Metastatic site influences driver gene function in pancreatic cancer

Kaloyan M. Tsanov[1], Francisco M. Barriga[1,2], Yu-Jui Ho[1], Direna Alonso-Curbelo[1,3], Geulah Livshits[1], Richard P. Koche[4], Timour Baslan[1,5], Janelle Simon[1], Sha Tian[1], Alexandra N. Wuest[1], Wei Luan[1], John E. Wilkinson[6], Ignas Masilionis[7], Nevenka Dimitrova[1], Christine A. Iacobuzio-Donahue[8,9], Ronan Chaligné[7], Dana Pe'er[7,10], Joan Massagué[1], Scott W. Lowe[1,10,#]

[1]Cancer Biology & Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[2]Vall d'Hebron Institute of Oncology (VHIO), Barcelona, Spain

[3]Institute for Research in Biomedicine (IRB), The Barcelona Institute of Science and Technology (BIST), Barcelona, Spain

[4]Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[5]Department of Biomedical Sciences, School of Veterinary Medicine, The University of Pennsylvania, Philadelphia, PA, USA

[6]Department of Pathology, University of Michigan School of Medicine, Ann Arbor, MI, USA

[7]Computational & Systems Biology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[8]David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[9]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, USA

[10]Howard Hughes Medical Institute, Chevy Chase, MD, USA

[#]Corresponding Author: Scott W. Lowe, Memorial Sloan Kettering Cancer Center, 417 East 68th Street, New York, NY 10065, USA. Phone: 646-888-3342. E-mail: lowes@mskcc.org.

ABSTRACT

Driver gene mutations can increase the metastatic potential of the primary tumor[1-3], but their role in sustaining tumor growth at metastatic sites is poorly understood. A paradigm of such mutations is inactivation of *SMAD4* – a transcriptional effector of TGFβ signaling – which is a hallmark of multiple gastrointestinal malignancies[4,5]. *SMAD4* inactivation mediates TGFβ's remarkable anti- to pro-tumorigenic switch during cancer progression and can thus influence both tumor initiation and metastasis[6-14]. To determine whether metastatic tumors remain dependent on *SMAD4* inactivation, we developed a mouse model of pancreatic ductal adenocarcinoma (PDAC) that enables *Smad4* depletion in the pre-malignant pancreas and subsequent *Smad4* reactivation in established metastases. As expected, *Smad4* inactivation facilitated the formation of primary tumors that eventually colonized the liver and lungs. By contrast, *Smad4* reactivation in metastatic disease had strikingly opposite effects depending on the tumor's organ of residence: suppression of liver metastases and promotion of lung metastases. Integrative multiomic analysis revealed organ-specific differences in the tumor cells' epigenomic state, whereby the liver and lungs harbored chromatin programs respectively dominated by the KLF and RUNX developmental transcription factors, with *Klf4* depletion being sufficient to reverse *Smad4*'s tumor-suppressive activity in liver metastases. Our results show how epigenetic states favored by the organ of residence can influence the function of driver genes in metastatic tumors. This organ-specific gene–chromatin interplay invites consideration of anatomical site in the interpretation of tumor genetics, with implications for the therapeutic targeting of metastatic disease.

MAIN

Metastatic disease – the growth of cancers beyond the primary tumor – accounts for 90% of cancer-related deaths[15,16]. Metastasis involves the acquisition of multiple traits that enable cells to leave the primary tumor, survive in the circulation, and ultimately reach and colonize other organs[16,17]. Despite the distinct capabilities that must be acquired for a tumor cell to successfully metastasize, genome-sequencing efforts have identified few driver gene mutations that are specific to metastatic tumors[5,18,19]. This has suggested that pro-metastatic traits arise from epigenetic programs that facilitate cell state changes such as the epithelial-to-mesenchymal transitions (EMTs)[16,17,20-22]. Although driver gene mutations can endow primary tumors with increased metastatic capacity[1-3], whether or how tumor evolution or the metastatic site itself influences their contribution to tumor maintenance is unknown. Such knowledge would have important implications for precision oncology and may guide the development of much needed metastasis-targeting therapies.

Among driver mutations, inactivation of the *SMAD4* tumor suppressor gene – a core mediator of TGFβ signaling – is a hallmark of several gastrointestinal malignancies that is found at highest frequency in PDAC[4,5,10]. During cancer progression, *SMAD4* inactivation shifts TGFβ's activity from tumor-suppressive to tumor-promoting by impairing its ability to trigger cell cycle arrest and EMT-coupled apoptosis[9,23]. Accordingly, *SMAD4*-mutant tumors have higher rates of metastasis in PDAC patients, an effect recapitulated in animal models[7,8,11,14]. However, it is unknown whether *SMAD4* inactivation maintains disease at metastatic sites, which is key to understand given that most PDAC patients are diagnosed after tumor cells have spread to distant organs[24]. In this study, we took advantage of a new murine model that enables inducible and reversible *Smad4* inactivation at different PDAC stages to interrogate the ongoing need for *Smad4* inactivation in metastatic disease. Our results reveal a diametrically opposed role for *Smad4* inactivation in sustaining liver and lung metastases and establish a critical interplay between driver mutations and organ-specific chromatin states that contributes to the heterogeneity of cancers driven by identical genetic lesions.

RESULTS

Smad4-restorable genetically engineered mouse model of PDAC

To enable reversible *Smad4* inactivation in PDAC, we generated a genetically engineered mouse model (GEMM) that harbors pancreas-specific, single-copy, doxycycline (Dox)-inducible short hairpin RNA (shRNA) against *Smad4* (or against *Renilla* luciferase as a control) on the background of oncogenic *Kras*^{G12D} (hereafter KC-shSmad4 and KC-shRen, respectively; see Methods) (**Fig. 1a**). This genetic strategy allows for tumor development in the setting of *Smad4* depletion (by Dox administration) and subsequent restoration of *Smad4* expression at physiological levels from its endogenous locus (by Dox withdrawal). The alleles also contain two fluorescent reporters that facilitate identification and isolation of tumor cells: a constitutive mKate2 and an shRNA-linked GFP (**Fig. 1a**).

Consistent with results from conventional knockout GEMMs[6,12-14], *Smad4* depletion promoted tumor initiation, shortened survival, and led to the development of tumors that metastasize to the liver and, less frequently, to the lungs (**Fig. 1b, c; ED Fig. 1a**). At late stage, primary and metastatic tumors expressed the fluorescent reporters (**ED Fig. 1b**) and maintained potent depletion of SMAD4 protein (**ED Fig. 1c**). Notably, tumor formation appeared to require further inactivation of the *Cdkn2a* tumor suppressor gene, as sparse whole-genome sequencing of tumor-derived cell lines revealed spontaneous homozygous deletion of the *Cdkn2a/b* locus in 9/10 cases[25] (**Fig. 1d**). This lesion (along with *Kras* gain) was the most prominent event in an otherwise largely stable genome (**ED Fig. 1d**), in agreement with the previously reported genome evolution of PDAC driven by alterations in the TGFβ pathway[26]. The *Cdkn2a/b* deletions were highly concordant between primary and metastatic tumors (**ED Fig. 1e**), and they mirrored the genetic association of *SMAD4* alterations with homozygous *CDKN2A/B* deletions in the MSK IMPACT cohort of human PDAC (**ED Fig. 1f**). Thus, the KC-shSmad4 GEMM recapitulates cardinal features of the human disease and further enables reversible *Smad4* inactivation.

***Smad4* restoration has organ-specific effects on tumor growth**

To study *Smad4* reactivation in metastatic tumors, we turned to transplantation assays using cell lines derived from primary GEMM tumors that were capable of metastasis, as this approach afforded longer experimental time before tumor burden necessitated mouse euthanasia. The tumor-derived cell lines maintained robust *Smad4* restorability (**Fig. 1e**), mounted an expected *Smad4*-dependent cytostatic response to TGFβ *in vitro* (**ED Fig. 2a, b**), and produced metastases with remarkably similar histopathology to those emerging in the GEMMs and in PDAC patients[27] (**ED Fig. 2c**). *Smad4*-dependent apoptotic responses induced by TGFβ in PDAC epithelial progenitors[9,28] were not captured in these cell lines at the analyzed timepoints.

KC-shSmad4 cells were stably transduced with firefly luciferase to facilitate tumor monitoring and delivered via orthotopic, intrasplenic, or tail vein injections into nude mice to respectively generate primary tumors, liver metastases, or lung metastases (**Fig. 2a**). After 4-6 weeks, which allowed for tumor formation under *Smad4*-depleted conditions (+Dox = Smad4 OFF), *Smad4* expression was restored by Dox withdrawal (–Dox = Smad4 ON) in a randomly selected half of each cohort, and tumor burden was assessed 30 days later (**Fig. 2a**). Strikingly, the response to *Smad4* restoration was different in each of the three organs: tumor burden was unchanged in the pancreas, decreased in the liver, and increased in the lungs (**Fig. 2b-d; ED Fig. 2d**). Of note, similar results were obtained in spontaneous metastases arising from the orthotopic KC-shSmad4 transplants but not in KC-shRen cells, thus ruling out artifactual effects due to the employed metastasis assays or Dox administration (**ED Fig. 2e-g**).

To determine the relevance of our findings to human PDAC, we queried data on metastatic recurrence in PDAC patients after primary tumor removal, where *SMAD4* status was evaluated by immunohistochemistry (IHC) staining of resected primary or metastatic tumors[29,30]. Corroborating a potentially tumor-suppressive vs. promoting function of *SMAD4* in the liver vs. lungs, 69% of cases with recurrent liver metastases lacked SMAD4 expression, in contrast to 50% of concurrent and 33% of isolated lung metastases (**ED Fig. 2h**). Thus, *Smad4* inactivation confers a selective advantage to liver metastases and a disadvantage to lung metastases, consistent with SMAD4 expression patterns in PDAC patients with metastatic recurrence.

Smad4 **induces different transcriptional programs in liver vs. lung metastases**

SMAD4 acts as a transcription factor (TF) by forming a complex with the SMAD2/3 TFs to activate gene expression programs downstream of TGFβ receptors[31]. Hence, we performed RNA sequencing (RNA-seq) to explore the transcriptional basis of the observed organ-specific phenotypes. Leveraging the mKate2 reporter in our model, we used fluorescent-activated cell sorting (FACS) to isolate tumor (mKate2$^+$) cells 7 and 14 days after Dox withdrawal, which allowed for assessment of *Smad4* output kinetics. Consistent with its organ-specific effects on tumor growth, *Smad4* restoration led to upregulation of partially overlapping but mostly distinct genes, as compared to tumor cells kept on Dox: 89% (1580/1773) and 74% (613/825) organ-specific genes at days 7 and 14, respectively (**ED Fig. 3a**). This phenomenon was particularly pronounced in the liver at the 7-day timepoint and was still observed across all three organs at day 14 (**Fig. 3a**). Importantly, these results were not confounded by differential *Smad4* expression or baseline TGFβ signaling, since the extent of *Smad4* depletion/restoration and the levels of phosphorylated SMAD2 (a SMAD4-independent readout of TGFβ signaling[31]) were similar between the three organs (**ED Fig. 3b, c**).

To better understand the organ-specific consequences of *Smad4* restoration, we next performed functional annotation of *Smad4*-activated genes in the liver and lungs, as these organs exhibited opposite tumor growth phenotypes. Gene ontology analysis revealed that tumors in both organs upregulated extracellular matrix (e.g. glycosaminoglycan, proteoglycan, and focal adhesion) and EMT-related transcriptional programs, while only liver metastases showed enrichment for cell cycle and senescence-related gene signatures (**Fig. 3a; ED Fig. 3d**). Intersection of these gene lists with available data from SMAD2/3 ChIP-seq (chromatin immune-precipitation followed by sequencing) of murine PDAC cells[9,28] confirmed differential engagement of SMAD4-dependent binding targets, further implicating direct effects of *Smad4* reactivation (**ED Fig. 3e, f**).

Importantly, the differentially expressed targets included genes that distinguish between SMAD4's tumor-suppressive and tumor-promoting functions. They prominently featured upregulation of the tumor suppressor gene and cell cycle inhibitor *Cdkn1c* (also known as p57$^{\mathrm{KIP2}}$)[32,33] specifically in the liver, as well as induction of SMAD4's tumor-promoting fibrogenic program (including *Il11, Has2, Serpine1, Col6a1, Itga5, Ccbe1* and *Wisp1*)[28] in both organs, albeit with a delayed kinetics in the lungs (**Fig. 3b**). Of note, the SMAD4-dependent target *Id1*, known to reflect TGFβ's pro-tumor mode of action[34], showed elevated expression in lung metastases but downregulation in liver metastases (**Fig. 3b**).

These transcriptomic data suggesting differential cytostatic and fibrogenic outputs were validated by immunostaining for the proliferation marker Ki67 and the TGFβ-dependent myofibroblast marker α-smooth muscle actin (α-SMA)[28]. In support of a liver-specific cytostatic response and liver/lung-shared fibrogenic response, *Smad4* reactivation reduced the proportion of Ki67$^+$ tumor cells only in the liver, while both organs exhibited increases in α-SMA levels (**Fig. 3c, d**). Taken together, our analysis reveals discordant engagement of *Smad4*'s tumor-suppressive vs.

promoting effectors in liver vs. lung metastases, thus demonstrating that anatomical site can uncouple TGFβ's anti- and pro-tumorigenic programs.

Liver and lung metastases harbor distinct chromatin states

Given the organ-specific transcriptional responses, we hypothesized that liver and lung metastases may harbor distinct chromatin states that afford different accessibility to SMAD4's target genes. To test this, we performed ATAC-seq (assay for transposase-accessible chromatin using sequencing) on FACS-isolated tumor cells from the pancreas, liver, and lungs +/- *Smad4* restoration. Unsupervised clustering of differentially accessible peaks linked distinct chromatin states to tumors residing in different organs, while *Smad4* status itself had a limited impact (**Fig. 4a; ED Fig. 4a, b**), consistent with the fact that it is not a pioneer TF[31,35]. However, the differential chromatin accessibility affected SMAD4-dependent target genes, including its cytostatic and fibrogenic effectors highlighted earlier (**ED Fig. 4c**).

In light of these data, we then asked which other TFs are predicted to bind in the liver- vs. lung-specific ATAC-open regions and may thereby facilitate SMAD4's organ-specific activity. Motif analysis revealed mostly distinct TF families, the top unique predictions being KLF, HNF, and ELF in the liver, and RUNX, FOX, and ETS in the lungs (**Fig. 4b**). To determine which of these may cooperate with SMAD4 to impact organ-specific gene expression, we integrated our RNA- and ATAC-seq datasets to identify TF families: (i) whose motifs were enriched in the differentially accessible regions, and (ii) whose predicted targets were upregulated upon *Smad4* restoration. This analysis narrowed down the list of organ-specific candidates to the KLF family in the liver and the RUNX family in the lungs (**Fig. 4c**), both of which are TF families with pioneer factor capabilities[35-37]. Finally, to assess which TF families are of highest relevance to human PDAC, we used RNA-seq data from metastatic human PDAC[38] to impute differential TF activity in liver or lung metastases based on enrichment or depletion of a given TF's target genes relative to primary tumors (see Methods for details). Corroborating our mouse data, KLF targets were enriched in liver metastases, while RUNX targets were depleted in liver and enriched in lung metastases (**Fig. 4d**). Overall, these results nominate the KLF and RUNX TF families as organ-specific determinants of chromatin-directed transcriptional programs responsive to SMAD4.

The KLF and RUNX families contain multiple TFs that have been implicated in PDAC biology. Among them, KLF4 and KLF5 are both enforcers of pancreatic epithelial identify[9,39-41], whereby KLF5 silencing by TGFβ compromises PDAC cell survival[9]. Interestingly, KLF4 can exhibit pro- or anti-tumorigenic effects in PDAC depending on context, despite binding to similar DNA motifs as KLF5[9,41,42]. On the other hand, RUNX1 and RUNX3 have been implicated as drivers of invasion and metastasis and as potential genetic dependencies in PDAC[14,43,44]. To define which specific TF(s) are most likely to underlie the observed organ-specific phenotypes, we performed IHC staining for each of these factors in KC-shSmad4 metastases. KLF4 exhibited strong specificity for liver and RUNX1 for lung metastases; at the same time, KLF5, KLF6, and RUNX2 presented signal in both organs, albeit to a different extent, and RUNX3 did not show tumor-specific signal but rather stained stromal cells in our model (**Fig. 4e; ED Fig. 4d**). Thus, our refined analysis further nominates KLF4 and RUNX1 as particular TFs that are likely to mediate organ-specific chromatin opening in liver vs. lung metastases.

Liver and lung metastasis-like cell states emerge in primary tumors

A growing body of evidence suggests that pro-metastatic epigenetic programs can arise early during tumorigenesis[45,46]. Interestingly, IHC staining for KLF4 and RUNX1 in KC-shSmad4 primary tumors was heterogeneous, indicating broad expression of both TFs yet in apparently non-overlapping subsets of cells (**Fig. 4f**). These data suggested that the KLF and RUNX-associated chromatin states – and their opposite responsiveness to SMAD4 – may already be present in sub-populations within the primary tumor. To further address this, we integrated our

ATAC- and RNA-seq data from primary tumors +/- *Smad4* restoration (for 14 days) to infer KLF and RUNX TF activities. The latter were defined by combining metrics of chromatin accessibility at the respective TF motif with transcriptional changes in the TF's target genes (see Methods for details). This analysis revealed that *Smad4* restoration caused a decrease of inferred KLF activity and an increase of inferred RUNX activity in the pancreas (**ED Fig. 4e**). These data support the concept that the KLF and RUNX-associated cell states pre-exist in the primary tumor, whereby SMAD4 antagonizes KLF activity while cooperating with RUNX activity.

