additional 5 min under the same conditions. The remaining 200 μl of supernatant was aspirated and nuclei were resuspended in 50 μl of transposition mix (25 μl 2× TD buffer (2 ml 1 M Tris-HCl, pH 7.6, 1 ml 1 M $MgCl_2$, 20 ml DMF and 77 ml ddH_2O aliquoted and stored at −20 °C), 2.5 μl transposase (100 nM final), 16.5 μl PBS, 0.5 μl 1% digitonin, 0.5 μl 10% Tween-20 and 5 μl ddH_2O). Transposition reactions were incubated at 37 °C for 30 min with 1,000 rpm 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 20 μl at 4 nM. Paired-end sequencing was performed on an Illumina NextSeq using 75 cycle kits.

ATAC–seq data processing and alignment. Adapter 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. Further information on detailed ATAC–seq data analysis can be found in the Supplemental Information. All available raw sequencing data, aligned fragments and bigwigs for all ATAC–seq samples is provided in Supplementary Table 2 using AWS. Furthermore, all matrices (peak matrix and chromVAR) are available in Supplementary Table 2 using AWS.

ATAC–seq data analysis, TCGA LUAD tumours. We downloaded the TCGA ATAC–seq data from https://gdc.cancer.gov/about-data/publications/ATACseq-AWG for tumours with matched RNA. We then scored each tumour as being high (top 10%), medium (middle 80%) or low (bottom 10%) in LKB1 expression (TPM). We also identified which tumours had a medium–high predicted mutation (VarScan2). We then unbiasedly identified the top 10,000 variable peaks and grouped them into five *k*-means clusters. We then plotted a heat map of the scaled $\log_2$-transformed accessibility as described above. To test the enrichment of specific gene mutations in each chromatin subtype (Extended Data Fig. 5a), we computed the proportion of medium–high predicted mutation burden of the gene (VarScan2) and computed a binomial enrichment versus the mutation frequency of all TCGA LUAD tumours ($n=230$). We then computed the FDR for the binomial enrichments with adjusted *P* values in R.

scRNA–seq analysis, LUAD metastases[31]. We downloaded the raw data from Laughney et al.[31] from 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. *z*-scores were then computed for all genes. We then averaged the scaled expression for all cells from each donor that belonged in cluster 'H0' and 'H3' (to increase the number of donors), which represent the most undifferentiated metastatic cells. The s.e.m. was computed for all cells from each donor in these clusters, and the mean and s.e.m. of SOX17 expression versus LKB1 expression was plotted for each of the donors.

Cloning and generating knockout and overexpression cell lines. To generate individual knockout 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 shown in Supplementary Table 4. sgRNA sequences were chosen on the basis of 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 2 days and then selected cells with 8 μg ml^{-1} puromycin for at least 3 days to enrich for cells transduced with the lentivirus, before initiating treatment with vehicle or 4-OHT.

To generate double and triple knockout cell lines, we used Gibson assembly to create lentiviral vectors with sgRNAs transcribed in series by the bovine U6 promoter, human U6 promoter and mouse U6 promoter, as previously described[41]. 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 mouse 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 2 days and then selected cells with 8 μg ml^{-1} puromycin for at least 3 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 mouse *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 2 days and then selected mouse cell lines with 8 μg ml^{-1} puromycin for at least 3 days to enrich for cells transduced with the lentivirus before initiating treatment with vehicle or 4-OHT. To increase the *Sox17* knockout efficiency of sgSox17-1 and sgSox17-2 cell lines, these cell lines were also 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* 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 2 μg ml^{-1} puromycin for at least 4 days to enrich for cells transduced with lentivirus, then let the cells recover in fresh medium 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) and NCI-H2009 (ATCC CRL-5911)) were either purchased directly from ATCC or were a gift from M. Bassik's laboratory, which had previously purchased them from ATCC. Human

cell lines were cultured in RPMI medium 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*$^{fl/fl}$) were bred into the metastatic *KPT* (*Kras*$^{LSL-G12D}$;*Trp53*$^{fl/fl}$;*Rosa26*$^{LSL-tdTomato}$) 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 polyethylenimine, the supernatant was collected at 48 and 72 h after transfection, ultracentrifuged at 25,000*g* for 90 min and resuspended in PBS. Tumours were initiated by intratracheal transduction of 10- to 12-week-old mice with lentiviral vectors that express Cre recombinase[42]. For ATAC–seq, tumours were collected and processed at staggered time points at which the tumour burden was similar between *KPT* and *KPT*;*Lkb1*$^{fl/fl}$ cohorts of mice ($n = 11$ female *KPT* mice, 3 male *KPT* mice, 2 female *KPT*;*Lkb1*$^{fl/fl}$, and 6 male *KPT*;*Lkb1*$^{fl/fl}$ mice). For the survival curve, mice were euthanized immediately after exhibiting physical symptoms of distress from lung tumour burden.

Tumour dissociation, cell sorting and freezing. Primary tumours and metastases were individually microdissected and dissociated using collagenase IV (Thermo Fisher), dispase (Corning) and trypsin (Invitrogen) at 37 °C for 30 min. After dissociation, the samples remained on ice and in the presence of 2 mM EDTA (Promega) and 1 U ml^{-1} 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 haematopoietic and endothelial linage (Lin$^+$) cells. DAPI was used to exclude dead cells. BD FACSAria sorters (BD Biosciences) were used for cell sorting. tdTomato$^+$Lin$^-$DAPI$^-$ cells were sorted by FACS into microcentrifuge tubes, spun down at 500*g* for 5 min at 4 °C in a fixed-angle centrifuge, resuspended in 250 µl freezing medium (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–seq library preparation for primary tumours and metastases. FACS-isolated samples were taken out of storage in liquid nitrogen, quickly thawed at 37 °C, diluted with 1 ml PBS and centrifuged at 300*g* for 5 min at 4 °C in a fixed-angle centrifuge. Primary tumours 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 fewer than 50,000 cells. For example, for a sample with 10,000 cells, one-fifth of the normal amount of transposase was used in the 50 µl transposition reaction. The remaining volume was replaced with ddH_2O.

RNA-seq library preparation for primary tumours and metastases. RNA was extracted from sorted cancer cells using the AllPrep DNA/RNA Micro Kit (Qiagen). The RNA quality of each tumour 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 analysis had an RNA integrity number > 8.0. RNA-seq libraries were generated as previously described[42] and sequenced using 200 cycle kits on an Illumina HiSeq 2000.

scATAC–seq library preparation for primary tumours and metastases. FACS-isolated samples were taken out of storage in liquid nitrogen, quickly thawed at 37 °C, diluted with 1 ml PBS and centrifuged at 300*g* for 5 min at 4 °C in a fixed-angle centrifuge. Cells were resuspended in PBS and 0.04% BSA, passed through a 40-µm Flowmi cell strainer (Sigma) to minimize cell debris, and the cell concentration was determined. Primary tumours 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 were converted to fastq format using cellranger atac mkfastq (10x Genomics, version 1.2.0). scATAC–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 from: 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 using AWS.