To further assess whether chromatin states that pre-exist in the primary tumor are linked to organ-specific metastasis, we performed coupled single-cell multiomics (ATAC/RNA-seq) on three independent KC-shSmad4 primary tumors and then mapped liver- and lung-specific open chromatin peaks identified in our bulk ATAC-seq on the single-cell space. This analysis identified distinct primary tumor cell sub-populations that harbored chromatin states resembling the organ-specific states observed in established liver and lung metastases (**Fig. 4g; ED Fig. 4f**). We also probed previously published single-cell ATAC-seq data from pre-malignant pancreatic tissue (harboring only *Kras*^{G12D})[45]. Remarkably, this analysis yielded similar results to the advanced KC-shSmad4 primary tumors, implying that the organ-specific chromatin states may arise even before cells acquire malignant potential (**ED Fig. 4g**).

Finally, we asked if the identified liver- and lung-specific metastatic states can also be found in human PDAC. To this end, we leveraged the multiomic nature of our mouse single-cell data to generate matching RNA-seq signatures of the cell populations that were enriched for liver- vs. lung-specific chromatin peaks in the ATAC-seq analysis. The resulting signatures were then used to query single-cell RNA-seq data from 16 human primary PDAC samples[47]. This analysis confirmed the existence of distinct cell sub-populations in the human primary tumors that resemble the organ-specific cell states in our mouse model (**ED Fig. 4h**). The resemblance was particularly strong for the liver-specific states, likely because liver metastasis is more common than lung metastasis in PDAC patients. Overall, our data suggest that the liver and lungs favor metastatic cells harboring different chromatin states which are similar to pre-existing states found in the primary tumor.

***Klf4* depletion is sufficient to reverse *Smad4* function in liver metastasis**

Finally, we functionally interrogated the *Smad4*-TF interplay, focusing on the setting of liver metastasis where *Smad4*-mediated tumor suppression could be successfully restored. To do so, we used stable shRNA-based knockdown to deplete *Klf4* or *Runx1* (or shRNAs to target *Renilla* luciferase as a neutral control) in KC-shSmad4 cells. While both sh*Klf4* and sh*Runx1* achieved potent reduction of the respective proteins (**Fig. 5a**), the corresponding cell lines had unaltered basal proliferation and response to TGFβ upon *Smad4* restoration *in vitro* in comparison to the sh*Ren* control (**ED Fig. 5a**). We then generated liver metastases by intrasplenic injection of mice on Dox (Smad4 OFF), as described earlier (**ED Fig. 5b**). Consistent with a general role for *Klf4* in supporting a liver-metastatic cell state, its depletion reduced baseline tumor burden; by contrast, *Runx1* depletion produced a smaller reduction in tumor burden that did not reach statistical significance (**ED Fig. 5c**).

The effect of *Klf4* depletion on the tumor-suppressive response to *Smad4* in liver metastasis was striking. Whereas the sh*Ren*- and sh*Runx1*-expressing metastases displayed expectedly similar tumor suppression upon *Smad4* reactivation, sh*Klf4*-expressing metastases now exhibited a tumor-promotion phenotype of a three-fold increase in metastasis burden (**Fig. 5b-d**). Thus, *Klf4* depletion is sufficient to reverse *Smad4* function in liver metastases, which demonstrates that KLF4 facilitates *Smad4*'s tumor-suppressive activity in this setting. These data provide functional validation of our model whereby KLF4 cooperates with *Smad4* inactivation to sustain liver metastases (**ED Fig. 5d**).

DISCUSSION

Our results demonstrate how driver gene mutations that are important for tumor initiation can show opposite requirements for maintenance at metastatic sites and place the paradigmatic duality of TGFβ signaling in an anatomical context. While it has been established that *SMAD4* mutation can switch TGFβ's activity from tumor-suppressive to tumor-promoting[23], here we show that this switch can also be mediated by the organ location of otherwise isogenic tumors. *SMAD4* loss itself thus does not universally favor tumor growth since its inactivation appears to be a liability rather than an advantage for lung metastases, at least in the clinically relevant setting of pre-seeded metastases modeled in our study. Hence, our data provide a possible reason for the unusually low rates of *SMAD4* inactivation in isolated lung metastases of PDAC patients[29,30,48].

Mechanistically, the involvement of KLF and RUNX factors is consistent with both the operational logic of TGFβ signaling – whose contextual effects are often defined by interplay with such developmental TFs[31] – and the reported functions of these TFs in PDAC and other settings. In particular, certain KLF factors promote the epithelial cell fate[39], which in turn is known to be essential for colonization of the liver[49,50], whereas RUNX proteins facilitate extracellular matrix remodeling that can create a pro-metastatic fibrogenic microenvironment in the lungs when unopposed by a tumor-suppressive program[46]. Future studies will determine the extent to which organ-specific effects apply to other common cancer drivers, many of which also enhance metastatic proclivity; such drivers include missense mutations of the *TP53* tumor suppressor gene[51], gains/amplifications of the mutant-*KRAS/PTHLH*[26,52] and *MYC*[53] loci, or deletions of the *CDKN2A*/type I interferon locus[25].

Our results also add to the growing appreciation that tissue context can influence the output of gene mutations in cancer, for example, as illustrated by the differential susceptibility of cells from particular tissues or tissue locations to certain oncogenic events[54-57]. Our findings extend this concept to metastasis by showing how organ site can have a profound impact on a single driver mutation in a tumor from the same tissue of origin. As an underlying mechanism, we show that such mutations synergize with or antagonize distinct chromatin states that emerge early during tumorigenesis and are favored by different metastatic sites. Additional work will define the contribution of immune and other factors in the organ microenvironment that likely influence this gene–chromatin interplay. Regardless, when extended to precision oncology, our results draw attention to potentially divergent responses to therapy based on the tumor's organ of residence. As such, they are in line with clinical observations of organ heterogeneity in therapy response[58-61] and invite consideration of organ-specific therapies for tumors driven by mutations that show such dependence on metastatic site.

METHODS

Animals and *in vivo* procedures

Animal care: All mouse experiments were approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee (IACUC). Mice were maintained under pathogen-free conditions, housed on a 12 h–12 h light–dark cycle under standard temperature and humidity of approximately 18–24°C and 40–60%, respectively. Food and water were provided *ad libitum*. GEMMs were generated in house. *Foxn1*^{nu} (athymic nude) mice used for transplants were purchased from Envigo or The Jackson Laboratory.

GEMMs: *Ptf1a*$^{Cre/+}$*;LSL-Kras*$^{G12D/+}$*;Rosa26*$^{LSL\text{-}rtTA3\text{-}IRES\text{-}mKate2/+}$*(RIK);Col1a1*$^{shRNA\text{-}Homing\text{-}Cassette/+}$*(CHC)* embryonic stem cells (ESCs)[62] were targeted with two independent GFP-linked Smad4 shRNAs (shSmad4.591: CAAAGATGAATTGGATTCTTT; shSmad4.1599: ACAGTTGGAATGTAAAGGT-GA) cloned into miR30-based targeting constructs, as previously described[62,63]. Targeted ESCs were selected and functionally tested for single integration of the GFP-linked shRNA element into the CHC locus, as previously described[62]. The KC-shRen ESC control clone used in this study has been described[62]. Before injection, ESCs were expanded briefly in KOSR+2i medium[64] and confirmed to be negative for mycoplasma. Mice were generated by 8-cell or blastocyst injection of KC-shSmad4 or KC-shRen ESCs, and short hairpin RNAs were induced by treatment of the resulting mice with doxycycline (625 mg/kg) in the drinking water starting at 5-6 weeks of age. The identity of the ESCs and ESC-derived mice were authenticated by genomic PCR using a common Col1a1 primer paired with an shRNA-specific primer, all yielding products of approximately 250 bp:

- Col1a1: 5’-CACCCTGAAAACTTTGCCCC-3’;
- shRen.713: 5’-GTATAGATAAGCATTATAATTCCTA-3’;
- shSmad4.591: 5’- GTATAAAGAATCCAATTCATCTT-3’;

- shSmad4.1599: 5'-TATTCACCTTTACATTCCAAC-3'.

Orthotopic transplantation assays: Mouse hosts were placed on doxycycline chow 5-7 days before transplantation. Mice were anesthetized and a survival surgery was performed to expose the pancreas. 1 x 10^5 tumor-derived cells were resuspended in 25 μL 1:1 DMEM (Gibco) : Matrigel (Corning) mix and injected in the tail of the pancreas of each mouse. Tumor engraftment and progression were monitored by palpation and ultrasound imaging. Where applicable, doxycycline withdrawal was done 4 weeks post-injection (corresponding to 3-5 mm tumor diameter) by switching the food source to regular chow. Primary tumor size was measured using ultrasound imaging (see below). At experimental endpoint, primary tumors, livers, and lungs were dissected and imaged under a dissection microscope for brightfield, mKate, and GFP fluorescence (Nikon SMZ1500 with NIS-Element software). Endpoint liver and lung metastasis burden were measured by calculating percent tumor area (mKate+) as a fraction of overall organ area. Euthanasia was performed upon reaching experimental or humane endpoints according to IACUC guidelines. All mice used were 6–8-week-old *Foxn1*^{nu} females.

Experimental metastasis assays: For liver metastasis assays, mice were anesthetized and a survival surgery was performed to expose the spleen. 4 x 10^5 tumor-derived cells resuspended in 20 μL PBS were injected in the splenic parenchyma of each mouse, followed by removal of the spleen and cauterization (splenectomy). Tumor engraftment and progression were monitored by palpation and ultrasound imaging. At experimental endpoint, livers were dissected and imaged under a dissection microscope for brightfield, mKate, and GFP fluorescence (Nikon SMZ1500 with NIS-Element software). Endpoint liver metastasis burden was measured by calculating percent tumor area (mKate+) as a fraction of overall organ area. For lung metastasis assays, mice were restrained, and 2.5 x 10^5 tumor-derived cells resuspended in 250 μL PBS were injected in the tail vein of each mouse. Bioluminescent imaging was used to monitor tumor engraftment and progression, and to measure tumor burden (see below). Where applicable, doxycycline withdrawal was done at 5 weeks (intrasplenic) or 6 weeks (tail vein) post-injection by switching the food source to regular chow. The occasional mice that developed tumors outside of the respective target organs were excluded from the analysis. Euthanasia was performed upon reaching experimental or humane endpoints according to IACUC guidelines. All mice used were 6–8-week-old *Foxn1*^{nu} females.

Animal imaging: For ultrasound, mice were anesthetized, then high-contrast ultrasound imaging was performed on a Vevo 2100 System with a MS250 13- to 24-MHz scanhead (VisualSonics). Images were acquired and tumor volume was measured using the Vevo 2100 software (VisualSonics). For bioluminescence, mice were injected with luciferin (5 mg/mouse, i.p.; Gold Technologies), anesthetized for 10 min, and then imaged on a IVIS Spectrum imager (Perkin Elmer). Images were acquired and bioluminescence signal was measured using the Living Image software (Perkin Elmer).

Histological, immunohistochemistry (IHC), and immunofluorescence (IF) analyses

Tissues were fixed overnight in 10% neutral buffered formalin (Fisher Scientific #22-050-105), embedded in paraffin, and cut into 5-µm sections. Hematoxylin & eosin staining was performed using standard protocols. For immunostaining, slides were heated for 30 min at 55°C, deparaffinized, rehydrated with an alcohol series, and subjected to antigen retrieval with citrate buffer (Vector Laboratories #H-3300) for 25 min in a pressure cooker set on high. Sections were treated with 3% H_2O_2 for 10 min followed by a wash in deionized water (for IHC only), washed in PBS, then blocked in PBS/0.1% Triton X-100/1% BSA. Primary antibodies were incubated overnight at 4°C in blocking buffer. The following primary antibodies were used: mKate2 (Evrogen #AB233, 1:1,000, IF), SMAD4 (Millipore #04-1033, 1:200, IHC), pSMAD2 (Cell Signaling #3108, 1:100, IHC), Ki67 (BD Pharmingen #550609, 1:200, IF), α–SMA (Sigma #A2547, 1:1,000, IF), KLF4 (Abcepta #AM2725A, 1:100, IHC), KLF5 (Abcam #ab137676, 1:500, IHC), KLF6 (Abcam #ab241385, 1:1,000, IHC), RUNX1 (Cell Signaling #8529, 1:500, IHC), RUNX2 (Cell Signaling #12556, 1:500, IHC), RUNX3 (Life Technologies #MA5-17169, 1:500, IHC).

For IHC, HRP-conjugated secondary antibodies (ImmPRESS kits, Vector Laboratories #MP7401 and #MP2400) were applied for 30-60 minutes at room temperature and visualized with DAB substrate (ImmPACT kit, Vector Laboratories #SK-4105). Tissues were then counterstained with hematoxylin, dehydrated, and mounted with Permount (Fisher Scientific #SP15-100). For IF, secondary Alexa Fluor 488 or 594 dye-conjugated antibodies (Life Technologies, 1:500) were applied for 60 min at room temperature. Tissues were then counterstained with DAPI and mounted with Prolong Gold Antifade Mountant (Life Technologies #P36930).

Images were acquired on a Zeiss AxioImager microscope using a 10X, 20X, or 40X objective, an ORCA/ER CCD camera (Hamamatsu Photonics), and ZEN 3.3 software (Zeiss). For pSMAD2 quantification, the number of pSMAD2+ cells per randomly chosen IHC-stained 40X field of view in a tumor region was manually counted. For Ki67 and α–SMA quantification, 20X fields of view co-stained for mKate2 were analyzed as follows: mKate2+ areas were selected and then (1) the corresponding Ki67+ cells were counted and calculated as a percentage of DAPI+ cells; or (2) the corresponding α–SMA+ area was measured and calculated as a percentage of the mKate2+ area. Image analyses were performed using ImageJ/FIJI (NIH, USA). The number of analyzed samples and statistical analyses used for each assay are specified in the respective figure legends.

Cloning

ESC-targeting plasmids were generated as described under 'GEMMs' above. The firefly luciferase reporter plasmid (pMSCV-Luc2-Blast) was generated by subcloning Luc2 from the pCDH-EF1-Luc2-P2A-tdTomato plasmid into the pMSCV-Blasticidin retroviral vector at the BglII/HpaI restriction sites with the inclusion of a Kozak sequence (GCCACC) upstream of the ATG start codon, using standard protocols. Constitutive *Renilla*, *Klf4*, and *Runx1* shRNAs were cloned in the pMSCV-mirE-SV40-Neomycin-BFP retroviral vector[65] at the XhoI/EcoRI restriction sites, using standard protocols. pCDH-EF1-Luc2-P2A-tdTomato was a gift from Kazuhiro Oka (Addgene #72486). pMSCV-Blasticidin was a gift from David Mu (Addgene #75085)[66]. All plasmids were authenticated by test digestion and Sanger sequencing. The following shRNA sequences were used:

- Ren: 5'-GTATAGATAAGCATTATAATTCCTA-3'
- Klf4: 5'-TATAAAAATAGACAATCAGCA-3'
- Runx1: 5'-AAATCAGAAGCATTCACAGTT-3'

Cell culture

All cells were maintained in a humidified incubator at 37°C with 5% CO_2.

Primary cell line derivation: Cell lines were generated from tumor-bearing pancreata, livers, or lungs of KC-shSmad4 or KC-shRen mice. Liver- and lung-derived lines were only used for sWGS analysis. Tumors were dissected, chopped with razor blades, and digested with 1 mg/ml collagenase V (Sigma-Aldrich) diluted in HBSS for 30-60 min, followed by 0.25% trypsin for 5-10 min. Digested tissues were washed with complete DMEM (DMEM, 10% FBS (Gibco), 1X penicillin–streptomycin), passed through a 100-µm filter, and cultured in complete DMEM on collagen-coated plates (PurCol, Advanced Biomatrix, 0.1 mg/ml) supplemented with 1 µg/ml doxycycline at 37°C. Cells were passaged at least five times to eliminate any non-tumor cells before using them in experiments. Primary cultures were authenticated by flow cytometry of engineered fluorescent alleles. All cultures were tested negative for mycoplasma.

Virus generation and transduction: For stable transduction of firefly luciferase and constitutive shRNA constructs, VSV-G pseudotyped retroviral supernatants were generated from transduced Phoenix-GP packaging cells and infections were performed as described elsewhere[67]. Infected cells were selected with 10 µg/ml Blasticidin for 5 days or 800 µg/ml G418 for 7 days, depending on the selection marker. Luciferase expression was confirmed by *in vitro* and *in vivo* bioluminescence. Knockdowns were confirmed by Western blot using standard procedures and the following antibodies: SMAD4 (Santa Cruz #sc-7966, 1:500), KLF4 (Abcepta #AM2725A, 1:1000), RUNX1 (Cell Signaling #8529, 1:1000), and Actin-HRP (Sigma #A3854, 1:20,000).

Tumor cell isolation

For RNA-seq, ATAC-seq, and scMultiomics analyses, tumor cells were freshly isolated from pancreata, livers, or lungs of KC-shSmad4 mice by FACS. Specifically, dissected tumors were finely chopped with scissors and incubated with digestion buffer containing 1 mg/ml collagenase V (Sigma #C9263), 2 U/ml dispase (Life Technologies #17105041) dissolved in HBSS with Mg^{2+} and Ca^{2+} (Thermo Fisher Scientific #14025076) supplemented with 0.1 mg/ml DNase I (Sigma #DN25-100MG) and 0.1 mg/ml soybean trypsin inhibitor (STI) (Sigma #T9003), in gentleMACS C Tubes (Miltenyi Biotec) for 42 min at 37 °C using the gentleMACS Octo Dissociator. After enzymatic dissociation, samples were washed with PBS and further digested with a 0.05% solution of Trypsin-EDTA (Gibco #15400) diluted in PBS for 5 min at 37°C. Trypsin digestion was neutralized with FACS buffer (10 mM EGTA and 2% FBS in PBS) containing DNase I and STI. Samples were then treated with RBC Lysis Buffer (Invitrogen #00-4333-57) for 5 min at room temperature, washed in FACS buffer containing DNase I and STI, and filtered through a 100-µm strainer. Cells were then resuspended in FACS buffer containing DNase I, STI and 300 nM DAPI as a live-cell marker, and filtered through a 40-μm strainer. Cells were sorted on a BD FACSAria I, BD FACSAria III (Becton Dickinson), or MA900 (Sony) cell sorter for mKate2 (co-expressing GFP for on-Dox shRNA mice), excluding DAPI+ cells. Cells were sorted directly into TRIzol LS (Thermo Fisher Scientific # 10296028) for RNA-seq or collected in 2% FBS in PBS for ATAC-seq.

Sparse whole-genome sequencing

Low-pass whole-genome sequencing was performed on genomic DNA freshly isolated from cultured cells as previously described[68]. Briefly, 1 μg of gDNA was sonicated on an E220 sonicator (Covaris; settings: 17Q, 75s), and libraries were prepared by standard procedure (end repair, addition of polyA, and adaptor ligation). Libraries were then purified (AMPure XP magnetic beads, Beckman Coulter), PCR enriched, and sequenced (Illumina HiSeq). Reads were mapped to the mouse genome, duplicates removed, and an average of 2.5 million reads were used for CNA determination with the Varbin algorithm[69].

RNA-sequencing

RNA extraction, RNA-seq library preparation and sequencing: Total RNA was isolated using TRIzol LS (Thermo Fisher Scientific # 10296028) followed by column clean-up using an RNeasy kit (Qiagen #74106). RNA was quantified using Nanodrop and its quality assessed by an Agilent 2100 BioAnalyzer. 100-500 ng of total RNA underwent polyA selection and TruSeq library preparation according to instructions provided by Illumina (TruSeq Stranded mRNA HT Kit, Illumina #20020595), with 15 cycles of PCR. Samples were barcoded and run on a HiSeq (Illumina) in a 75-bp SE run, with an average of 50 million reads per sample.

RNA-seq read mapping, differential expression analysis and heatmap visualization: RNA-seq data were analyzed by removing adaptor sequences using Trimmomatic[70]. RNA-seq reads were then aligned to GRCm38.91 (mm10) with STAR[63], and transcript count was quantified using featureCounts[71] to generate raw count matrix. Differential gene expression analysis was performed using the DESeq2 package[72] implemented in R (http://cran.r-project.org/). Principal component analysis (PCA) was performed using the DESeq2 package in R. Differentially expressed genes (DEGs) were determined by >2-fold change in gene expression with Benjamini-Hochberg adjusted p-value <0.05. For heatmap visualization of DEGs, samples were z-score normalized and plotted using the 'pheatmap' package in R.

Functional annotations of gene sets: Pathway enrichment analysis was performed using the indicated gene sets in enrichR[73]. Significance of the tests was assessed using combined score, described as $c = \log(p) * z$, where c is the combined score, p is Fisher's exact test p-value, and z is z-score for deviation from expected rank (shown in Figure 3 and ED Table 1), as well as by adjusted p-values defined by enrichR.

Intersection of RNA-seq and publicly available ChIP-seq data: To analyze transcriptional dynamics of SMAD2/3/4-binding targets, we used two publicly available PDAC datasets[9,28]. First, we extracted SMAD4-dependent SMAD2/3 ChIP-seq peaks that were significantly enriched in both studies (based on p-value < 1e-8 and an enrichment cut-off of > 8-fold increase of SMAD2/3 ChIP signal). These peaks were then associated with genes based on UCSC.mm10.knownGene using ChIPseeker package[74]: they were analyzed for genic location (annotatePeaks) and the nearest TSS was identified in order to annotate the peak to that gene. Finally, the resulting gene list was intersected with SMAD4-depednent differentially expressed genes in the present study. Log2FoldChange values between Smad4 ON vs. OFF in each organ were plotted, and the type of genomic binding region was color-annotated on the left (ED Fig. 3e).

Bulk ATAC-sequencing

Cell preparation, transposition reaction, ATAC-seq library construction and sequencing: A total of 60,000 mKate2$^+$ cells were isolated by FACS, washed once with 50 µl cold PBS and resuspended in 50 µl cold lysis buffer[75]. Cells were then centrifuged immediately for 10 min at 500*g* at 4°C, and the pellet of nuclei was subjected to transposition with Nextera Tn5 transposase (Illumina #FC-121–1030) for 30 min at 37°C, according to the manufacturer's instructions. DNA was eluted using a DNA Clean & Concentrator Kit in 21 µl elution buffer (Zymo Research #D4013). ATAC-seq libraries were prepared using the NEBNext High-Fidelity 2X PCR Master Mix (NEB M0541) as previously described[76]. Purified libraries were assessed using a Bioanalyzer High-Sensitivity DNA Analysis kit (Agilent). Approximately 50 million paired-end 150-bp reads (25 million each side) were sequenced per replicate on a HiSeq instrument (Illumina).

Mapping, peak calling and dynamic peak calling: FASTQ files were trimmed with trimGalore and cutadapt[77], and the filtered, pair-ended reads were aligned to mm10 with Bowtie2[78]. Peaks were called over input using MACS2[79], and only peaks with a p-value of ≤ 0.001 and outside the ENCODE blacklist region were kept. All peaks from all samples were merged by combining peaks within 500 bp of each. featureCounts[71] was used to count the mapped reads for each sample. The resulting peak atlas was normalized using DESeq2[72]. For comparison to DepthNorm, samples were normalized to 10 million mapped reads. Normalized bigWig files were created using the normalization factors from DESeq2 as previously described[80] and BEDTools genomeCoverageBed[81]. Dynamic ATAC peaks were called if they had an absolute $\log_2$-transformed fold change ≥ 0.58 and FDR ≤ 0.1.

ATAC-seq heat map clustering: The dynamic peaks determined by comparing pancreas, liver and lung +/-Dox (=Smad4 OFF/ON) conditions were clustered using *z*-scores and *k*-means of 4 and plotted using ComplexHeatmap[82].

TF motif enrichment analyses: Motif enrichment analysis was performed on differentially expressed ATAC peaks between liver and lung with the HOMER *de novo* motif discovery tool[83] using findMotifsGenome command with the parameters size = given and length = 8-12. Motif enrichment scores of the *de novo* predicted motifs identified from this analysis were calculated for ATAC gain or loss regions between liver and lung +/-Dox (=Smad4 OFF/ON) by applying the findMotifsGenome command with size = given and length = 8-12 in each peak-set.

Integration of ATAC-seq and RNA-seq data: We adapted a previously described workflow[84] that combines RNA-seq and ATAC-seq data with TF motif information to predict dominant TFs in liver vs. lung, as follows. For each organ, differential gene expression analysis between DoxOFF and DoxON was first performed individually (pancreas, liver, or lungs). Next, the resulting gene lists were intersected to identify organ-specific Smad4-responsive targets. EnrichR[73] was used to calculate enrichment scores for annotated TF targets using the ChEA_2016 database. RNA-score was defined as -log10(adjusted p-value). Separately, HOMER was used to compare ATAC-seq data between Liver(DoxON) and Lung(DoxON) samples to identify differentially expressed peaks (DEPs). Custom TF motifs were curated by combining all the pairwise DoxON comparisons between any two organs, and these motifs were then used to reannotate the DEPs to get consistent enrichment scores across known TFs. ATAC-score was defined as -log10(p-value). Finally, Combined-scores were calculated by multiplying the respective RNA- and ATAC-scores. To determine the net change in ATAC-RNA combined scores for KLF and RUNX in the pancreas (ED Fig. 4e), each TF's RNA- and ATAC-scores were calculated for Pancreas(DoxON) and Pancreas(DoxOFF) samples, multiplied to obtain combined scores for each Dox condition, and then the DoxON combined score was subtracted from the DoxOFF combined score.

scMultiome-sequencing

Cell preparation, transposition reaction, scATAC-seq library construction and sequencing: Single Cell Multiome ATAC + Gene Expression was performed with 10X Genomics system using Chromium Next GEM Single Cell Multiome Reagent Kit A (catalog #1000282) and ATAC Kit A (catalog #1000280) following Chromium Next GEM Single Cell Multiome ATAC + Gene Expression Reagent Kits User Guide and demonstrated protocol, Nuclei Isolation for Single Cell Multiome ATAC + Gene Expression Sequencing. Briefly, cells (viability 95%) were lysed for 4 min and resuspended in Diluted Nuclei Buffer (10X Genomics #PN-2000207). Lysis efficiency and nuclei concentration was evaluated on Countess II automatic cell counter by trypan blue and DAPI staining. 11,000 nuclei were loaded per transposition reaction, targeting recovery of 7,000 nuclei after sequencing. After transposition reaction, nuclei were encapsulated and barcoded. Next-generation sequencing libraries were constructed following User Guide and sequenced on an Illumina NovaSeq 6000 system.

Quality control and cell filtering: Nucleosome signal score and TSS enrichment score for each cell were computed. Criteria for retaining individual cells were: TSS enrichment score >0.3, a nucleosome signal score <1.5, total ATAC-seq counts between 1,000 and 200,000 (based on the 10X Cell Ranger ATAC-seq count matrix) and total RNA counts between 1,000 and 50,000.

scRNA-seq data preprocessing and cell annotation: Gene expression UMI count data were normalized using SCTransform, percent mitochondrial RNA content was regressed out, and PCA was performed on the SCTransform Pearson residual matrix using the RunPCA function in Seurat[86]. Nearest neighbors were identified using FindNeighbors with dims = 1:30. The R package BBKNN was used to remove batch effects between mouse samples and cell types were annotated using R packages celldex, SingleR, Azimuth, and custom gene sets[85,86]. Only cells annotated as Ductal or Acinar cells were retained for downstream analysis.

scATAC-seq data preprocessing: scATAC-seq peaks were identified using MACS2 (ref. 79) using default parameters. Peak calling was performed using the CallPeaks function in Signac. Any peaks overlapping annotated genomic blacklist regions for the mm10 genome were removed. Counts for the resulting peak set were quantified for each cell using the FeatureMatrix function in Signac. For scATAC-seq analysis, data were normalized using the RunTFIDF function, and the top features were identified using FindTopFeatures with min.cutoff = '10'. The RunSVD function was used to create the LSI space, and the resulting visualization was generated with UMAP using dims = 2:30. Clusters were identified using the "algorithm = 3" option.

Mapping of bulk ATAC-seq signatures: Differentially accessible peaks from bulk ATAC-seq were used to overlap with accessible peaks from scATAC-seq using intersectbed from bedtools[81]. Peaks with at least 1-bp overlaps were kept, and the top 5000 scATAC-seq peaks sorted based on significance were used to calculate ATAC signature scores using AddChromatinModule from Signac. LiverOPEN and LungOPEN cells were identified based on the signature score, and mutual exclusivity was calculated using the Chi-squared test.

Differential gene expression of scRNA-seq data: FindMarkers function from Seurat[86] with “min.pct = 0.1” was used to identify differentially expressed genes (DEGs) between LiverOPEN and LungOPEN cells identified from scATAC-seq data. These DEGs were used for calculating gene signature scores for human PDAC scRNA-seq data.

Public scATAC-seq analysis

Raw scATAC-seq data from ref. 45 were downloaded from the Gene Expression Omnibus (GSE137069). We kept cells using filters “min.cells > 10, min.features < 200”. Peaks that overlap with annotated genomic blacklist regions for the mm10 genome were identified. We retained cells using filters “peak_region_fragments > 3000, peak_region_fragments < 20000, pct_reads_in_peaks > 15, blacklist_ratio < 0.05, nucleosome_signal < 4, TSS.enrichment > 2”. Data were normalized using the RunTFIDF function, and the top features were identified using FindTopFeatures. We utilized the RunSVD function to create the LSI space, and the resulting visualization was generated with UMAP using dims = 2:30. Clusters were identified using the following options: "algorithm = 3 and resolution = 0.5 ". GeneActivity from Signac was used to derive an approximate gene activity matrix, and custom gene signatures were used to annotate cell types. Ductal and acinar cells were retained for downstream analysis. Differentially accessible peaks from bulk ATAC-seq were used to intersect with accessible peaks from scATAC-seq using intersectbed from bedtools[81]. Peaks with at least 1-bp overlap were kept, and top5000 scATAC-seq peaks sorted based on significance were used to calculate ATAC signature scores using AddChromatinModule from Signac.

Human PDAC analysis

Metastasis recurrence analysis: Previously reported data[29,30] were re-analyzed by site of recurrence (liver or lungs) and annotation of SMAD4 IHC status (positive or negative).

MSK-IMPACT analysis: Human datasets were obtained through the MSK Clinical Sequencing Cohort (MSK-IMPACT) via cBioPortal[87,88]. Samples were selected as follows: (1) Cancer Type: Pancreatic cancer, (2) Cancer Type Detailed: Pancreatic Ductal Adenocarcinoma, and (3) Genotype (SMAD4: MUT HOMDEL). Comparison of genomic annotations between SMAD4-altered and unaltered (i.e. the rest) samples, along with corresponding statistical analyses, were generated and visualized by the cBioPortal.

Bulk RNA-seq analysis: Bulk RNA-seq data from ref. 38 was retrieved from the Gene Expression Omnibus (GSE71729). First, we compared liver or lung metastasis samples to primary tumor samples to derive liver or lung metastasis-specific differentially expressed genes. Second, we performed analogous comparison of normal liver or lung vs. normal pancreas samples. The tumor and normal gene sets were then intersected to identify liver and lung tumor-specific signatures by filtering out differentially expressed genes in the normal organs. Pathway enrichment analysis was performed on the resulting organ/tumor-specific gene sets using enrichR[73], and the top enriched TFs were plotted using combined scores in bar plot format.

scRNA-seq analysis: Raw scRNA-seq data for ref. 47 were downloaded from the Gene Expression Omnibus (GSE155698). We kept cells using filters “min.cells > 100, nFeature_RNA > 500, nCount_RNA > 2500, percent.mt < 25”. SCTransform was used to regress out percent mitochondrial RNA, and nearest neighbors were found using FindNeighbors with dims = 1:30. Clusters were identified using resolution = 0.8, and cell types were annotated using R packages celldex, SingleR, Azimuth, and custom gene sets[85,86]. Only Ductal and acinar cells were retained for calculating gene signature scores from our mouse multiomic data using AddModuleScore from Seurat[86]. FeaturePlot was used to visualize the differential peak openings from our mouse LiverOPEN and LungOPEN ATAC signatures.

Statistics and reproducibility

Statistical analyses were performed with GraphPad Prism (v.9), R (v4.3.1) and Python (v.3.6.4). Pooled data are presented as mean ± s.e.m. Sample size, error bars and statistical methods are reported in the figure legends. Exact p-values are shown in figures or associated legends. RNA- and ATAC-seq data were analyzed as described in the respective sections above. No statistical methods were used to predetermine sample size in the mouse studies. Mice were randomized into different treatment groups (for example, +Dox vs. –Dox). The investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

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

DATA AVAILABILITY

All sequencing data will be deposited to GEO prior to publication. All other data supporting the findings of this study will be made available upon reasonable request.

CODE AVAILABILITY

No unique code was developed for this study.

ACKNOWLEDGEMENTS

We thank M. Chalarca, K. Rybczyk, and the MSKCC Research Animal Resource Center for technical assistance with mouse work; Z. Zhao, A. Kahn, Y. Zhao, S. Yang, S. Young, Y. Furuta, and the MSKCC Mouse Genetics Core for assistance with the generation of ESC-derived mouse models; R. Gardner and the MSKCC Flow Cytometry Core facility staff for assistance with cell sorting; O. Chaudhary and the MSKCC Single-cell Analytics Innovation Lab (SAIL) for technical assistance with scMultiomics; C. Burdziak and R. Sharma for advice on scMultiome analysis; F.J. Sánchez-Rivera, R. Mezzadra, J. P. Morris IV, J. Reyes and the rest of the Lowe laboratory for advice and helpful discussions; and J. Novak for manuscript editing. Mouse and human illustrations were created with BioRender (https://biorender.com/). We acknowledge Core funding by an NCI Cancer Center Support Grant (P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology. K.M.T. was supported by the Jane Coffin Childs Memorial Fund for Medical Research, a Shulamit Katzman Endowed Postdoctoral Research Fellowship, and an NIH K99/R00 Transition to Independence Award (5K99CA266939). F.M.B. was supported by a GMTEC Postdoctoral Fellowship, an MSKCC Translational Research Oncology Training Fellowship (5T32CA160001) and a Young Investigator Award by the Edward

P. Evans Foundation. D.A.C. was supported by the Spanish Fundación Ramón Areces Postdoctoral Fellowship and is recipient of the La Caixa Postdoctoral Junior Leader Fellowship (LCF/BQ/PI20/11760006). G.L. was supported by an NIH F32 grant (1F32CA177072) and an American Cancer Society Fellowship (ACS PF-13-037-01-DMC). T.B. received support from the William C. and Joyce C. O'Neil Charitable Trust and the MSK Single Cell Sequencing Initiative. This work was supported by MSKCC's David Rubenstein Center for Pancreatic Research Pilot Project (S.W.L.); the Alan and Sandra Gerry Metastasis and Tumor Ecosystems Center (J.M. and S.W.L.); the Lustgarten Foundation Research Investigator Award (S.W.L.); the Agilent Thought Leader Program (S.W.L.); and NIH grants P01CA13106 (S.W.L.) and R35CA252978 (J.M.). S.W.L. is the Geoffrey Beene Chair of Cancer Biology, D.P. is the Alan and Sandra Gerry Endowed Chair of Computational Biology, and J.M. is the Marie-Josée and Henry R. Kravis Chair. S.W.L. and D.P. are Investigators of the Howard Hughes Medical Institute.