ArchR for scATAC–seq data analysis. We used ArchR[43] for all downstream scATAC–seq analysis (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'. In addition, we filtered cells that had a TSS enrichment below 6, fewer than 1,000 fragments or more than 50,000 fragments. Dimensionality was reduced with 'addIterativeLSI' excluding chrX and chrY from this analysis. We then added clusters with 'addClusters' with a resolution of 0.4. A UMAP was added 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', and quantified the number of Tn5 insertions per peak per cell using 'addPeakMatrix'. Motif annotations were added using 'addMotifAnnotations' with chromVAR mouse motifs version 1 'mouse_pwms_v1'. We then computed chromVAR deviations for each single cell with 'addDeviationsMatrix'. For transcription factor footprinting of NKX2-1 and SOX17 we used 'plotFootprints' with normalization method 'subtract' which substracts the Tn5 bias from the ATAC footprint.

To further characterize the 12 scATAC–seq clusters on the basis of their metastatic state, we computed differential peaks from the LKB1-deficient bulk primary tumours and metastases. We took significantly differential peaks (|$\log_2$-transformed fold change| > 3 and FDR < 0.01) and overlapped these peaks with the scATAC–seq peaks. The mean accessibility and s.e.m. across these overlapping regions was then plotted for peaks specific to primary tumours and peaks specific to metastases independently.

Statistics and reproducibility. Experimental data were plotted and analysed 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.

Material availability. Plasmids generated in this study are available upon request (M.M.W.).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

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 dataset derived from this resource that supports the findings of this study is publicly available at https://gdc.cancer.gov/about-data/publications/ATACseq-AWG. All other data supporting the findings of this study are available from the corresponding authors on request. Transcription factor binding motifs were derived from CIS-BP (http://cisbp.ccbr.utoronto.ca/index.php). Source data are provided with this paper.

Code availability

All custom code used in this work is available from the corresponding authors upon request. We also host a Github website that includes the main analysis code used in this study (https://github.com/GreenleafLab/LKB1_2021)[44].

References

38. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
39. Morgens, D. W., Deans, R. M., Li, A. & Bassik, M. C. Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat. Biotechnol.* **34**, 634–636 (2016).
40. Li, W. et al. MAGeCK enables robust identification of essential genes from genome-scale CRISPR/Cas9 knockout screens. *Genome Biol.* **15**, 554 (2014).
41. Adamson, B. et al. A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882 (2016).
42. 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).
43. Granja, J. M. et al. ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nat. Gen.* **53**, 403–411 (2021).
44. Granja, J. M. GreenleafLab/LKB1_2021: Release_1.0.1 *Zenodo* https://doi.org/10.5281/zenodo.5035694 (2021).

Acknowledgements

We thank J. Sage, A. Trevino and members of the Greenleaf and Winslow laboratories for comments. We thank the Stanford Shared FACS facility and the Veterinary Service Center 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 (grant number T31DT1900). This work was supported by National Institutes of Health (NIH) grant numbers R01-CA204620 and

R01-CA230919 (to M.M.W.), 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 grant P30-CA124435).

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 contributions 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, which 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.

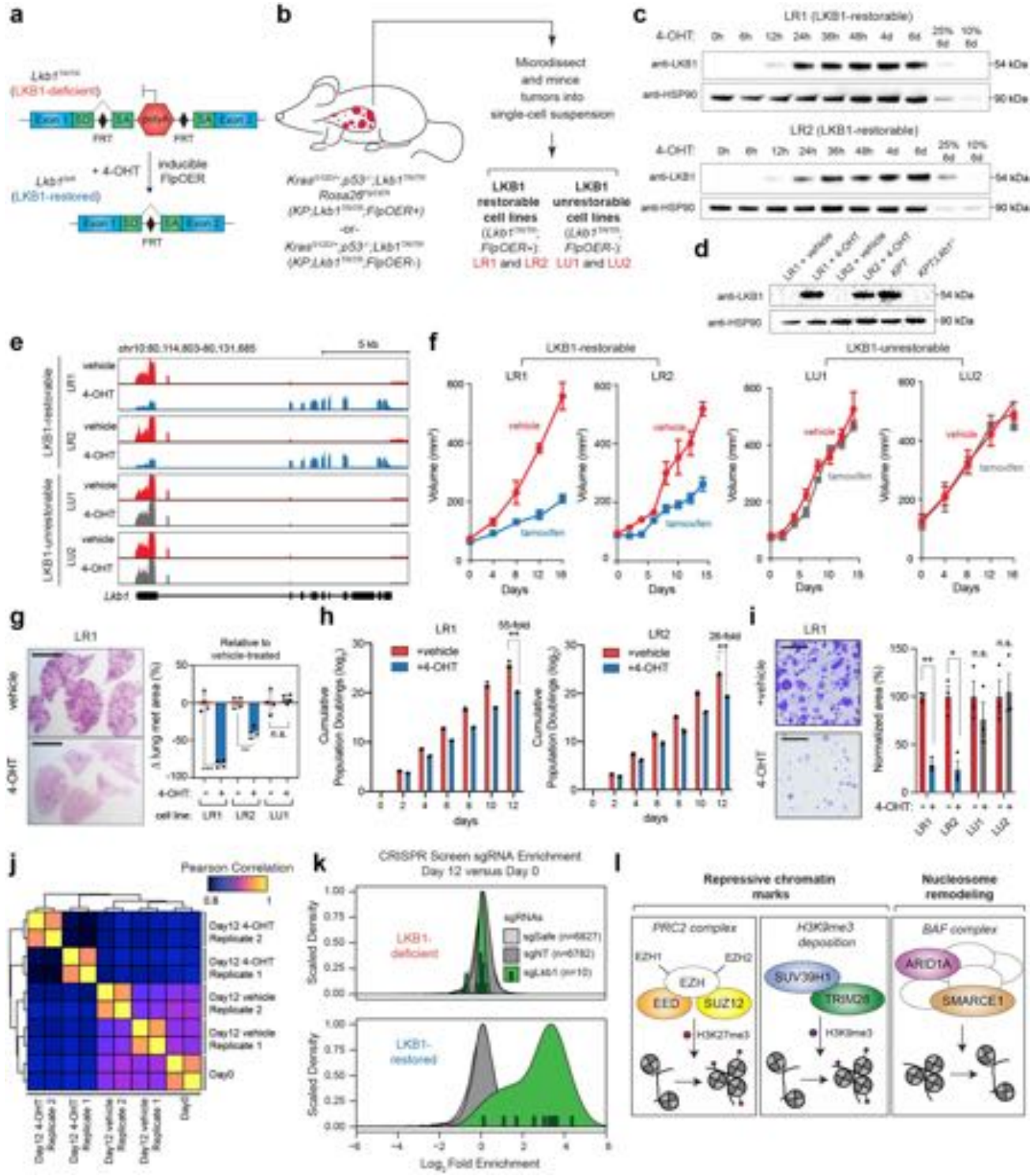
Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41556-021-00728-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-021-00728-4.