AUTHOR CONTRIBUTIONS

K.M.T. conceived the study, designed and performed experiments, analyzed data, and wrote the manuscript. F.M.B., D.A.C. and G.L. designed and performed experiments, and analyzed data. Y.-J.H. analyzed all bulk RNA-seq and scRNA/ATAC-seq data, including bulk RNA-ATAC-seq integration and analysis of public datasets. R.K. analyzed bulk ATAC-seq data. T.B. analyzed sparse WGS data. J.S., S.T., A.N.W. and W.L. assisted in experiments. J.E.W. helped interpret histopathology data. I.M. performed scRNA/ATAC-seq, with supervision and critical input from D.P. and R.C. N.D. assisted with data analysis. J.M., D.P. and C.A.I.-D. provided critical intellectual input and guidance in experimental design and data interpretation. S.W.L. conceived and supervised the study and wrote the manuscript. All authors read the manuscript.

COMPETING INTERESTS

S.W.L. is a consultant for Fate Therapeutics, and is a consultant and holds equity in Blueprint Medicines, ORIC Pharmaceuticals, Mirimus, PMV Pharmaceuticals, Faeth Therapeutics, and Senescea Therapeutics. R.P.K. is a co-founder of and consultant for Econic Biosciences. D.P. is on the scientific advisory board of Insitro. J.M. holds equity in Scholar Rock. All other authors have no competing interests. None of these affiliations represent a conflict of interest with respect to the design or execution of this study or interpretation of data presented in this report.

FIGURE LEGENDS

Figure 1. A *Smad4*-restorable genetically engineered mouse model of PDAC.

(a) Schematic of GEMM alleles. rtTA3 = reverse tetracycline transactivator, 3rd generation; Dox = doxycycline; TRE = tetracycline response element.

(b) Overall survival of KC-shRen mice and KC-shSmad4 mice (expressing one of two independent Smad4 shRNAs: #591 or #1599) after Dox administration (n=7 KC-shRen mice; n=11 KC-shSmad4 mice). Statistical analysis by log-rank test.

(c) Representative hematoxylin and eosin (H&E) staining of primary tumors and liver and lung metastases from KC-shSmad4 mice.

(d) Sparse whole-genome sequencing (sWGS) analysis of the *Cdkn2a/b* locus in KC-shSmad4 tumor-derived cell lines (n=10 independent mice).

(e) Representative Western blot analysis of Dox response *in vitro* in two independent KC-shSmad4 cell lines.

Extended Data Figure 1.

(a) Overall (pie chart) and organ-specific (stacked bar graph) frequency of metastasis in KC-shSmad4 mice.

(b) Representative macroscopic images of tumor-bearing pancreas, liver, and lungs from KC-shSmad4 mice (+Dox) at endpoint. Fluorescent images match the corresponding brightfield images. Data are representative of 10 (pancreas), 6 (liver), and 3 (lungs) independent mice.

(c) Representative immunohistochemistry (IHC) staining for SMAD4 in primary and metastatic KC-shSmad4 tumors (+Dox). Dashed lines demarcate metastases. Data are representative of 3 independent mice.

(d) sWGS analysis of genome-wide copy number alterations in KC-shSmad4 primary tumor-derived cell lines (n=10 independent mice). Frequency plot is shown on the top and individual sample tracks are provided on the bottom.

(e) Representative sWGS analysis of genome-wide copy number alterations in KC-shSmad4 tumor-derived cell lines from matched primary tumors, liver and lung metastases. The *Kras* and *Cdkn2a/b* loci are highlighted (arrows). Data are representative of 3 independent mice.

(f) Frequency of the indicated homozygous deletions in *SMAD4*-altered (mutated or homozygously deleted, n=959) or wild-type (WT, n=3188) PDAC tumors in the MSK-IMPACT cohort (*CDKN2A/B*, $p<10^{-10}$; $q<10^{-10}$). *CDKN2A/B* and their adjacent gene *MTAP* are highlighted.

Figure 2. *Smad4* restoration has organ-specific effects on tumor growth.

(a) Schematic of orthotopic, intrasplenic, and tail vein injection experiments.

(b) Analysis of primary tumor growth after orthotopic transplantation of KC-shSmad4 cells and subsequent *Smad4* restoration. (Left) Fold-change quantifications of tumor volume at day 30 vs. day 0 of Dox withdrawal (n=8 independent mice per group). Different color shading indicates independent cell lines. (Right) Representative ultrasound images of tumors (demarcated by dashed yellow lines).

(c) Analysis of liver metastasis burden after intrasplenic injections of KC-shSmad4 cells and subsequent *Smad4* restoration. (Left) Percent-area quantifications at day 30 after Dox withdrawal (n=6 independent mice per group). Different color shading indicates independent cell lines. (Right) Representative macroscopic images of tumor-bearing livers.

(d) Analysis of lung metastasis burden after tail vein injections of KC-shSmad4 cells and subsequent *Smad4* restoration. (Left) Fold-change quantifications of bioluminescent signal at day 30 vs. day 0 of Dox withdrawal (n=8, 7 independent mice per group, respectively). Different color shading indicates independent cell lines. (Right) Representative bioluminescent images of tumor-bearing mice.

Statistical analysis: (b) Unpaired two-tailed t-test; (c, d) Mann-Whitney test.

Extended Data Figure 2.

(a) Proliferation assays of KC-shRen and KC-shSmad4 tumor-derived cell lines +/-Dox +/-TGFβ *in vitro* (n=3 and 4 independently derived cell lines for KC-shRen and KC-shSmad4, respectively).

(b) Representative brightfield images of KC-shRen and KC-shSmad4 cells after TGFβ treatment +/-Dox for 6 days. Data are representative of 3 independent cell lines per genotype.

(c) Representative H&E staining of liver and lung metastases generated by intrasplenic or tail vein injections of KC-shSmad4 cell lines. Data are representative of 20 metastases across 4 independent mice per organ site.

(d) Representative H&E staining of liver and lung metastases from KC-shSmad4 cells without (Smad4 OFF) or with *Smad4* restoration (Smad4 ON). Data are representative of 6 independent mice per organ site.

(e) Schematic of orthotopic experiments with KC-shRen and KC-shSmad4 cells for analysis of metastasis burden.

(f) Analysis of liver metastasis burden after orthotopic injections of KC-shRen or KC-shSmad4 cells with or without subsequent Dox withdrawal. (Left) Percent-area quantifications at endpoint (n=5, 3, 8, and 8 independent mice for groups shown, left to right). Different color shading indicates independent cell lines. (Right) Representative macroscopic images of tumor-bearing livers. Insets show tumor-constitutive mKate2 reporter.

(g) Analysis of lung metastasis burden after orthotopic injections of KC-shRen or KC-shSmad4 cells with or without subsequent Dox withdrawal. (Left) Percent-area quantifications at endpoint (n=4, 5, 8, and 8 independent mice for groups shown, left to right). Different color shading indicates independent cell lines. (Right) Representative macroscopic images of tumor-bearing lungs. Insets show tumor-constitutive mKate2 reporter.

(h) (Left) Schematic of metastasis recurrence studies in human PDAC[29,30]. (Right) Fraction of patients whose tumors stained positive or negative for SMAD4 by IHC, according to site of recurrence. Exact numbers are indicated on each bar.

Statistical analysis: (a) Two-way ANOVA; (f, g) Unpaired two-tailed t-test; (h) Fisher's exact test.

Figure 3. *Smad4* induces different transcriptional programs in liver vs. lung metastases.

(a) Gene ontology (GO) analysis of upregulated genes in liver or lung metastases at 7 or 14 days after Dox withdrawal. Combined scores and p-values for the top KEGG (left) or Hallmark (right) pathways are shown for Smad4 ON vs. OFF in each organ and timepoint. Complete GO lists are provided in ED Table 1.

(b) Heatmap of representative genes from *Smad4*'s cytostatic/apoptotic (tumor-suppressive) and fibrogenic (tumor-promoting) transcriptional programs. Average RNA-seq $\log_2$ fold-change and p-values are shown for Smad4 ON vs. OFF in each organ and timepoint.

(c-d) Representative immunofluorescence staining for Ki67 (c) or α-SMA (d) in KC-shSmad4 liver and lung metastases +/-*Smad4* restoration. mKate2 labels tumor cells. Quantifications are shown on the right (n=10-16 independent tumors from 3-4 mice). Ki67 analysis was performed 7 days after Dox withdrawal; α-SMA analysis was performed at experimental endpoint (30 days for liver; 45 days for lungs). Statistical analysis by unpaired two-tailed t-test (Ki67) or Mann-Whitney test (α-SMA).

Extended Data Figure 3.

(a) Overlap between upregulated genes in tumor cells from the pancreas, liver, or lungs at 7 and 14 days after Dox withdrawal. Numbers reflect genes in each category. Complete gene lists are provided in ED Table 2. DEGs = differentially expressed genes.

(b) Normalized *Smad4* mRNA levels (RNA-seq data; tpm = transcripts per million) in KC-shSmad4 tumor cells from the pancreas, liver, or lungs +/-Dox for 7 or 14 days.

(c) Representative IHC staining for pSMAD2 in KC-shSmad4 tumors (+Dox) from the pancreas, liver, or lungs. Quantifications (%pSMAD2 cells) are shown on the right (n=6 independent metastases from two mice). Statistical analysis by one-way ANOVA.

(d) Heatmap of differentially expressed genes from the indicated KEGG pathways. Average log2 fold-change and p-values are shown for each organ and timepoint.

(e) Heatmap of SMAD4-dependent SMAD2/3 ChIP target genes that are differentially upregulated in liver vs. lung metastases at 7 or 14 days after Dox withdrawal. Average RNA-seq $\log_2$ fold-change values are shown for Smad4 ON vs. OFF in each organ and timepoint. The type of genomic binding region is color-annotated on the left. Complete gene lists are provided in ED Table 2.

(f) Fraction of upregulated genes in liver and/or lung metastases at 7 or 14 days of Dox withdrawal that are SMAD4-dependent SMAD2/3 ChIP-seq targets. Absolute number of genes per group is shown in parentheses.

Figure 4. Liver and lung metastases harbor distinct chromatin states.

(a) Heatmap of differentially accessible chromatin regions in tumor (mKate2$^+$) cells from the pancreas, liver, or lungs. Smad4 ON corresponds to 14 days of Dox withdrawal. Each column represents an independent mouse. A complete list of differentially accessible ATAC-seq peaks is provided in ED Table 3.

(b) Top-scoring TF motifs identified by HOMER *de novo* motif analysis of ATAC-seq peaks enriched in liver vs. lung metastases. The numbers in parentheses indicate enrichment p-values.

(c) ATAC-RNA combined scores for the indicated TF families in liver vs. lung metastases. This metric infers the probability that a given TF family with a significantly enriched motif in the ATAC-open regions impacts SMAD4-induced gene expression changes, based on a consistent RNA-seq change in the Smad4 ON vs. OFF comparison (see Methods for details). Top TFs scored using the HOMER *de novo* motif analysis in liver vs. lungs are shown.

(d) RNA-seq combined scores for the indicated TF families in liver and lung metastases from human PDAC patients[38]. This metric infers the activity of a given TF family, based on enrichment/depletion of its predicted target genes in the respective metastases vs. primary tumors but not in the corresponding normal tissues (see Methods for details). Top TFs scored using JASPAR/TRANSFAC position weight matrix (PWM) data are shown.

(e) Representative IHC staining for KLF4 and RUNX1 in KC-shSmad4 (+Dox) liver or lung metastases. Higher magnifications of dashed areas are shown on the right to highlight tumor-specific nuclear signal. Data are representative of 20 metastases across 4 independent mice.

(f) Representative IHC staining for KLF4 and RUNX1 in serial sections of a KC-shSmad4 (+Dox) primary tumor. Higher magnifications of dashed areas are shown on the right to highlight mutual exclusivity. Data are representative of 4 independent mice.

(g) UMAP (uniform manifold approximation and projection) visualization of scATAC-seq profiles of KC-shSmad4 (+Dox) primary tumor cells (mKate2$^+$). Signature scores based on liver- or lung-specific ATAC-open peaks from bulk ATAC-seq data are displayed in color per individual cell.

Data are representative of 3 independent mice.

Extended Data Figure 4.

(a) Principal component analysis (PCA) of ATAC-seq data from tumor (mKate2$^+$) cells isolated from the pancreas, liver, or lungs. Analysis is based on peak normalization. Each sample corresponds to an independent mouse. Circled areas highlight separation based on organ site.

(b) Number of statistically significant differentially accessible ATAC-seq peaks. Exact number of gained/lost peaks are shown for each comparison (absolute fold change $\geq$1.5; FDR $\leq$0.1).

(c) Representative ATAC-seq tracks at loci with liver-specific, lung-specific, or liver/lung-shared chromatin opening in KC-shSmad4 tumor cells (mKate2$^+$GFP$^+$). Housekeeping genes are shown as a reference. Data are representative of 3 independent tumors. y-axis scale range is indicated per lane as normalized read counts.

(d) Representative IHC staining for the indicated TFs in KC-shSmad4 (+Dox) liver or lung metastases. Data are representative of 20 metastases across 4 independent mice.

(e) Net change in ATAC-RNA combined scores for the KLF and RUNX TF families in Smad4 ON vs. OFF primary tumors at day 14. This metric infers the probability that a given TF family is impacted by Smad4 restoration, based on a combined change in ATAC-seq motif accessibility and RNA-seq levels of its predicted target genes (see Methods for details).

(f) Mutual exclusivity of liver- and lung-specific open chromatin signatures in primary tumors. The plot shows the distribution of cells from KC-shSmad4 (+Dox) primary tumors according to their enrichment of liver- (LiverOPEN) and lung-specific (LungOPEN) open chromatin signatures derived from bulk ATAC-seq (see Methods for details). Each dot corresponds to an independent cell. The proportions of cells with liver-only (blue) and lung-only (red) signatures were compared to those with both liver and lung signatures (green) and those with neither signature (purple) using Chi-squared test.

(g) UMAP visualization of scATAC-seq profiles of *Kras*-mutant pancreatic epithelial cells (mKate2$^+$) +/- cerulein injury (n=1 mouse for each) as described in ref. 45 (see Methods for details). Signature scores based on liver- or lung-specific ATAC-open peaks from bulk ATAC-seq data are displayed in color per individual cell.

(h) UMAP visualization of scRNA-seq profiles of human primary PDAC tumor cells from ref. 47 (see Methods for details). Signature scores based on scRNA-seq profiles of mouse liver- or lung-specific ATAC-open cell populations (from the matching scATAC-seq multiomics data) are displayed in color per individual cell.

Figure 5. *Klf4* depletion is sufficient to reverse *Smad4* function in liver metastasis.

(a) Western blot analysis of shRNA efficiency. Actin was used as a loading control.

(b) Quantifications of metastasis burden (fold-change of % tumor area) at day 30 after Dox withdrawal (n=8, 8, 7, 8, 9, 9 independent mice per group, respectively). Statistical analysis by unpaired two-tailed t-test.

(c) Representative macroscopic images of tumor-bearing livers +/-*Smad4* restoration on the background of the indicated shRNA-based knockdowns. Inset shows shSmad4-linked GFP reporter. Data are representative of the number of independent mice per group specified in (b).

(d) Representative H&E staining of tumor-bearing livers +/-*Smad4* restoration on the background of the indicated shRNA-based knockdowns. Data are representative of 6 independent mice per group.

Extended Data Figure 5.

(a) Proliferation of KC-shSmad4 cell lines harboring the indicated stable shRNAs +/-Dox and +/-TGFβ *in vitro* (average of three technical replicates). Statistical analysis by one-way ANOVA.

(b) Schematic of intrasplenic experiments with KC-shSmad4 cells harboring TF knockdowns.

(c) Quantifications of baseline metastasis burden (% tumor area) after intrasplenic injections of KC-shSmad4 cells under continuous Dox administration (n=8, 7, and 9 independent mice per group, respectively). Livers were harvested 64-66 days after injection. Statistical analysis by unpaired two-tailed t-test.

(d) Model of the organ-specific interplay between *SMAD4* and KLF4/RUNX1-associated chromatin states. In the liver, active KLF4 allows SMAD4 to act at genes that promote cytostasis, thus necessitating the inactivation of *Smad4*. In the lungs, more speculatively, active RUNX1 allows SMAD4 to act at pro-fibrotic genes unopposed by its cytostatic program.

REFERENCES

1 Birkbak, N. J. & McGranahan, N. Cancer Genome Evolutionary Trajectories in Metastasis. *Cancer Cell* **37**, 8-19, doi:10.1016/j.ccell.2019.12.004 (2020).

2 Patel, S. A., Rodrigues, P., Wesolowski, L. & Vanharanta, S. Genomic control of metastasis. *Br J Cancer* **124**, 3-12, doi:10.1038/s41416-020-01127-6 (2021).

3 Turajlic, S. & Swanton, C. Metastasis as an evolutionary process. *Science* **352**, 169-175, doi:10.1126/science.aaf2784 (2016).

4 Hoadley, K. A. *et al.* Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* **173**, 291-304 e296, doi:10.1016/j.cell.2018.03.022 (2018).

5 Nguyen, B. *et al.* Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell* **185**, 563-575 e511, doi:10.1016/j.cell.2022.01.003 (2022).

6 Bardeesy, N. *et al.* Smad4 is dispensable for normal pancreas development yet critical in progression and tumor biology of pancreas cancer. *Genes Dev* **20**, 3130-3146, doi:10.1101/gad.1478706 (2006).

7 Blackford, A. *et al.* SMAD4 gene mutations are associated with poor prognosis in pancreatic cancer. *Clin Cancer Res* **15**, 4674-4679, doi:10.1158/1078-0432.CCR-09-0227 (2009).

8 Dai, C. *et al.* SMAD4 represses FOSL1 expression and pancreatic cancer metastatic colonization. *Cell Rep* **36**, 109443, doi:10.1016/j.celrep.2021.109443 (2021).

9 David, C. J. *et al.* TGF-beta Tumor Suppression through a Lethal EMT. *Cell* **164**, 1015-1030, doi:10.1016/j.cell.2016.01.009 (2016).

10 Iacobuzio-Donahue, C. A. *et al.* Dpc4 protein in mucinous cystic neoplasms of the pancreas: frequent loss of expression in invasive carcinomas suggests a role in genetic progression. *Am J Surg Pathol* **24**, 1544-1548, doi:10.1097/00000478-200011000-00011 (2000).

11 Iacobuzio-Donahue, C. A. *et al.* DPC4 gene status of the primary carcinoma correlates with patterns of failure in patients with pancreatic cancer. *J Clin Oncol* **27**, 1806-1813, doi:10.1200/JCO.2008.17.7188 (2009).

12 Izeradjene, K. *et al.* Kras(G12D) and Smad4/Dpc4 haploinsufficiency cooperate to induce mucinous cystic neoplasms and invasive adenocarcinoma of the pancreas. *Cancer Cell* **11**, 229-243, doi:10.1016/j.ccr.2007.01.017 (2007).

13 Kojima, K. *et al.* Inactivation of Smad4 accelerates Kras(G12D)-mediated pancreatic neoplasia. *Cancer Res* **67**, 8121-8130, doi:10.1158/0008-5472.CAN-06-4167 (2007).

14 Whittle, M. C. *et al.* RUNX3 Controls a Metastatic Switch in Pancreatic Ductal Adenocarcinoma. *Cell* **161**, 1345-1360, doi:10.1016/j.cell.2015.04.048 (2015).

15 Ganesh, K. & Massague, J. Targeting metastatic cancer. *Nat Med* **27**, 34-44, doi:10.1038/s41591-020-01195-4 (2021).

16 Lambert, A. W., Pattabiraman, D. R. & Weinberg, R. A. Emerging Biological Principles of Metastasis. *Cell* **168**, 670-691, doi:10.1016/j.cell.2016.11.037 (2017).

17 Massague, J. & Ganesh, K. Metastasis-Initiating Cells and Ecosystems. *Cancer Discov* **11**, 971-994, doi:10.1158/2159-8290.CD-21-0010 (2021).

18 Reiter, J. G. *et al.* Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033-1037, doi:10.1126/science.aat7171 (2018).

19 Priestley, P. *et al.* Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature* **575**, 210-216, doi:10.1038/s41586-019-1689-y (2019).

20 Gupta, P. B., Pastushenko, I., Skibinski, A., Blanpain, C. & Kuperwasser, C. Phenotypic Plasticity: Driver of Cancer Initiation, Progression, and Therapy Resistance. *Cell Stem Cell* **24**, 65-78, doi:10.1016/j.stem.2018.11.011 (2019).

21 Lu, W. & Kang, Y. Epithelial-Mesenchymal Plasticity in Cancer Progression and Metastasis. *Dev Cell* **49**, 361-374, doi:10.1016/j.devcel.2019.04.010 (2019).

22 Yuan, S., Norgard, R. J. & Stanger, B. Z. Cellular Plasticity in Cancer. *Cancer Discov* **9**, 837-851, doi:10.1158/2159-8290.CD-19-0015 (2019).
23 Massague, J. & Sheppard, D. TGF-beta signaling in health and disease. *Cell* **186**, 4007-4037, doi:10.1016/j.cell.2023.07.036 (2023).
24 Kleeff, J. *et al.* Pancreatic cancer. *Nat Rev Dis Primers* **2**, 16022, doi:10.1038/nrdp.2016.22 (2016).
25 Barriga, F. M. *et al.* MACHETE identifies interferon-encompassing chromosome 9p21.3 deletions as mediators of immune evasion and metastasis. *Nat Cancer* **3**, 1367-1385, doi:10.1038/s43018-022-00443-5 (2022).
26 Mueller, S. *et al.* Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. *Nature* **554**, 62-68, doi:10.1038/nature25459 (2018).
27 Zhong, Y. *et al.* Mutant p53 Together with TGFbeta Signaling Influence Organ-Specific Hematogenous Colonization Patterns of Pancreatic Cancer. *Clin Cancer Res* **23**, 1607-1620, doi:10.1158/1078-0432.CCR-15-1615 (2017).
28 Su, J. *et al.* TGF-beta orchestrates fibrogenic and developmental EMTs via the RAS effector RREB1. *Nature* **577**, 566-571, doi:10.1038/s41586-019-1897-5 (2020).
29 Arnaoutakis, G. J. *et al.* Pulmonary resection for isolated pancreatic adenocarcinoma metastasis: an analysis of outcomes and survival. *J Gastrointest Surg* **15**, 1611-1617, doi:10.1007/s11605-011-1605-8 (2011).
30 Herman, J. M. *et al.* Smad4 Loss Correlates With Higher Rates of Local and Distant Failure in Pancreatic Adenocarcinoma Patients Receiving Adjuvant Chemoradiation. *Pancreas* **47**, 208-212, doi:10.1097/MPA.0000000000000985 (2018).
31 David, C. J. & Massague, J. Contextual determinants of TGFbeta action in development, immunity and cancer. *Nat Rev Mol Cell Biol* **19**, 419-435, doi:10.1038/s41580-018-0007-0 (2018).
32 Lee, M. H., Reynisdottir, I. & Massague, J. Cloning of p57KIP2, a cyclin-dependent kinase inhibitor with unique domain structure and tissue distribution. *Genes Dev* **9**, 639-649, doi:10.1101/gad.9.6.639 (1995).
33 Scandura, J. M., Boccuni, P., Massague, J. & Nimer, S. D. Transforming growth factor beta-induced cell cycle arrest of human hematopoietic cells requires p57KIP2 up-regulation. *Proc Natl Acad Sci U S A* **101**, 15231-15236, doi:10.1073/pnas.0406771101 (2004).
34 Huang, Y. H. *et al.* ID1 Mediates Escape from TGFbeta Tumor Suppression in Pancreatic Cancer. *Cancer Discov* **10**, 142-157, doi:10.1158/2159-8290.CD-19-0529 (2020).
35 Zaret, K. S. Pioneer Transcription Factors Initiating Gene Network Changes. *Annu Rev Genet* **54**, 367-385, doi:10.1146/annurev-genet-030220-015007 (2020).
36 Tetreault, M. P., Yang, Y. & Katz, J. P. Kruppel-like factors in cancer. *Nat Rev Cancer* **13**, 701-713, doi:10.1038/nrc3582 (2013).
37 Yi, H., He, Y., Zhu, Q. & Fang, L. RUNX Proteins as Epigenetic Modulators in Cancer. *Cells* **11**, doi:10.3390/cells11223687 (2022).
38 Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nat Genet* **47**, 1168-1178, doi:10.1038/ng.3398 (2015).
39 Ferrer, J. & Real, F. X. The cis-regulatory switchboard of pancreatic ductal cancer. *EMBO J* **35**, 558-560, doi:10.15252/embj.201693999 (2016).
40 He, P., Yang, J. W., Yang, V. W. & Bialkowska, A. B. Kruppel-like Factor 5, Increased in Pancreatic Ductal Adenocarcinoma, Promotes Proliferation, Acinar-to-Ductal Metaplasia, Pancreatic Intraepithelial Neoplasia, and Tumor Growth in Mice. *Gastroenterology* **154**, 1494-1508 e1413, doi:10.1053/j.gastro.2017.12.005 (2018).
41 Wei, D. *et al.* KLF4 Is Essential for Induction of Cellular Identity Change and Acinar-to-Ductal Reprogramming during Early Pancreatic Carcinogenesis. *Cancer Cell* **29**, 324-338, doi:10.1016/j.ccell.2016.02.005 (2016).

42 Ghaleb, A. M. *et al.* Kruppel-like factors 4 and 5: the yin and yang regulators of cellular proliferation. *Cell Res* **15**, 92-96, doi:10.1038/sj.cr.7290271 (2005).

43 Cheng, Y. *et al.* RUNX1 promote invasiveness in pancreatic ductal adenocarcinoma through regulating miR-93. *Oncotarget* **8**, 99567-99579, doi:10.18632/oncotarget.20433 (2017).

44 Doffo, J. *et al.* NOXA expression drives synthetic lethality to RUNX1 inhibition in pancreatic cancer. *Proc Natl Acad Sci U S A* **119**, doi:10.1073/pnas.2105691119 (2022).

45 Alonso-Curbelo, D. *et al.* A gene-environment-induced epigenetic program initiates tumorigenesis. *Nature*, doi:10.1038/s41586-020-03147-x (2021).

46 LaFave, L. M. *et al.* Epigenomic State Transitions Characterize Tumor Progression in Mouse Lung Adenocarcinoma. *Cancer Cell* **38**, 212-228 e213, doi:10.1016/j.ccell.2020.06.006 (2020).

47 Steele, N. G. *et al.* Multimodal Mapping of the Tumor and Peripheral Blood Immune Landscape in Human Pancreatic Cancer. *Nat Cancer* **1**, 1097-1112, doi:10.1038/s43018-020-00121-4 (2020).

48 Vitellius, C. *et al.* Impact of Driver Mutations on the Evolution of Isolated Metachronous Lung Metastasis of Pancreatic Ductal adenocarcinoma. *Mol Diagn Ther* **24**, 443-449, doi:10.1007/s40291-020-00472-9 (2020).

49 Carstens, J. L. *et al.* Stabilized epithelial phenotype of cancer cells in primary tumors leads to increased colonization of liver metastasis in pancreatic cancer. *Cell Rep* **35**, 108990, doi:10.1016/j.celrep.2021.108990 (2021).

50 Reichert, M. *et al.* Regulation of Epithelial Plasticity Determines Metastatic Organotropism in Pancreatic Cancer. *Dev Cell* **45**, 696-711 e698, doi:10.1016/j.devcel.2018.05.025 (2018).

51 Tang, Q., Su, Z., Gu, W. & Rustgi, A. K. Mutant p53 on the Path to Metastasis. *Trends Cancer* **6**, 62-73, doi:10.1016/j.trecan.2019.11.004 (2020).

52 Pitarresi, J. R. *et al.* PTHrP Drives Pancreatic Cancer Growth and Metastasis and Reveals a New Therapeutic Vulnerability. *Cancer Discov* **11**, 1774-1791, doi:10.1158/2159-8290.CD-20-1098 (2021).

53 Maddipati, R. *et al.* MYC Levels Regulate Metastatic Heterogeneity in Pancreatic Adenocarcinoma. *Cancer Discov* **12**, 542-561, doi:10.1158/2159-8290.CD-20-1826 (2022).

54 Baggiolini, A. *et al.* Developmental chromatin programs determine oncogenic competence in melanoma. *Science* **373**, eabc1048, doi:10.1126/science.abc1048 (2021).

55 Patel, S. A. *et al.* The renal lineage factor PAX8 controls oncogenic signalling in kidney cancer. *Nature* **606**, 999-1006, doi:10.1038/s41586-022-04809-8 (2022).

56 Sack, L. M. *et al.* Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* **173**, 499-514 e423, doi:10.1016/j.cell.2018.02.037 (2018).

57 Weiss, J. M. *et al.* Anatomic position determines oncogenic specificity in melanoma. *Nature* **604**, 354-361, doi:10.1038/s41586-022-04584-6 (2022).

58 Osorio, J. C. *et al.* Lesion-Level Response Dynamics to Programmed Cell Death Protein (PD-1) Blockade. *J Clin Oncol* **37**, 3546-3555, doi:10.1200/JCO.19.00709 (2019).

59 Pires da Silva, I. *et al.* Site-specific response patterns, pseudoprogression, and acquired resistance in patients with melanoma treated with ipilimumab combined with anti-PD-1 therapy. *Cancer* **126**, 86-97, doi:10.1002/cncr.32522 (2020).

60 Seifert, H. *et al.* Extrinsic factors can mediate resistance to BRAF inhibition in central nervous system melanoma metastases. *Pigment Cell Melanoma Res* **29**, 92-100, doi:10.1111/pcmr.12424 (2016).

61 Zhou, J., Li, Q. & Cao, Y. Spatiotemporal Heterogeneity across Metastases and Organ-Specific Response Informs Drug Efficacy and Patient Survival in Colorectal Cancer. *Cancer Res* **81**, 2522-2533, doi:10.1158/0008-5472.CAN-20-3665 (2021).

62 Saborowski, M. *et al.* A modular and flexible ESC-based mouse model of pancreatic cancer. *Genes Dev* **28**, 85-97, doi:10.1101/gad.232082.113 (2014).

63 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
64 Gertsenstein, M. *et al.* Efficient generation of germ line transmitting chimeras from C57BL/6N ES cells by aggregation with outbred host embryos. *PLoS One* **5**, e11260, doi:10.1371/journal.pone.0011260 (2010).
65 Loizou, E. *et al.* A Gain-of-Function p53-Mutant Oncogene Promotes Cell Fate Plasticity and Myeloid Leukemia through the Pluripotency Factor FOXH1. *Cancer Discov* **9**, 962-979, doi:10.1158/2159-8290.CD-18-1391 (2019).
66 Kendall, J. *et al.* Oncogenic cooperation and coamplification of developmental transcription factor genes in lung cancer. *Proc Natl Acad Sci U S A* **104**, 16663-16668, doi:10.1073/pnas.0708286104 (2007).
67 Zuber, J. *et al.* RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524-528, doi:10.1038/nature10334 (2011).
68 Baslan, T. *et al.* Optimizing sparse sequencing of single cells for highly multiplex copy number profiling. *Genome Res* **25**, 714-724, doi:10.1101/gr.188060.114 (2015).
69 Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-94, doi:10.1038/nature09807 (2011).
70 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
71 Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923-930, doi:10.1093/bioinformatics/btt656 (2014).
72 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
73 Chen, E. Y. *et al.* Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics* **14**, 128, doi:10.1186/1471-2105-14-128 (2013).
74 Yu, G., Wang, L. G. & He, Q. Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382-2383, doi:10.1093/bioinformatics/btv145 (2015).
75 Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol* **109**, 21 29 21-21 29 29, doi:10.1002/0471142727.mb2129s109 (2015).
76 Livshits, G. *et al.* Arid1a restrains Kras-dependent changes in acinar cell identity. *Elife* **7**, doi:10.7554/eLife.35216 (2018).
77 Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10-12 (2011).
78 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-359, doi:10.1038/nmeth.1923 (2012).
79 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
80 Pronier, E. *et al.* Targeting the CALR interactome in myeloproliferative neoplasms. *JCI Insight* **3**, doi:10.1172/jci.insight.122703 (2018).
81 Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-842, doi:10.1093/bioinformatics/btq033 (2010).
82 Al Bakir, M. *et al.* The evolution of non-small cell lung cancer metastases in TRACERx. *Nature* **616**, 534-542, doi:10.1038/s41586-023-05729-x (2023).
83 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).

84 Zhang, Z. *et al.* Loss of CHD1 Promotes Heterogeneous Mechanisms of Resistance to AR-Targeted Therapy via Chromatin Dysregulation. *Cancer Cell* **37**, 584-598 e511, doi:10.1016/j.ccell.2020.03.001 (2020).