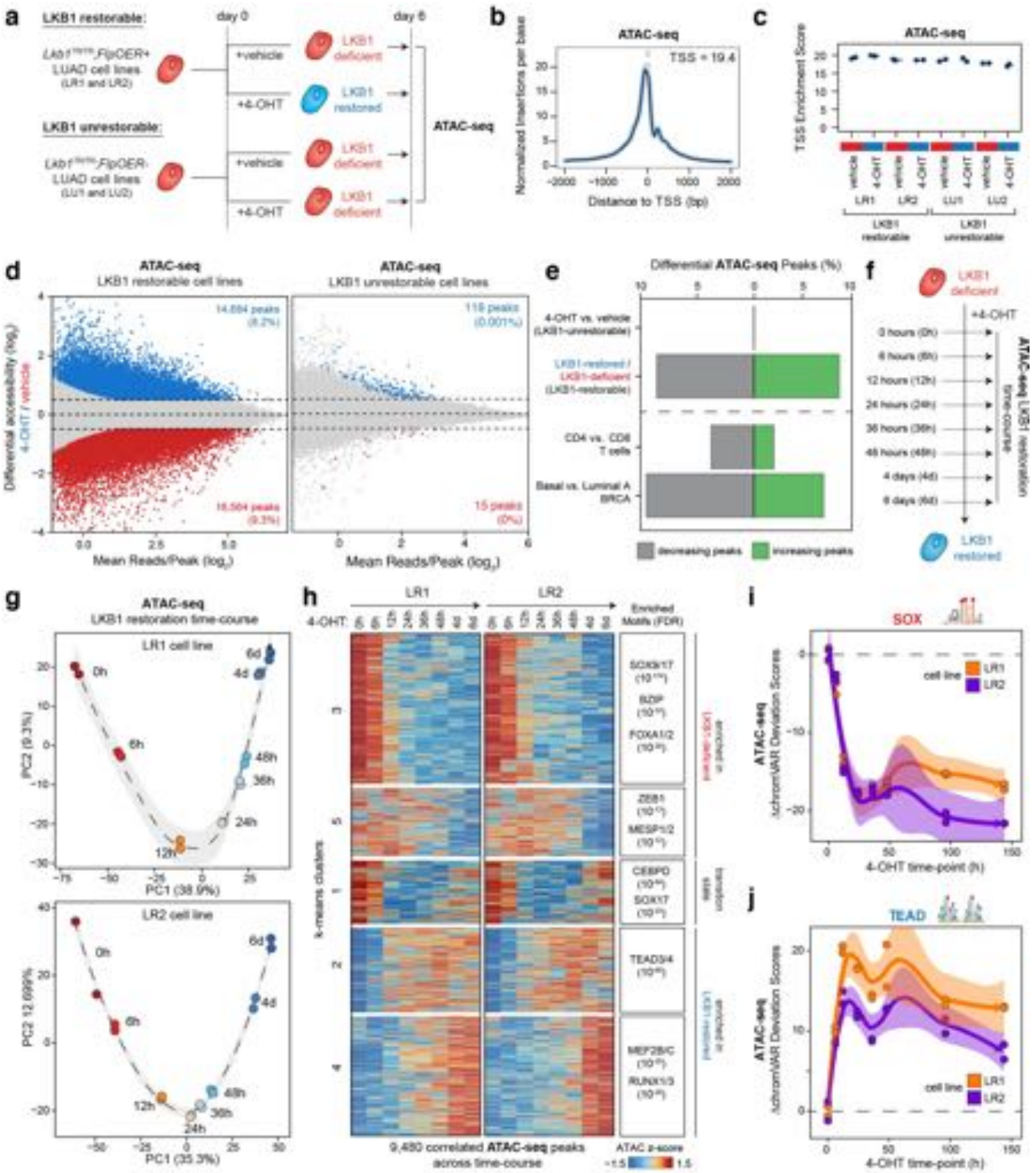
Correspondence and requests for materials should be addressed to S.E.P., W.J.G. or M.M.W.

Peer review Information *Nature Cell Biology* thanks Kwon-Sik Park, Tomi Makela and Skirmantas Kriaucionis for their contribution to the peer review of this work.



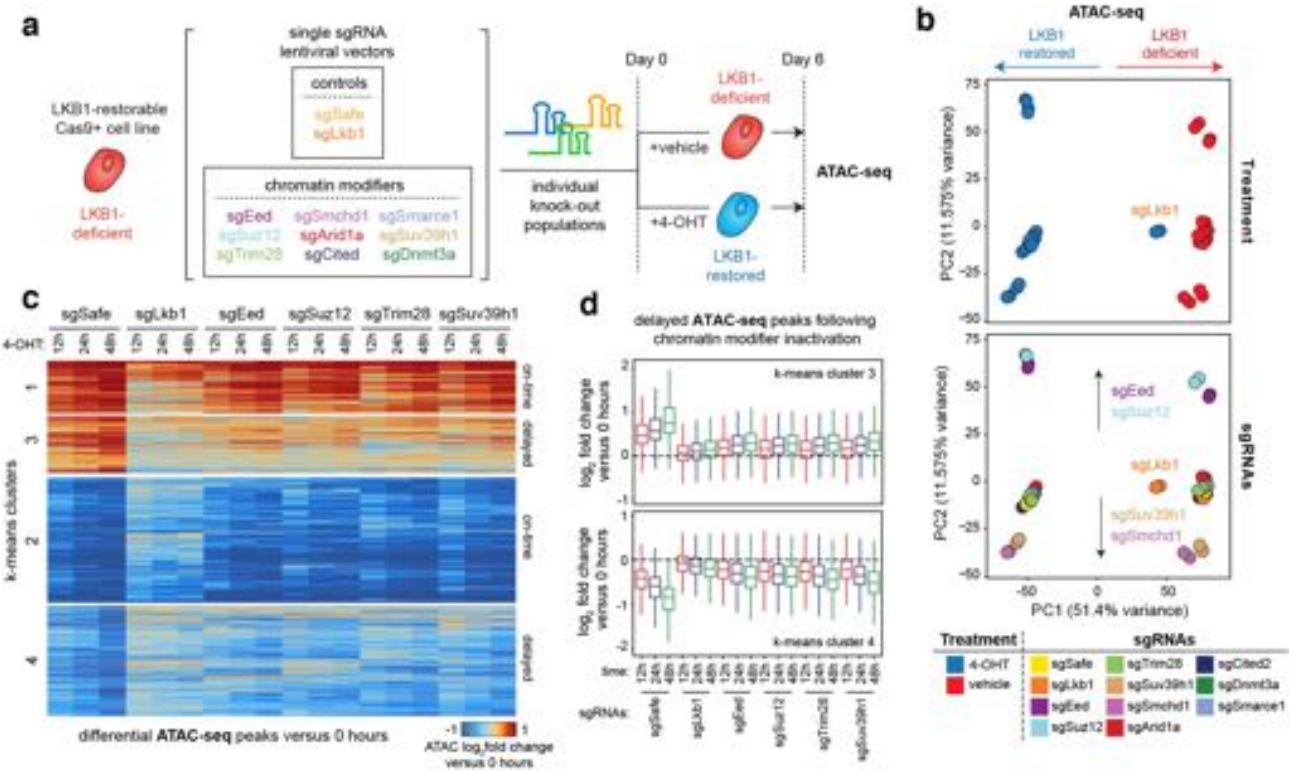
Extended Data Fig. 1 | See next page for caption.