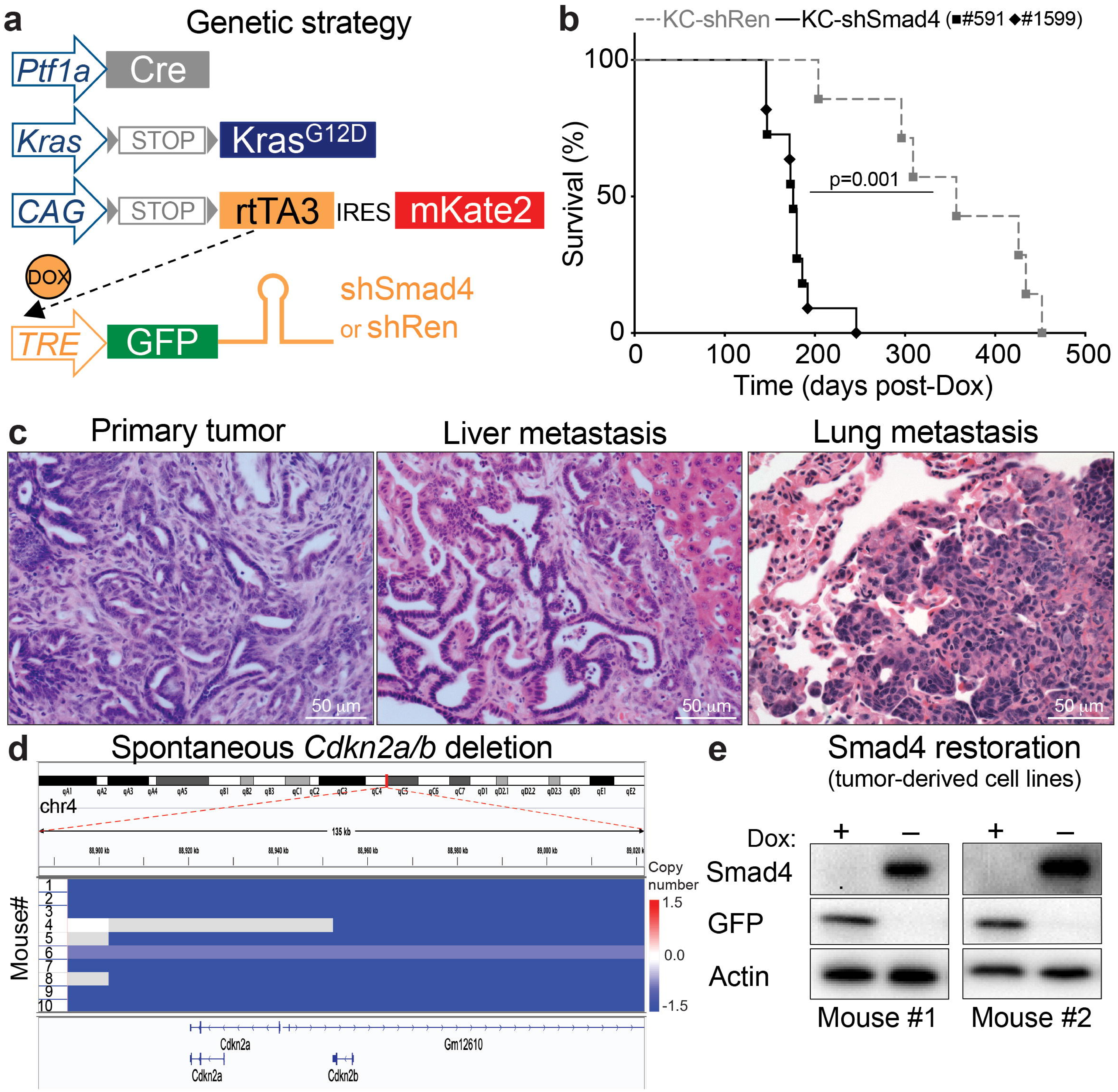
85 Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat Immunol* **20**, 163-172, doi:10.1038/s41590-018-0276-y (2019).

86 Hao, Y. *et al.* Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).

87 Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).

88 Gao, J. *et al.* Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. *Sci Signal* **6**, pl1, doi:10.1126/scisignal.2004088 (2013).

Figure 1


a
Genetic strategy
Ptf1a
Cre
Kras
STOP
Kras^{G12D}
CAG
STOP
rtTA3
IRES
mKate2
DOX
TRE
GFP
shSmad4
or shRen
b
KC-shRen
KC-shSmad4 (■#591 ◆#1599)
100
50
0
Survival (%)
p=0.001
0
100
200
300
400
500
Time (days post-Dox)
c
Primary tumor
Liver metastasis
Lung metastasis
50 µm
50 µm
50 µm
d
Spontaneous Cdkn2a/b deletion
qA1
qA2
qA3
qA4
qA5
qB1
qB2
qB3
qC1
qC2
qC3
qC4
qC5
qC6
qC7
qD1
qD2.1
qD2.2
qD2.3
qD3
qE1
qE2
chr4
135 kb
88,900 kb
88,920 kb
88,940 kb
88,960 kb
88,980 kb
89,000 kb
89,020 k
Mouse#
1
2
3
4
5
6
7
8
9
10
Copy number
1.5
0.0
-1.5
Cdkn2a
Gm12610
Cdkn2a
Cdkn2b
e
Smad4 restoration
(tumor-derived cell lines)
Dox:
+
–
+
–
Smad4
GFP
Actin
Mouse #1
Mouse #2

ED Figure 1

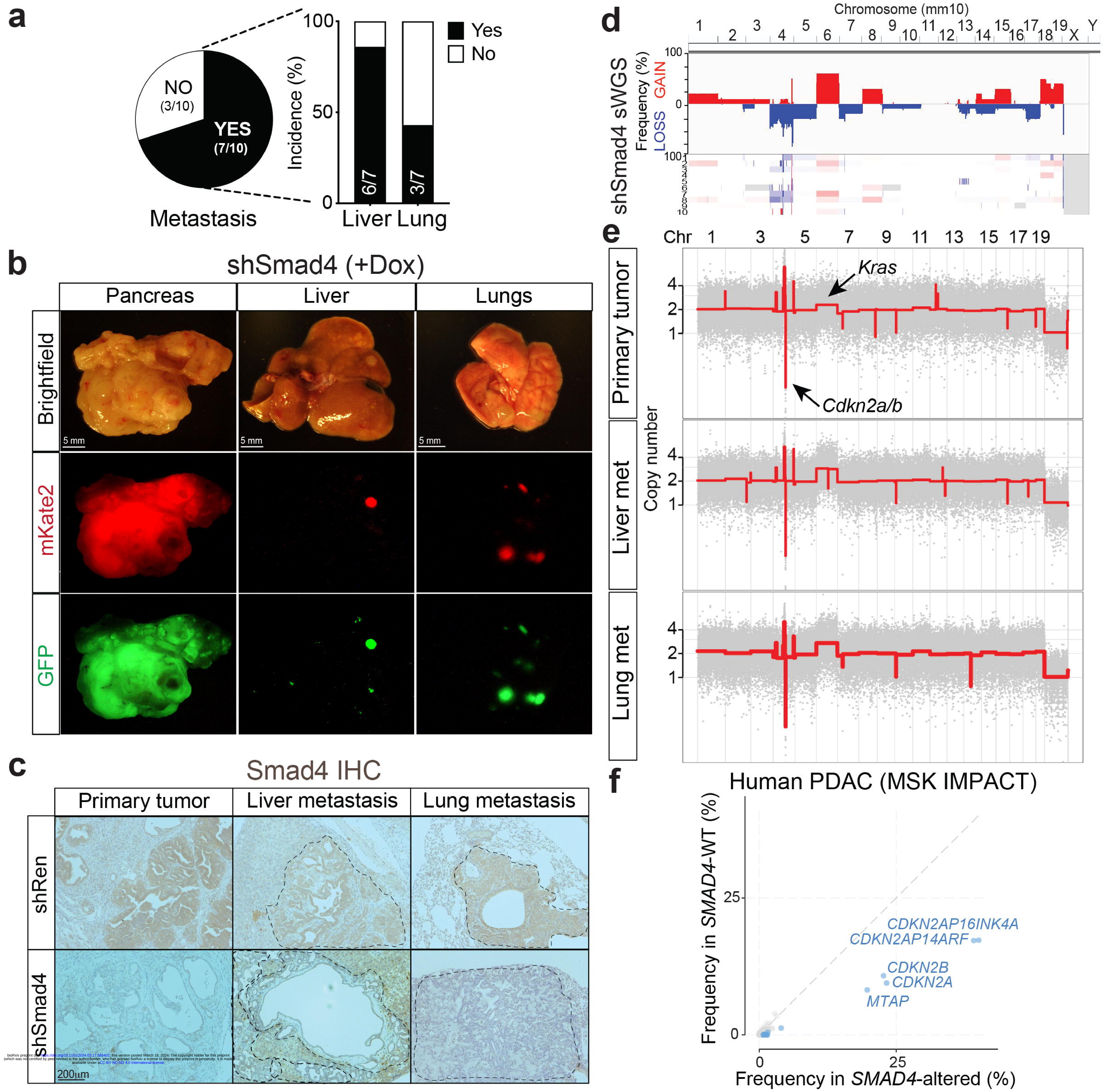
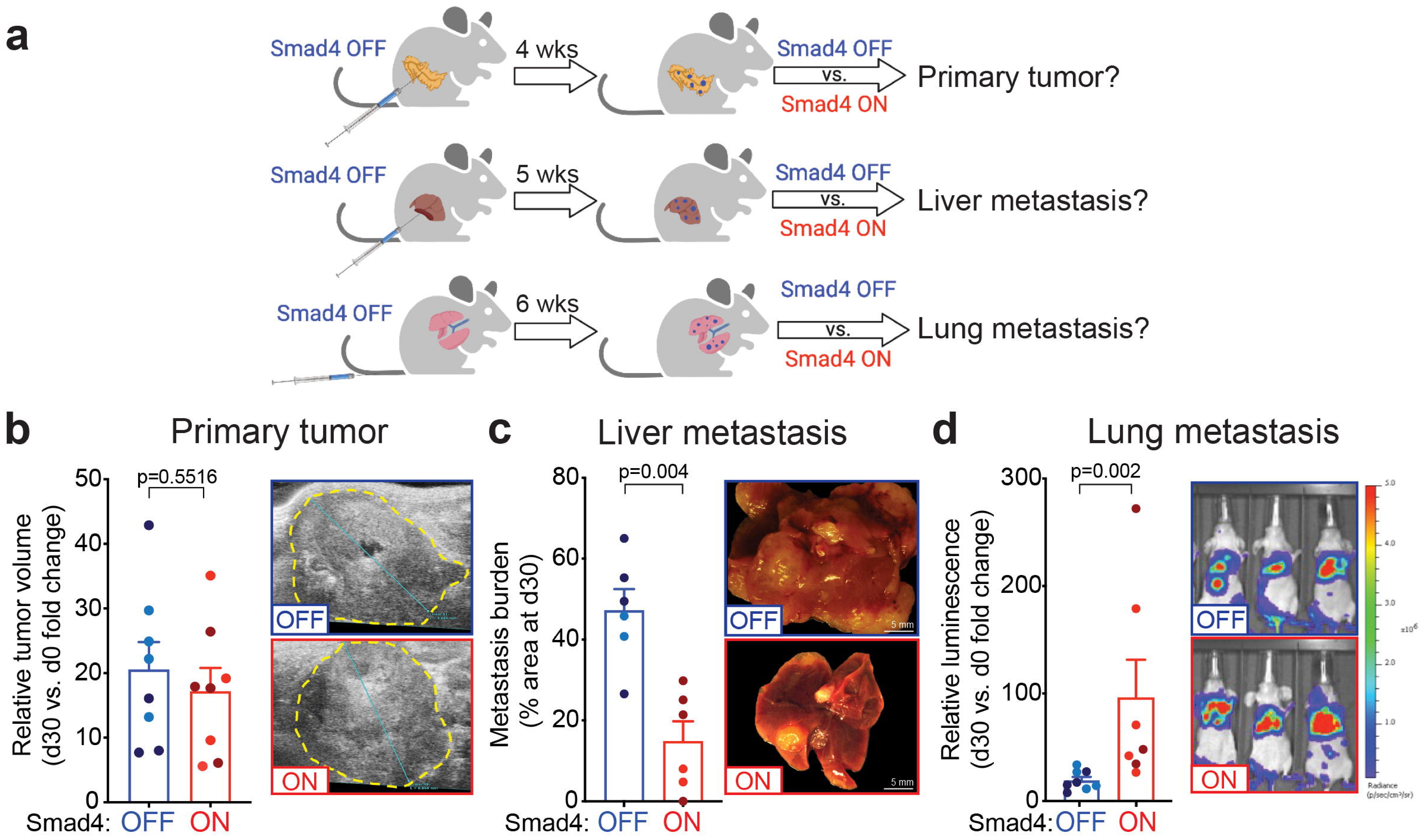
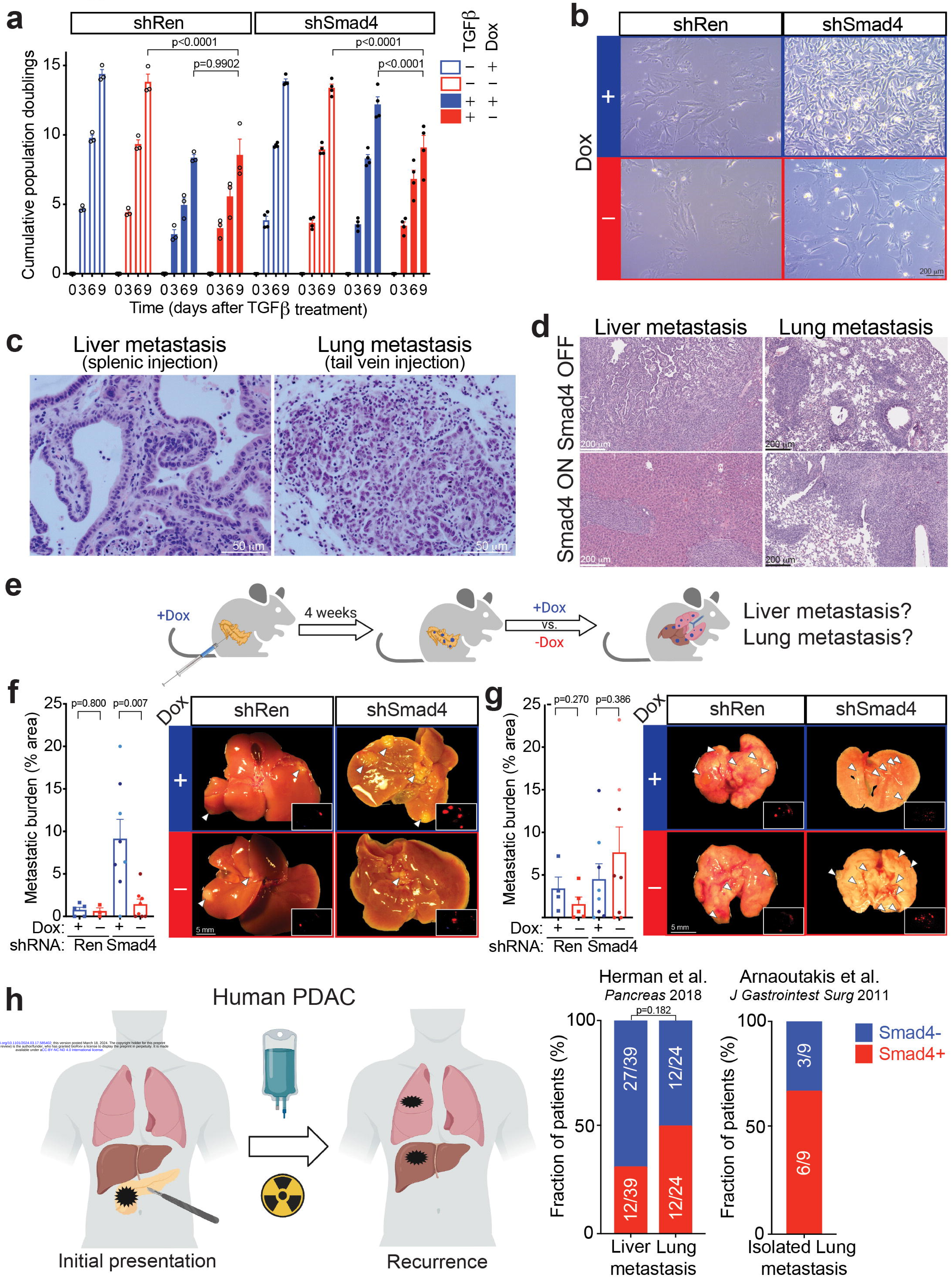

a
NO
(3/10)
YES
(7/10)
Metastasis
Incidence (%)
100
50
0
6/7
3/7
Liver
Lung
Yes
No
b
shSmad4 (+Dox)
Pancreas
Liver
Lungs
Brightfield
mKate2
GFP
5 mm
5 mm
5 mm
c
Smad4 IHC
Primary tumor
Liver metastasis
Lung metastasis
shRen
shSmad4
200µm
d
Chromosome (mm10)
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 X Y
shSmad4 sWGS
Frequency (%)
GAIN
LOSS
100
0
100
1 2 3 4 5 6 7 8 9 10
e
Chr 1 3 5 7 9 11 13 15 17 19
Primary tumor
Liver met
Lung met
Copy number
4
2
1
Kras
Cdkn2a/b
f
Human PDAC (MSK IMPACT)
Frequency in SMAD4-WT (%)
Frequency in SMAD4-altered (%)
0
25
CDKN2AP16INK4A
CDKN2AP14ARF
CDKN2B
CDKN2A
MTAP