Extended Data Fig. 1 | Validation and quality control of inducible LKB1 restoration model and genome-scale CRISPR/Cas9 screen. a. Schematic of restorable *Lkb1*$^{TR/TR}$ 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*$^{-/-}$ 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 +/− 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 +/− SEM is shown. $^{**}p = 0.001$, $^{***}p = 0.0001$, n.s. = not significant with a two-sided t-test. Scale bars represent 5 mm. **h**. Cumulative population doublings recorded over 12 days of 4-OHT treatment. Each cell line and condition was cultured and analyzed in triplicate. Mean +/− 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 +/− SEM is shown. $^{*}p < 0.01$, $^{**}p < 0.001$, n.s. = not significant with a two-sided t-test. Scale bars represent 10 mm. $p = 0.0001$ 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_2 fold enrichment of negative control sgRNAs and sgRNAs targeting *Lkb1* at day 12 versus day 0.

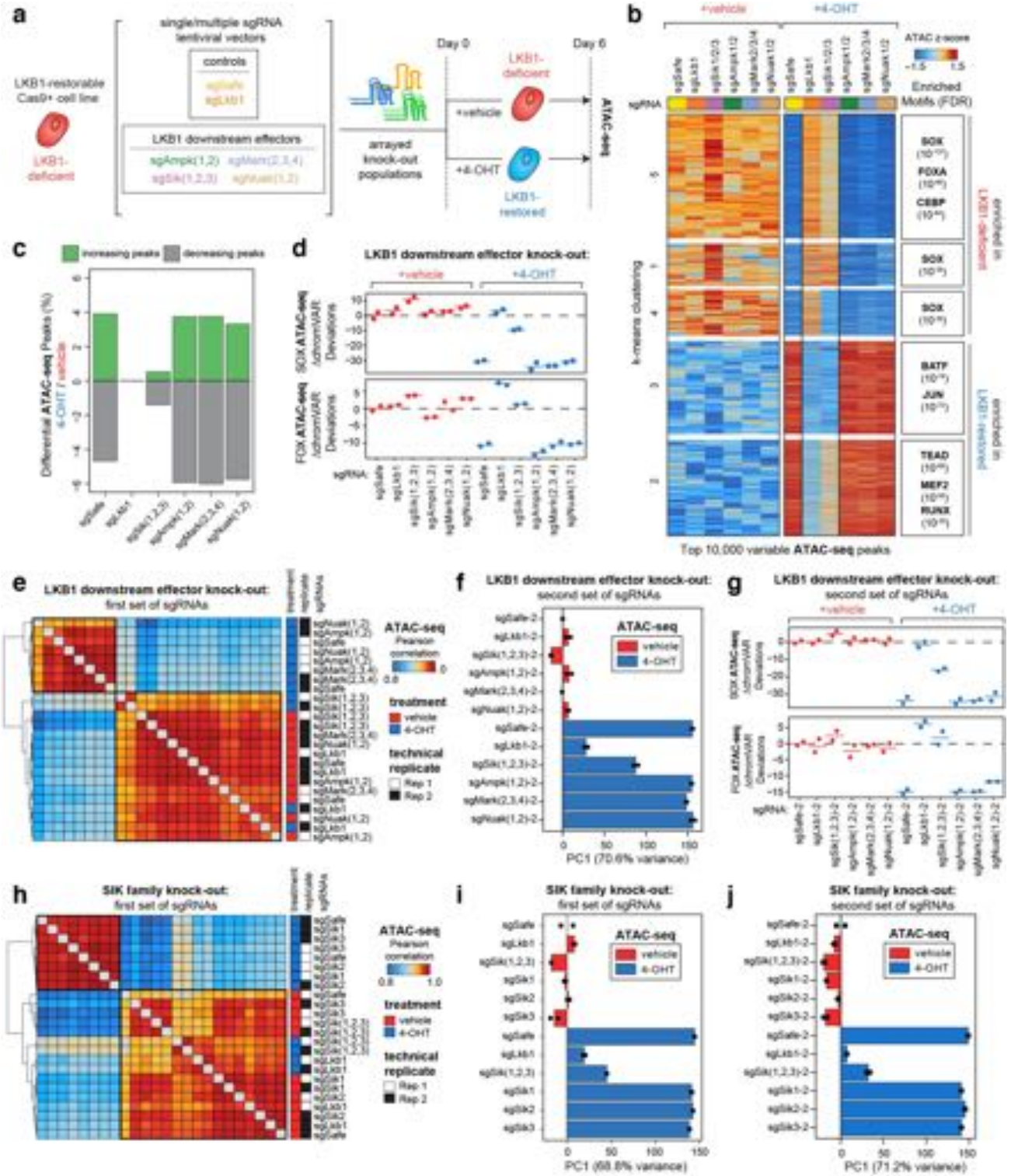


Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 2 | LKB1 restoration drives widespread changes in chromatin accessibility in lung adenocarcinoma cells. a. 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. **b**. 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. **c**. TSS enrichment scores for 16 ATAC-seq libraries with technical replicates. **d**. 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 (ΔchromVAR 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.

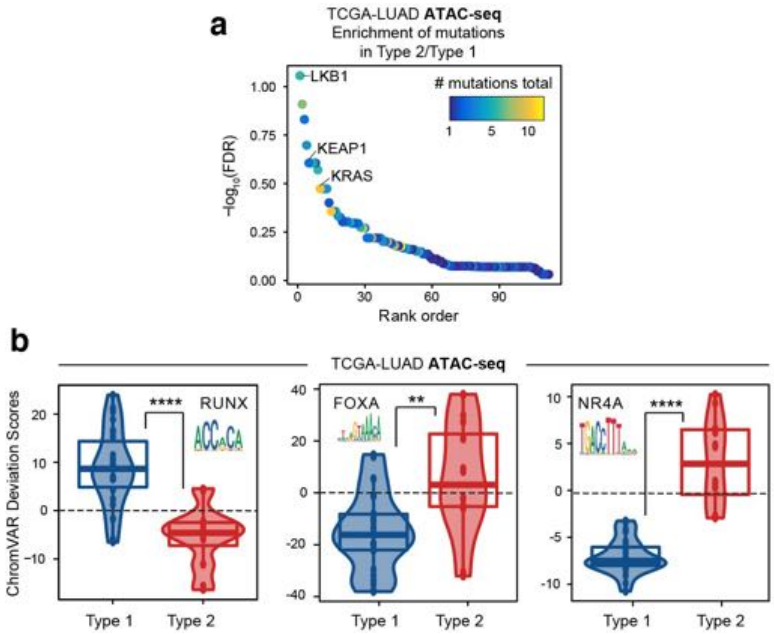


Extended Data Fig. 3 | Inactivating chromatin modifiers only delays LKB1-induced chromatin changes. a. Schematic of generating single knock-out populations of chromatin modifiers identified in the CRISPR screen, treating with 4-OHT or vehicle for six days, and processing for ATAC-seq. **b**. Principle 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 log2 fold change of each genotype and time-point versus the same genotype's initial time-point (day 0). **d**. Log_2 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.