Figure 2

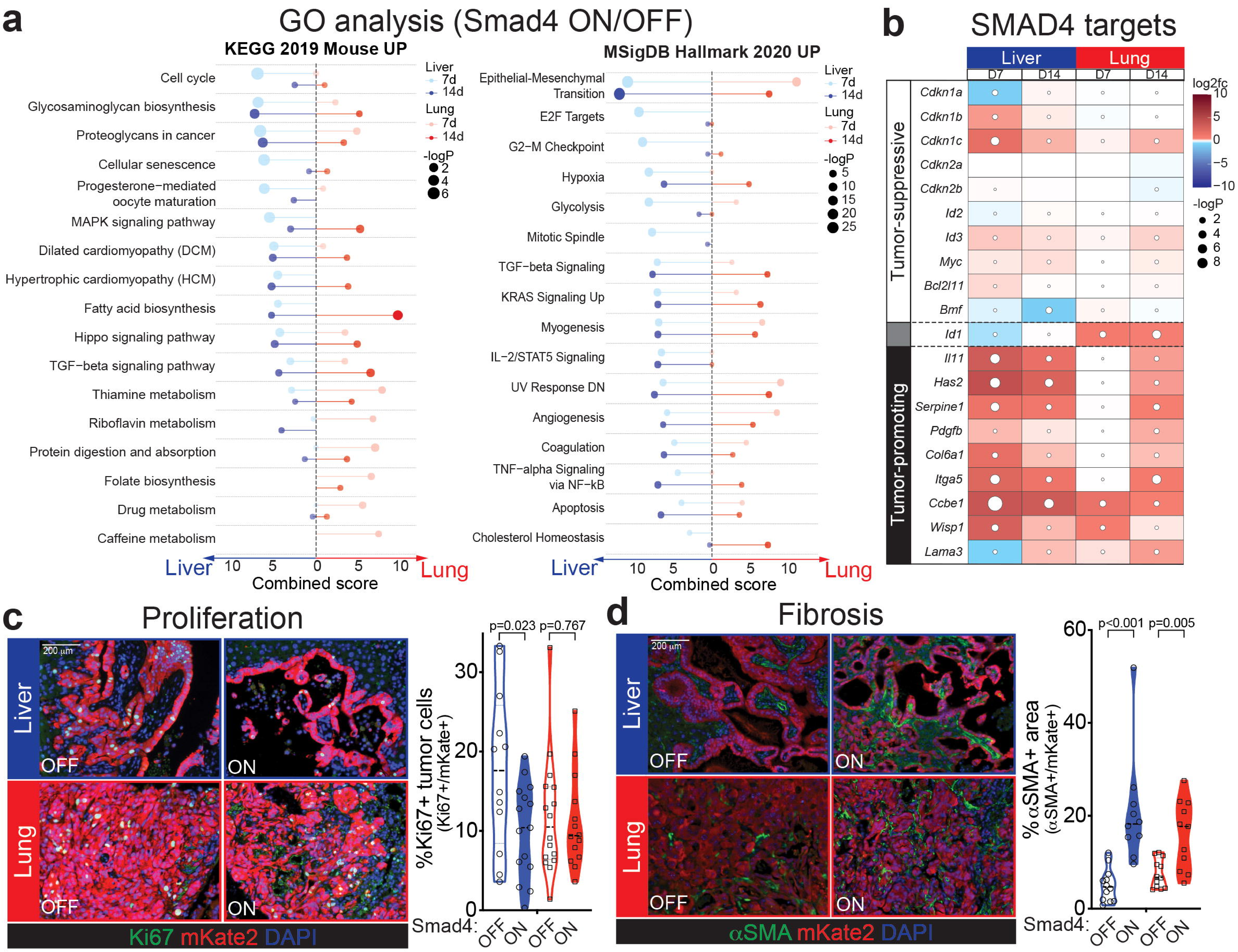

a
Smad4 OFF
4 wks
Smad4 OFF
vs.
Smad4 ON
Primary tumor?
Smad4 OFF
5 wks
Smad4 OFF
vs.
Smad4 ON
Liver metastasis?
Smad4 OFF
6 wks
Smad4 OFF
vs.
Smad4 ON
Lung metastasis?
b
Primary tumor
Relative tumor volume
(d30 vs. d0 fold change)
0
10
20
30
40
50
p=0.5516
Smad4: OFF ON
OFF
ON
c
Liver metastasis
Metastasis burden
(% area at d30)
0
20
40
60
80
p=0.004
Smad4: OFF ON
OFF
ON
5 mm
5 mm
d
Lung metastasis
Relative luminescence
(d30 vs. d0 fold change)
0
100
200
300
p=0.002
Smad4: OFF ON
OFF
ON
5.0
4.0
3.0
2.0
1.0
x10^6
Radiance
(p/sec/cm²/sr)

ED Figure 2



a
shRen
shSmad4
TGFβ
Dox
− +
− −
+ +
+ −
p<0.0001
p=0.9902
p<0.0001
p<0.0001
Cumulative population doublings
15
10
5
0
0369 0369 0369 0369 0369 0369 0369 0369
Time (days after TGFβ treatment)
b
shRen
shSmad4
Dox
+
−
200 µm
c
Liver metastasis
(splenic injection)
Lung metastasis
(tail vein injection)
50 µm
50 µm
d
Liver metastasis
Lung metastasis
Smad4 ON Smad4 OFF
200 µm
200 µm
200 µm
200 µm
e
+Dox
4 weeks
+Dox
vs.
-Dox
Liver metastasis?
Lung metastasis?
f
p=0.800
p=0.007
Metastatic burden (% area)
25
20
15
10
5
0
Dox: + − + −
shRNA: Ren Smad4
Dox
shRen
shSmad4
+
−
5 mm
g
p=0.270
p=0.386
Metastatic burden (% area)
25
20
15
10
5
0
Dox: + − + −
shRNA: Ren Smad4
Dox
shRen
shSmad4
+
−
5 mm
h
Human PDAC
Initial presentation
Recurrence
Herman et al.
Pancreas 2018
p=0.182
Fraction of patients (%)
100
50
0
27/39
12/39
12/24
12/24
Liver
Lung
metastasis
Arnaoutakis et al.
J Gastrointest Surg 2011
Fraction of patients (%)
100
50
0
3/9
6/9
Isolated Lung
metastasis
Smad4-
Smad4+

Figure 3


a
GO analysis (Smad4 ON/OFF)
KEGG 2019 Mouse UP
Cell cycle
Glycosaminoglycan biosynthesis
Proteoglycans in cancer
Cellular senescence
Progesterone−mediated oocyte maturation
MAPK signaling pathway
Dilated cardiomyopathy (DCM)
Hypertrophic cardiomyopathy (HCM)
Fatty acid biosynthesis
Hippo signaling pathway
TGF−beta signaling pathway
Thiamine metabolism
Riboflavin metabolism
Protein digestion and absorption
Folate biosynthesis
Drug metabolism
Caffeine metabolism
Liver
7d
14d
Lung
7d
14d
-logP
2
4
6
Liver 10 5 0 5 10 Lung
Combined score
MSigDB Hallmark 2020 UP
Epithelial-Mesenchymal Transition
E2F Targets
G2−M Checkpoint
Hypoxia
Glycolysis
Mitotic Spindle
TGF-beta Signaling
KRAS Signaling Up
Myogenesis
IL−2/STAT5 Signaling
UV Response DN
Angiogenesis
Coagulation
TNF−alpha Signaling via NF−kB
Apoptosis
Cholesterol Homeostasis
Liver
7d
14d
Lung
7d
14d
-logP
5
10
15
20
25
Liver 10 5 0 5 10 Lung
Combined score
b
SMAD4 targets
Liver
Lung
D7
D14
D7
D14
Tumor-suppressive
Cdkn1a
Cdkn1b
Cdkn1c
Cdkn2a
Cdkn2b
Id2
Id3
Myc
Bcl2l11
Bmf
Id1
Tumor-promoting
Il11
Has2
Serpine1
Pdgfb
Col6a1
Itga5
Ccbe1
Wisp1
Lama3
log2fc
10
5
0
−5
−10
-logP
2
4
6
8
c
Proliferation
200 µm
Liver
Lung
OFF
ON
OFF
ON
Ki67 mKate2 DAPI
p=0.023
p=0.767
%Ki67+ tumor cells (Ki67+/mKate+)
0
10
20
30
Smad4: OFF ON OFF ON
d
Fibrosis
200 µm
Liver
Lung
OFF
ON
OFF
ON
αSMA mKate2 DAPI
p<0.001
p=0.005
%αSMA+ area (αSMA+/mKate+)
0
20
40
60
Smad4: OFF ON OFF ON

ED Figure 3

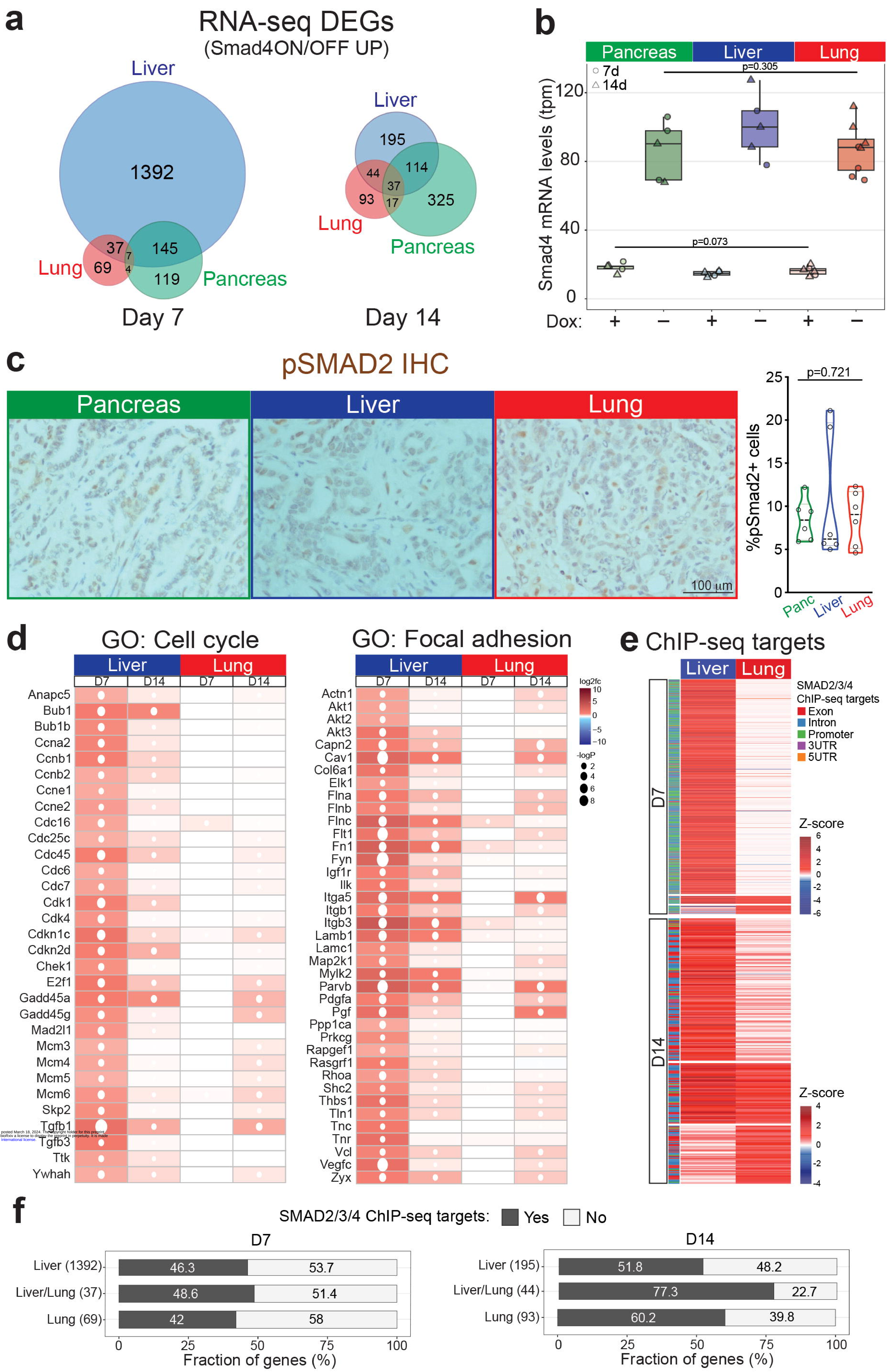
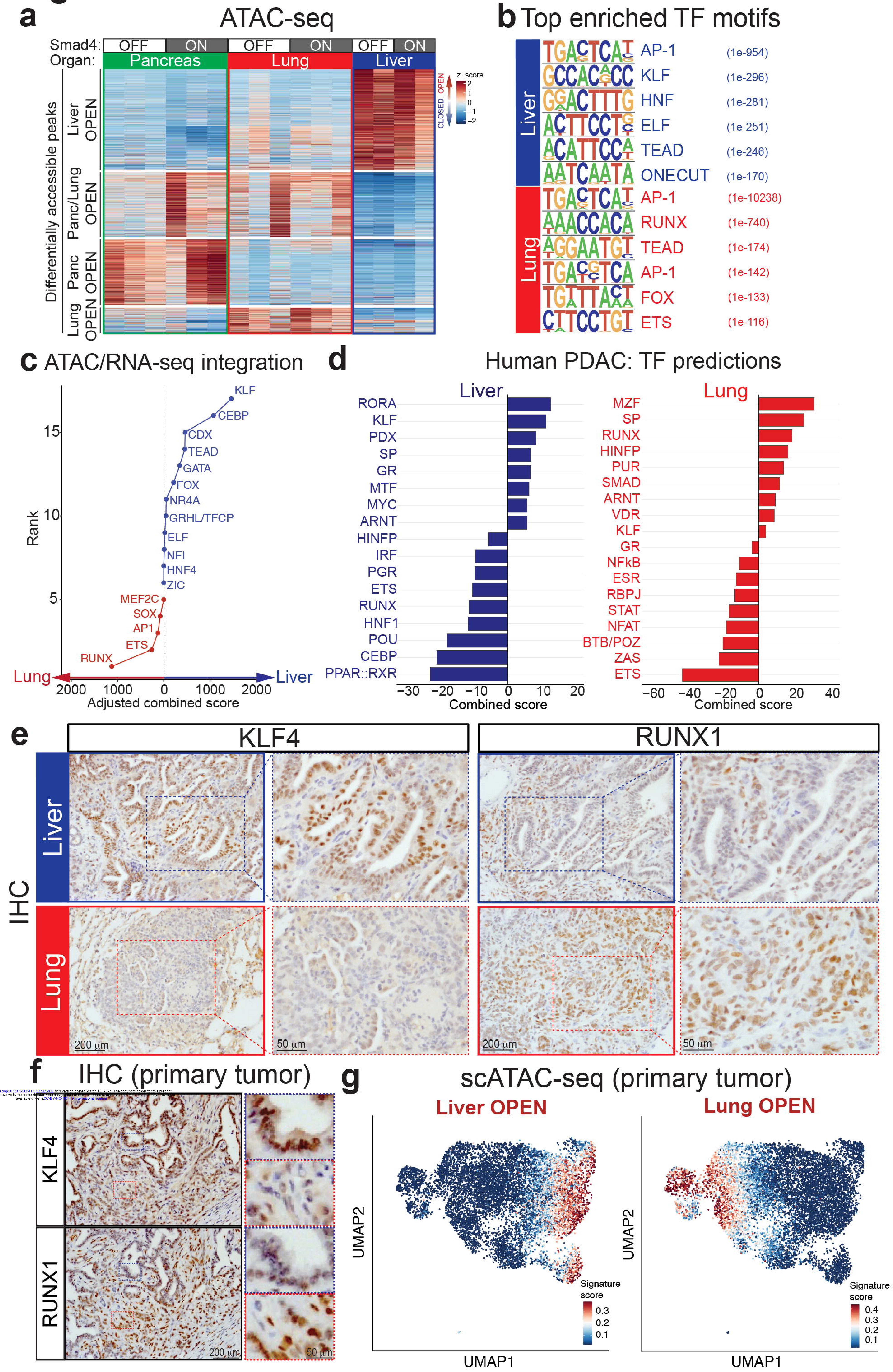
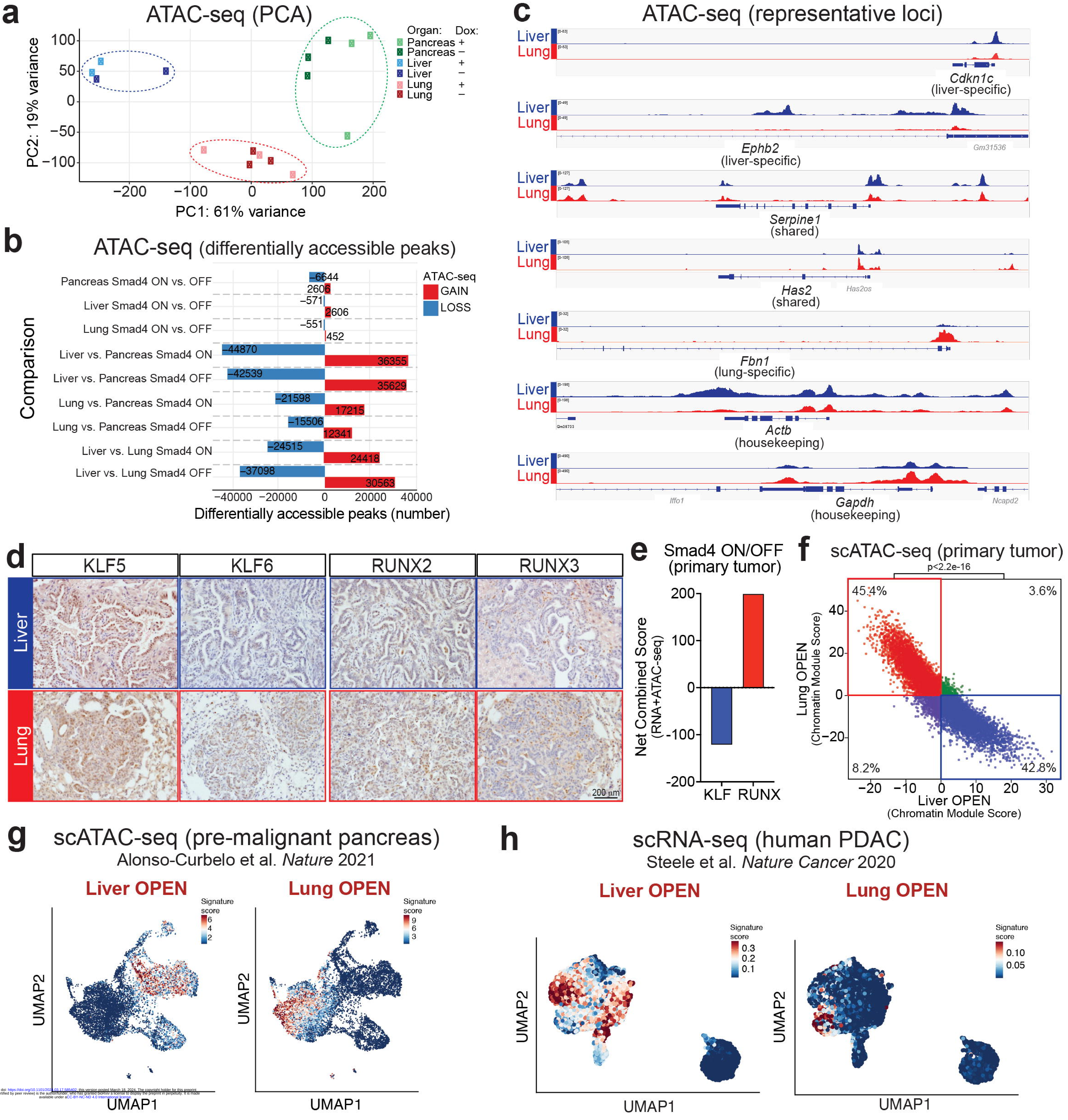