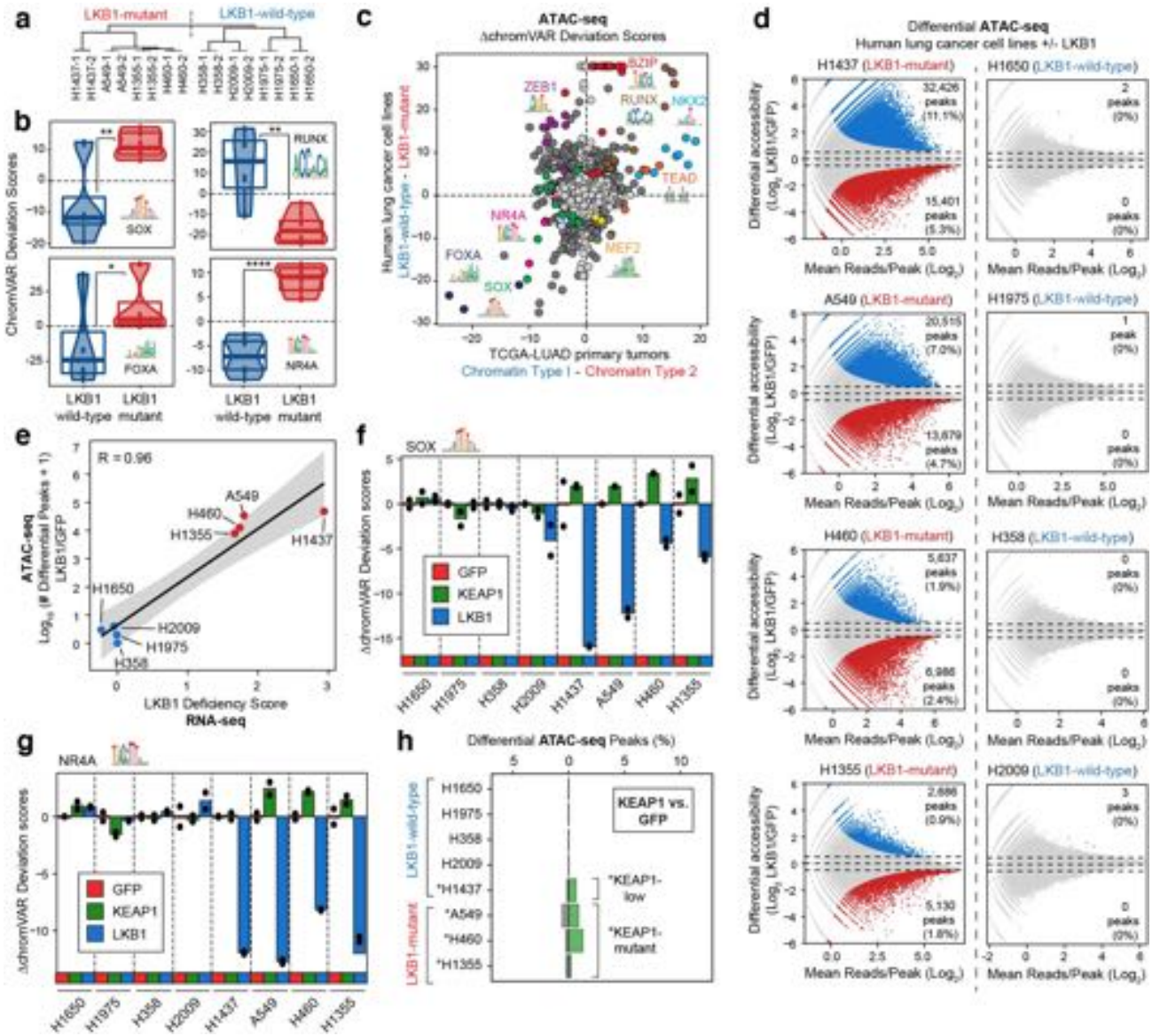


Extended Data Fig. 4 | See next page for caption.

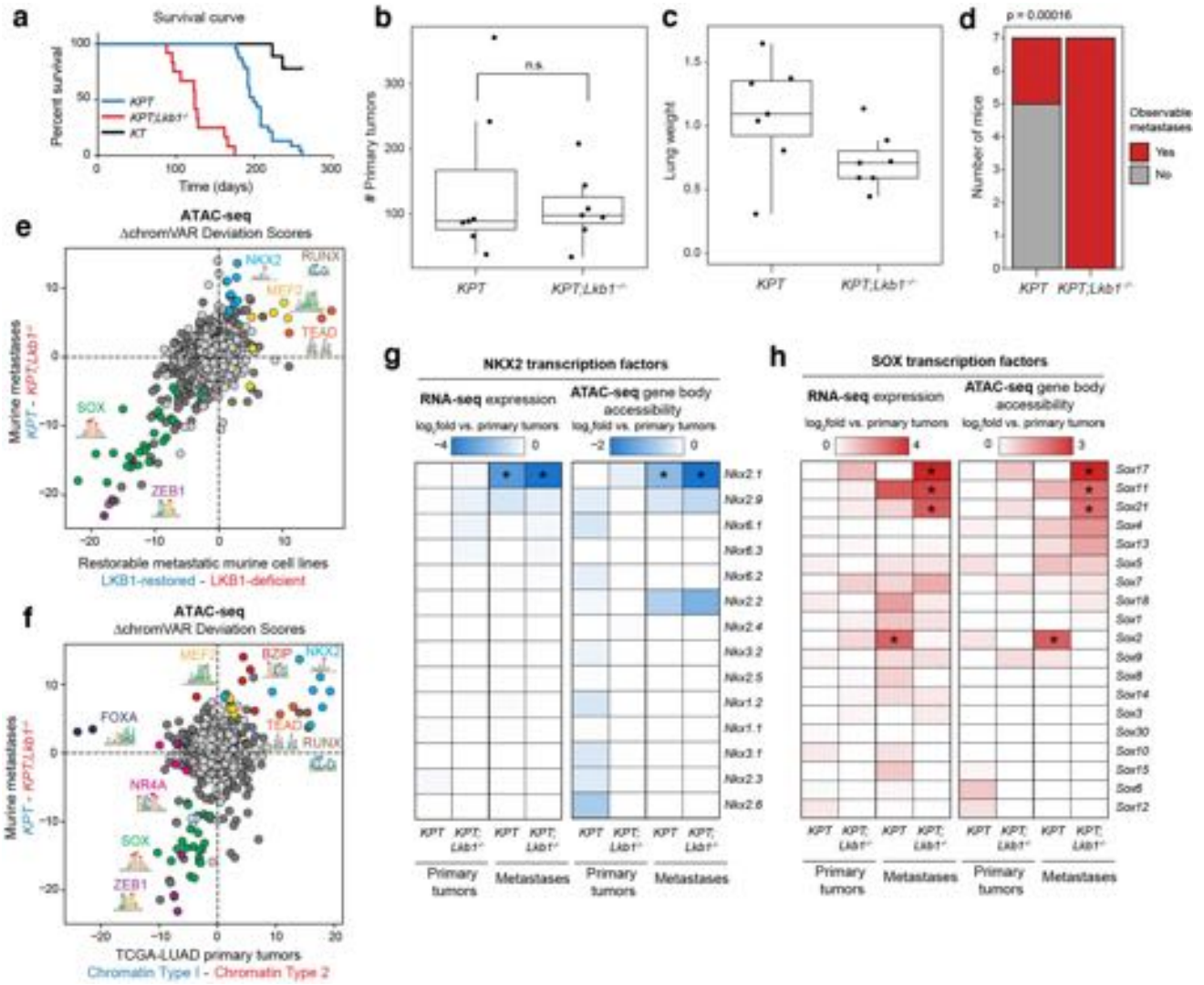
Extended Data Fig. 4 | SIK family members mediate LKB1-induced chromatin changes. a. Schematic of generating single and multiple sgRNA knock-out cell lines and processing for ATAC-seq. LR1;Cas9 cells were treated with 4-OHT or vehicle for six days. **b**. Left: Heatmap of peak accessibility between each 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 (ΔchromVAR 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 (ΔchromVAR 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.



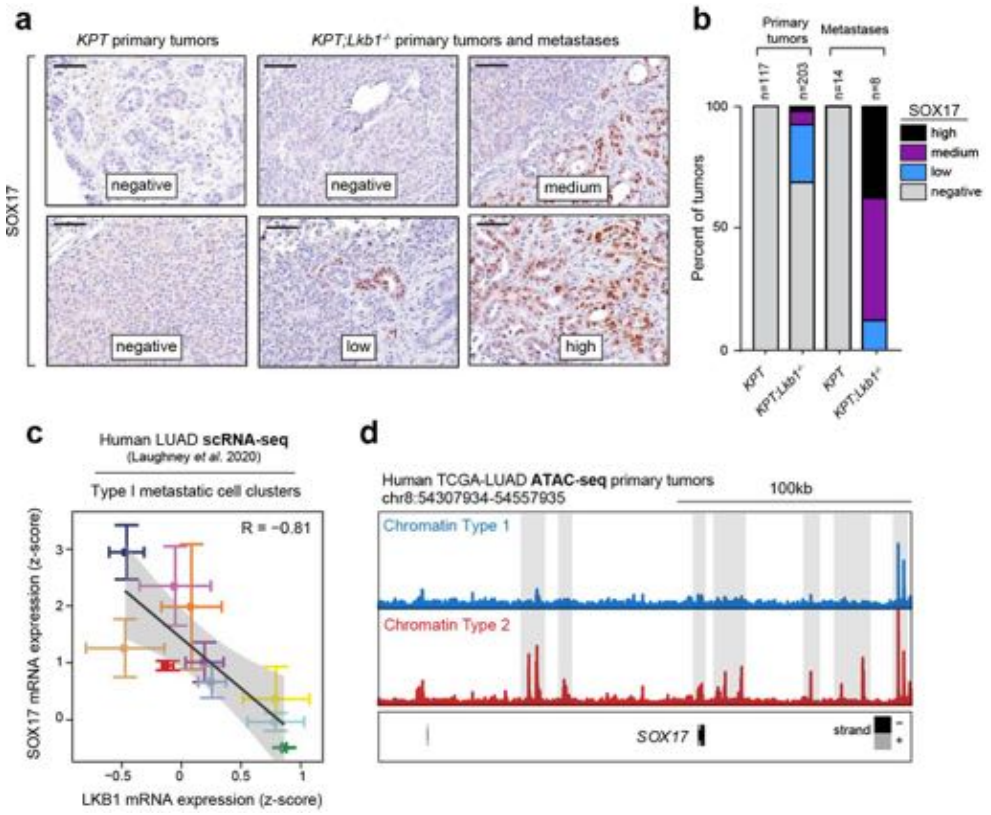
Extended Data Fig. 5 | Loss of LKB1 partitions human lung adenocarcinoma primary tumors into two chromatin accessibility sub-types. a. Enrichment of mutations in Chromatin Type 2 tumors compared to Chromatin Type 1 tumors. Genes are ranked according to -$\log_{10}$(FDR), with Rank 1 (LKB1) being the 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. $p = 1\times10^{-7}$ for RUNX, $p = 0.002$ for FOXA, and $p = 1\times10^{-7}$. $N = 13$ 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.