a
RNA-seq DEGs
(Smad4ON/OFF UP)
Liver
1392
37
7
4
145
Lung
69
119
Pancreas
Day 7
Liver
195
44
114
37
93
17
325
Lung
Pancreas
Day 14
b
Pancreas
Liver
Lung
○ 7d
△ 14d
p=0.305
p=0.073
Smad4 mRNA levels (tpm)
120
80
40
0
Dox: + − + − + −
c
pSMAD2 IHC
Pancreas
Liver
Lung
100 μm
p=0.721
%pSmad2+ cells
25
20
15
10
5
0
Panc
Liver
Lung
d
GO: Cell cycle
Liver
Lung
D7
D14
D7
D14
Anapc5
Bub1
Bub1b
Ccna2
Ccnb1
Ccnb2
Ccne1
Ccne2
Cdc16
Cdc25c
Cdc45
Cdc6
Cdc7
Cdk1
Cdk4
Cdkn1c
Cdkn2d
Chek1
E2f1
Gadd45a
Gadd45g
Mad2l1
Mcm3
Mcm4
Mcm5
Mcm6
Skp2
Tgfb1
Tgfb3
Ttk
Ywhah
GO: Focal adhesion
Liver
Lung
D7
D14
D7
D14
Actn1
Akt1
Akt2
Akt3
Capn2
Cav1
Col6a1
Elk1
Flna
Flnb
Flnc
Flt1
Fn1
Fyn
Igf1r
Ilk
Itga5
Itgb1
Itgb3
Lamb1
Lamc1
Map2k1
Mylk2
Parvb
Pdgfa
Pgf
Ppp1ca
Prkcg
Rapgef1
Rasgrf1
Rhoa
Shc2
Thbs1
Tln1
Tnc
Tnr
Vcl
Vegfc
Zyx
log2fc
10
5
0
−5
−10
-logP
2
4
6
8
e
ChIP-seq targets
Liver
Lung
SMAD2/3/4
ChIP-seq targets
Exon
Intron
Promoter
3UTR
5UTR
D7
D14
Z-score
6
4
2
0
-2
-4
-6
Z-score
4
2
0
-2
-4
f
SMAD2/3/4 ChIP-seq targets:
Yes
No
D7
Liver (1392)
Liver/Lung (37)
Lung (69)
46.3
53.7
48.6
51.4
42
58
0
25
50
75
100
Fraction of genes (%)
D14
Liver (195)
Liver/Lung (44)
Lung (93)
51.8
48.2
77.3
22.7
60.2
39.8
0
25
50
75
100
Fraction of genes (%)

Figure 4

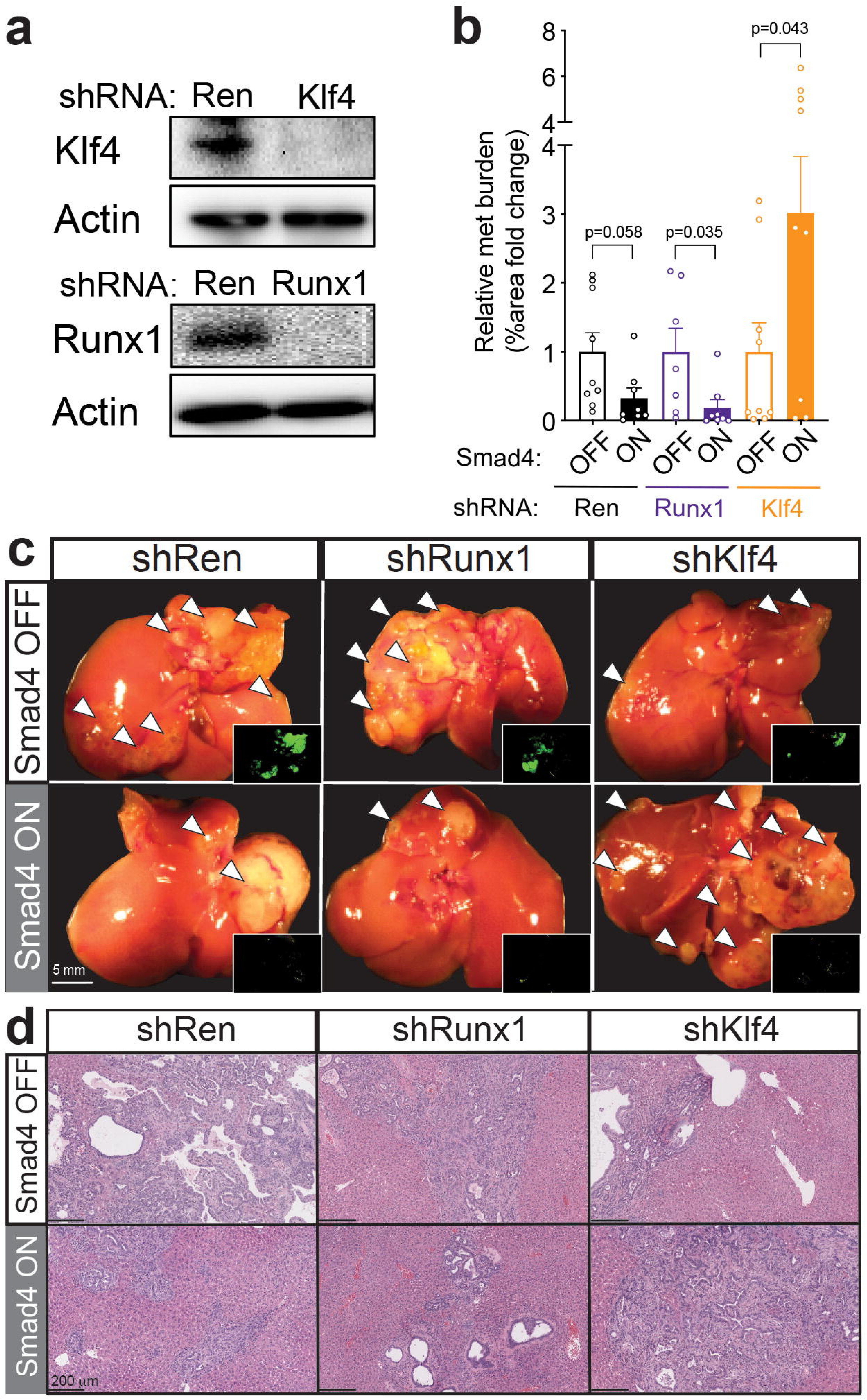

a
ATAC-seq
Smad4:
OFF
ON
OFF
ON
OFF
ON
Organ:
Pancreas
Lung
Liver
Differentially accessible peaks
Liver OPEN
Panc/Lung OPEN
Panc OPEN
Lung OPEN
z-score
2
1
0
−1
−2
OPEN
CLOSED
b
Top enriched TF motifs
Liver
AP-1 (1e-954)
KLF (1e-296)
HNF (1e-281)
ELF (1e-251)
TEAD (1e-246)
ONECUT (1e-170)
Lung
AP-1 (1e-10238)
RUNX (1e-740)
TEAD (1e-174)
AP-1 (1e-142)
FOX (1e-133)
ETS (1e-116)
c
ATAC/RNA-seq integration
Rank
15
10
5
KLF
CEBP
CDX
TEAD
GATA
FOX
NR4A
GRHL/TFCP
ELF
NFI
HNF4
ZIC
MEF2C
SOX
AP1
ETS
RUNX
Lung
Liver
2000
1000
0
1000
2000
Adjusted combined score
d
Human PDAC: TF predictions
Liver
RORA
KLF
PDX
SP
GR
MTF
MYC
ARNT
HINFP
IRF
PGR
ETS
RUNX
HNF1
POU
CEBP
PPAR::RXR
−30
−20
−10
0
10
20
Combined score
Lung
MZF
SP
RUNX
HINFP
PUR
SMAD
ARNT
VDR
KLF
GR
NFkB
ESR
RBPJ
STAT
NFAT
BTB/POZ
ZAS
ETS
−60
−40
−20
0
20
40
Combined score
e
KLF4
RUNX1
IHC
Liver
Lung
200 µm
50 µm
200 µm
50 µm
f
IHC (primary tumor)
KLF4
RUNX1
200 µm
50 µm
g
scATAC-seq (primary tumor)
Liver OPEN
Lung OPEN
UMAP2
UMAP1
Signature score
0.3
0.2
0.1
UMAP2
UMAP1
Signature score
0.4
0.3
0.2
0.1

ED Figure 4

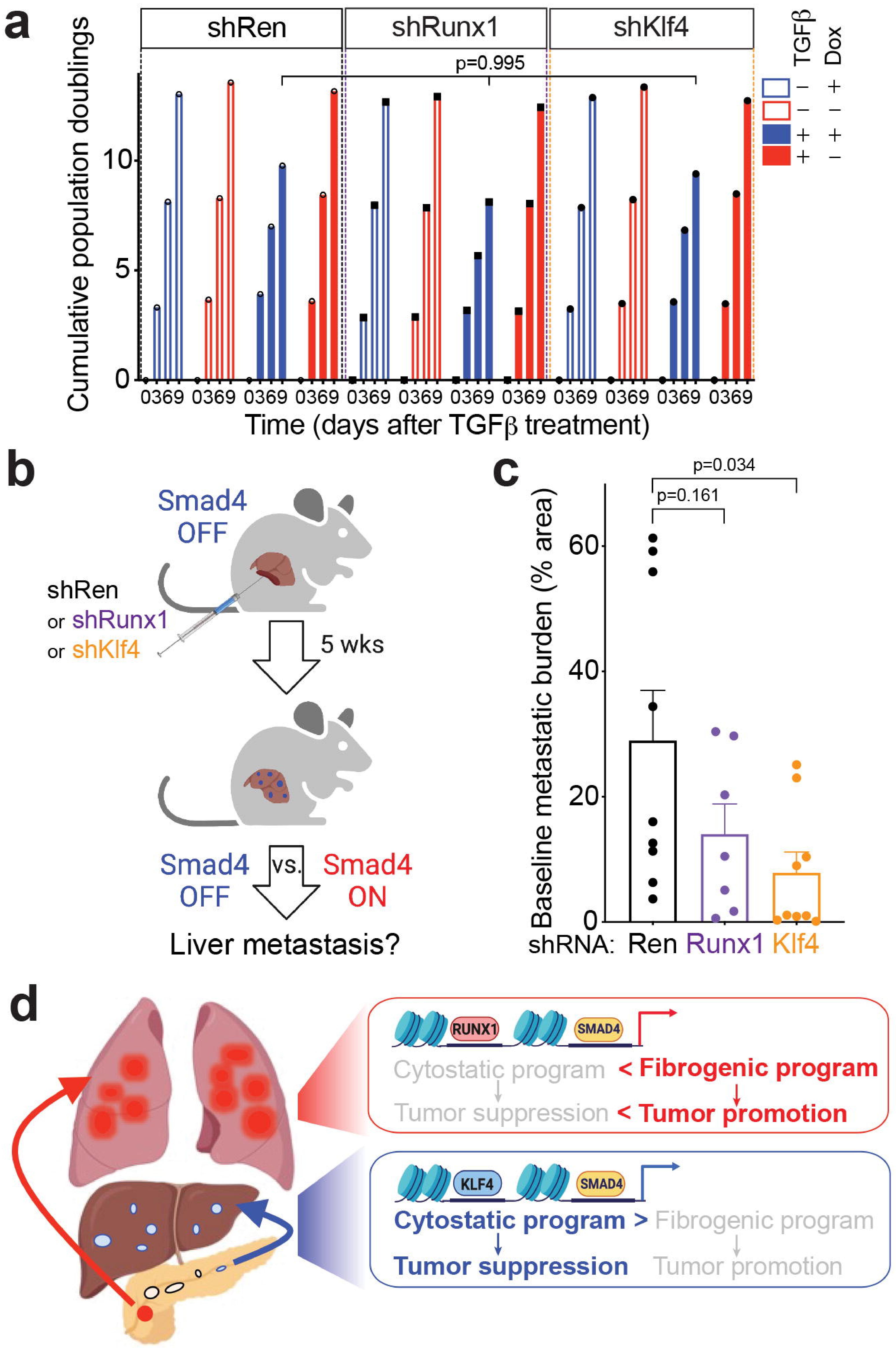


a
ATAC-seq (PCA)
PC2: 19% variance
PC1: 61% variance
100
50
0
−50
−100
−200
−100
0
100
200
Organ:
Dox:
Pancreas +
Pancreas −
Liver +
Liver −
Lung +
Lung −
b
ATAC-seq (differentially accessible peaks)
Comparison
Pancreas Smad4 ON vs. OFF
Liver Smad4 ON vs. OFF
Lung Smad4 ON vs. OFF
Liver vs. Pancreas Smad4 ON
Liver vs. Pancreas Smad4 OFF
Lung vs. Pancreas Smad4 ON
Lung vs. Pancreas Smad4 OFF
Liver vs. Lung Smad4 ON
Liver vs. Lung Smad4 OFF
−6644
2606
−571
2606
−551
452
−44870
36355
−42539
35629
−21598
17215
−15506
12341
−24515
24418
−37098
30563
ATAC-seq
GAIN
LOSS
−40000
−20000
0
20000
40000
Differentially accessible peaks (number)
c
ATAC-seq (representative loci)
Liver
Lung
[0-83]
Cdkn1c
(liver-specific)
[0-49]
Ephb2
(liver-specific)
Gm31536
[0-127]
Serpine1
(shared)
[0-105]
Has2
(shared)
Has2os
[0-32]
Fbn1
(lung-specific)
[0-198]
Gm26723
Actb
(housekeeping)
[0-490]
Iffo1
Gapdh
(housekeeping)
Ncapd2
d
KLF5
KLF6
RUNX2
RUNX3
Liver
Lung
200 µm
e
Smad4 ON/OFF
(primary tumor)
Net Combined Score
(RNA+ATAC-seq)
200
100
0
-100
-200
KLF
RUNX
f
scATAC-seq (primary tumor)
p<2.2e-16
45.4%
3.6%
8.2%
42.8%
Lung OPEN
(Chromatin Module Score)
40
20
0
−20
−20
−10
0
10
20
30
Liver OPEN
(Chromatin Module Score)
g
scATAC-seq (pre-malignant pancreas)
Alonso-Curbelo et al. Nature 2021
Liver OPEN
Lung OPEN
Signature score
6
4
2
9
6
3
UMAP2
UMAP1
h
scRNA-seq (human PDAC)
Steele et al. Nature Cancer 2020
Liver OPEN
Lung OPEN
Signature score
0.3
0.2
0.1
0.10
0.05
UMAP2
UMAP1

Figure 5


a
shRNA: Ren Klf4
Klf4
Actin
shRNA: Ren Runx1
Runx1
Actin
b
Relative met burden
(%area fold change)
8
6
4
4
3
2
1
0
p=0.043
p=0.058
p=0.035
Smad4: OFF ON OFF ON OFF ON
shRNA: Ren Runx1 Klf4
c
shRen
shRunx1
shKlf4
Smad4 OFF
Smad4 ON
5 mm
d
shRen
shRunx1
shKlf4
Smad4 OFF
Smad4 ON
200 µm

ED Figure 5


a
shRen
shRunx1
shKlf4
p=0.995
TGFβ
Dox
– +
– –
+ +
+ –
Cumulative population doublings
0
5
10
0369 0369 0369 0369 0369 0369 0369 0369 0369 0369 0369 0369
Time (days after TGFβ treatment)
b
Smad4 OFF
shRen
or shRunx1
or shKlf4
5 wks
Smad4 OFF
vs.
Smad4 ON
Liver metastasis?
c
p=0.034
p=0.161
Baseline metastatic burden (% area)
0
20
40
60
shRNA: Ren Runx1 Klf4
d
RUNX1
SMAD4
Cytostatic program < Fibrogenic program
Tumor suppression < Tumor promotion
KLF4
SMAD4
Cytostatic program > Fibrogenic program
Tumor suppression Tumor promotion