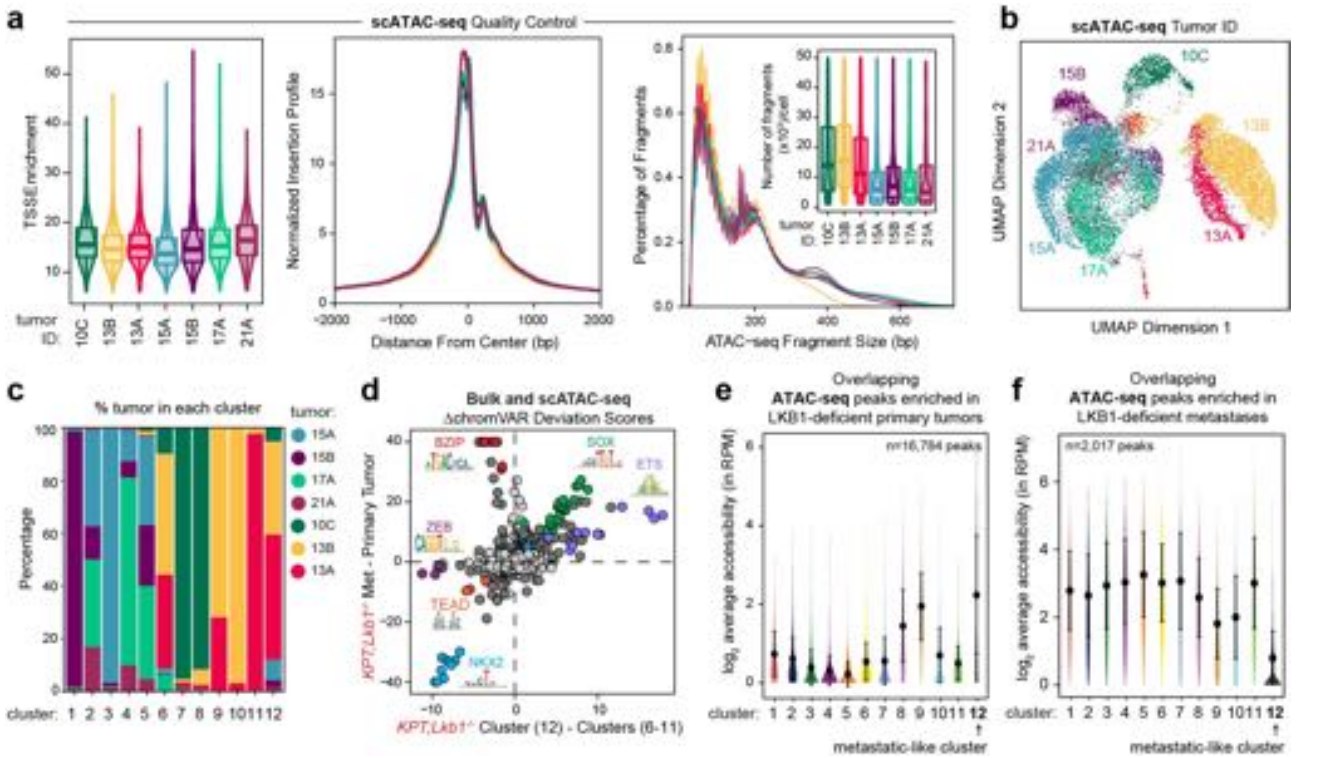
Extended Data Fig. 6 | Loss of LKB1 drives a unique chromatin accessibility state in human lung adenocarcinoma cell lines. a. Hierarchical clustering of human lung cancer cell lines using the Euclidian distance within the first three principle components from Fig. 2d. **b**. ChromVAR deviation scores for the indicated transcription factor motifs in eight human lung cancer cell lines at baseline. *$p < 0.1$, **$p < 0.005$, ****$p < 10\text{-}6$ 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 (Δ 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 ($|\log_2$ fold change$| > 0.5$, FDR < 0.05). **e**. LKB1-deficiency score by RNA-seq (using 16-gene signature from Kaufmann et al., 2017) compared to $\log_{10}$(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) an 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 ($|\log_2$ fold change$| > 0.5$, FDR < 0.05) in cells transduced to express KEAP1 compared to GFP.



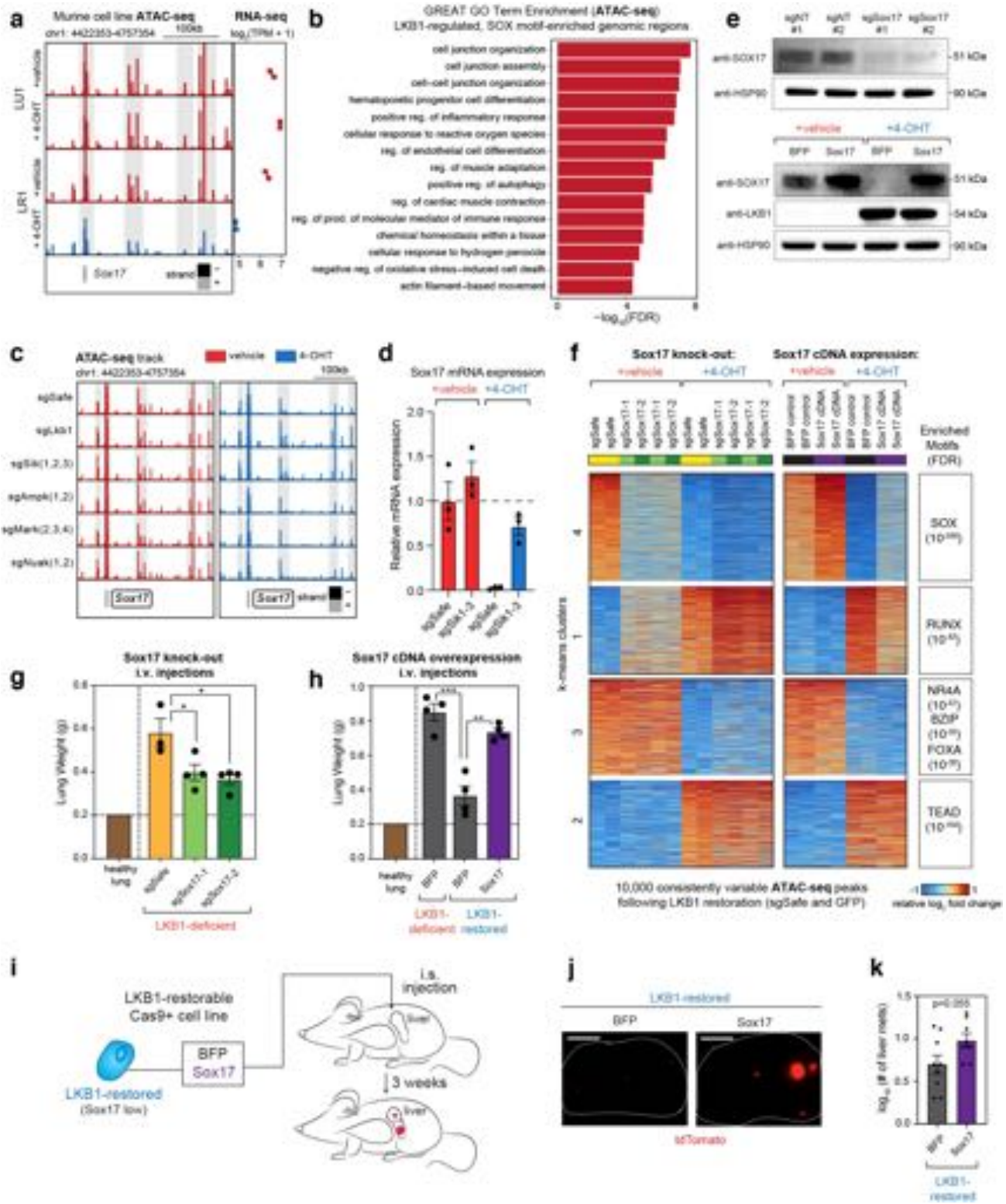
Extended Data Fig. 7 | Genotype-specific activation of SOX17 in LKB1-deficient metastatic cells. a. Percent survival of *KPT* and *KPT;Lkb1*$^{-/-}$ mice compared to *KT* mice. **b** and **c**. Number of primary tumors observed in *KPT* and *KPT;Lkb1*$^{-/-}$ mice (b). Lung weights of *KPT* and *KPT;Lkb1*$^{-/-}$ 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*$^{-/-}$ (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 (ΔchromVAR deviation scores) between murine LKB1-proficient (*KPT*) and LKB1-deficient (*KPT;Lkb1*$^{-/-}$) 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*$^{-/-}$ 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*$^{-/-}$ primary tumors, 14 *KPT* metastases, and 8 *KPT;Lkb1*$^{-/-}$ metastases, as quantified in (b). **b**. Quantitation of SOX17 protein expression in LKB1-proficient *KPT* and LKB1-deficient *KPT;Lkb1*$^{-/-}$ 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 +/− 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 10 C, 3556 cells for 13B, 1467 cells for 13 A, 3373 cells for 15 A, 1310 cells for 15B, 2858 cells for 17 A, and 851 cells for 21 A. 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 (ΔchromVAR 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 +/− 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 | See next page for caption.

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 differential ATAC-seq peaks in grey (log_2 fold change < −0.5, FDR < 0.05). *Sox17* also has significantly decreased mRNA expression (log_2 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_2fold 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 5 mm. **k**. Log_{10} (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.

nature research

Corresponding author(s): Sarah E. Pierce, William J. Greenleaf, Monte M. Winslow

Last updated by author(s): June 27th, 2021

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	cutadapt 1.15 - bulkATAC fastq read trimming Bowtie2 2.3.0 - bulkATAC Alignment samtools 1.6 - bulkATAC Alignment and filtering Cell Ranger 1.2.0 – scATAC Barcode Identification, Alignment, Filter, Deduplication
Data analysis	casTLE Version 1 - Screen Analysis Software. mageck 0.5.9 - Screen Analysis Software. macs2 2.1.1.20160309 – Peak Calling ATAC-Seq. R version 3.6.1 – R environment for all custom code edgeR - 3.26.8 - Software for bulk ATAC and RNA differential analysis. ArchR - 0.9.6 - Software for analysis of scATAC-seq data. rhdf5 - 2.30.1 - Software for HDF5 formatted analysis. Irlba 2.3.3 – Running PCA/SVD on large matrices. Rcpp 1.0.4 – Used for writing helpful C++ code to speed up operations. Rtsne 0.15 – Used for t-SNE embeddings. matrixStats 0.56.0 – Used for mathematical operations on large matrices. chromVAR_1.8.0 – Calculating TF deviation scores which can be associated with TF activity. Summarized Experiment 1.16.1 – R Data Class Environment used throughout analyses. Motifmatchr 1.8.0 – Matching TF Motifs within peak regions Seurat_3.1.2 – SNN Graph Clustering Implementation GenomicFeatures 1.32.2 – Genomic Ranges Operations used for overlap analyses GenomicRanges 1.38.0 - Genomic Ranges Operations used for overlap analyses Matrix 1.2-14 – Sparse Matrix math implementations. BSgenome 1.54.0 – Toolkit used for getting Genomic DNA sequences for motif matching and footprinting.

Rsamtools 2.2.3 – For manipulating BAM files within R
ggplot2_3.3.2 - R package for plotting.
GREAT version 4 - Gene Ontologies for ATAC-seq Peaks.
Picard Tools 2.20.3 - Aligning deduplicate ATAC-seq data.
Bowtie2 2.3.2 - Aligning ATAC-seq data.
Kallisto v0.46.1 - Aligning RNA-seq data.
TxDb.Mmusculus.UCSC.Mm10.knownGene 3.10 - Package with Gene locations for Mouse.
TxDb.Hsapiens.UCSC.hg38.knownGene 3.13 - Package with Gene locations for Human.
ComplexHeatmap 2.8.0 - Package for plotting heatmaps.
rtracklayer 1.51.0 - R interface to genome annotation files.
cellranger ATAC 1.2.0 - 10x Genomics scATAC software for aligning and counting.

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)

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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.ccbr.utoronto.ca/index.php.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Based on the current standard of the field, for CRISPR screens, ATAC-seq, and RNA-seq experiments, experiments were performed with two technical replicates each. For all other experiments, sample size is indicated in the figure legend for each experiment. While sample size was not predetermined with any statistical methods, the sample size was determined based on previous experience for each experiment to detect specific effects.
Data exclusions	No data were excluded from the manuscript.
Replication	All results presented in manuscript were reliably reproduced. The sample size and number of replicates for each experiment is included in the figure legends. Detailed information, particularly for the mouse experiments, is also provided in the Methods section.
Randomization	For all mouse experiments, mice were randomly allocated to each experimental group.
Blinding	No blinding was intentionally used but all samples had unique IDs that were not associated with any specific condition, such that the person collecting data was not aware of what the outcome might be.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

anti-LKB1 (Cell signaling, 13031S, D60C5F10 Rabbit mAb, dilution 1:1000 for Western blot), anti-HSP90 (BD Biosciences, 610418, Clone 68 Mouse mAb, dilution 1:2000 for Western blot), anti-SOX17 (Abcam, ab224637, Clone EPR20684 Rabbit mAb, dilution 1:500 for Western blot and 1:1000 for IHC), anti-Rabbit IgG HRP-linked antibody (Cell signaling 7074, Goat, dilution 1:2000 for Western blot), anti-Mouse IgG HRP-linked antibody (Santa Cruz, sc-2005, Goat anti-mouse polyclonal, dilution 1:2000 for Western blot).

Validation

The anti-LKB1 antibody we used is a high concentration version of another anti-LKB1 antibody (Cell Signaling, 3047S) that is more typically used for Western blotting. This antibody has been validated to work on murine L929 cells and we additionally showed that the appropriate band is not seen in LKB1-deficient cells but is seen in LKB1-restored cells (Extended Data Fig. 1c).

The anti-HSP90 antibody we used has been validated in many other publications on murine cells, including (Miyamoto et al 2002).

The anti-SOX17 antibody we used has been validated as part of the Protein Atlas and also has validation in murine tissue on the manufacturer's website. The band also appropriately goes away after knocking out SOX17 genetically with either of two sgRNAs (please see Extended Data Fig. 10e).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

All human cell lines were originally purchased from ATCC (NCIH1437, A549, NCIH460, NCIH1355, NCIH1650, NCIH1975, NCIH358, NCIH2009, 293T). Mouse cell lines (LR1, LR2, LU1, and LU2) were derived in our lab from primary tumors and metastases, as noted in the Methods section.

Authentication

None of the cell lines used were directly authenticated after purchase.

Mycoplasma contamination

All cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used in this study.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Mus musculus, 129:C57BL/6 mixed background and NSG, males and females, enrolled in the experiments at 2-4 months old and euthanized before 8 months (see Supplementary Table 5 for more information). Mice were kept at temperatures between 65-75 degrees Fahrenheit with 40-60% humidity under 12/12 light/dark cycles (light turns on at 7:00am and turns off at 7:00pm).

Wild animals

This study does not involve wild animals.

Field-collected samples

This study does not involve samples collected from the field.

Ethics oversight

Use of laboratory mice in this manuscript is approved by Administrative Panel on Laboratory Animal Care (APLAC) at Stanford University under Protocol 26696.

Note that full information on the approval of the study protocol must also be provided in the manuscript.