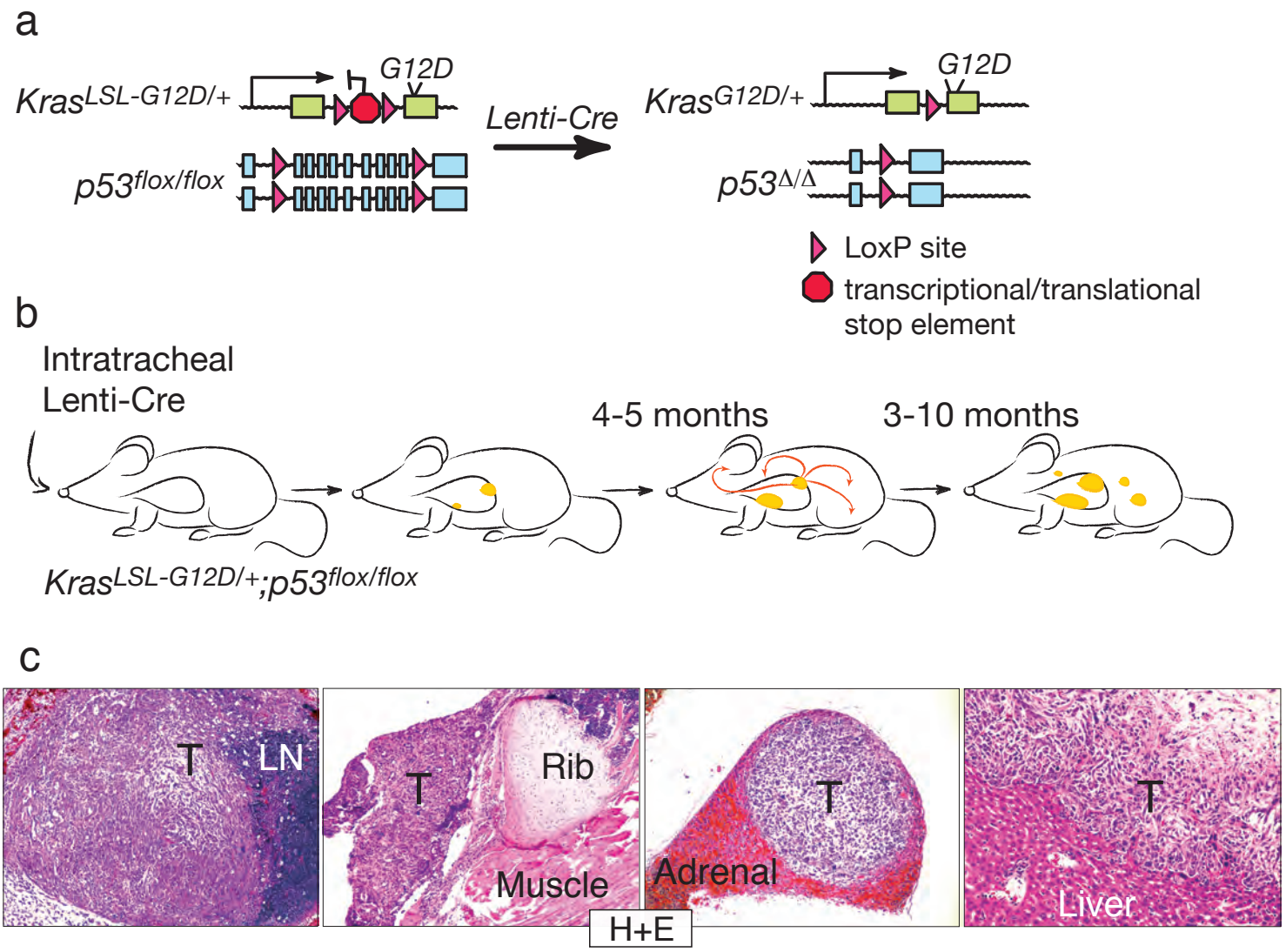
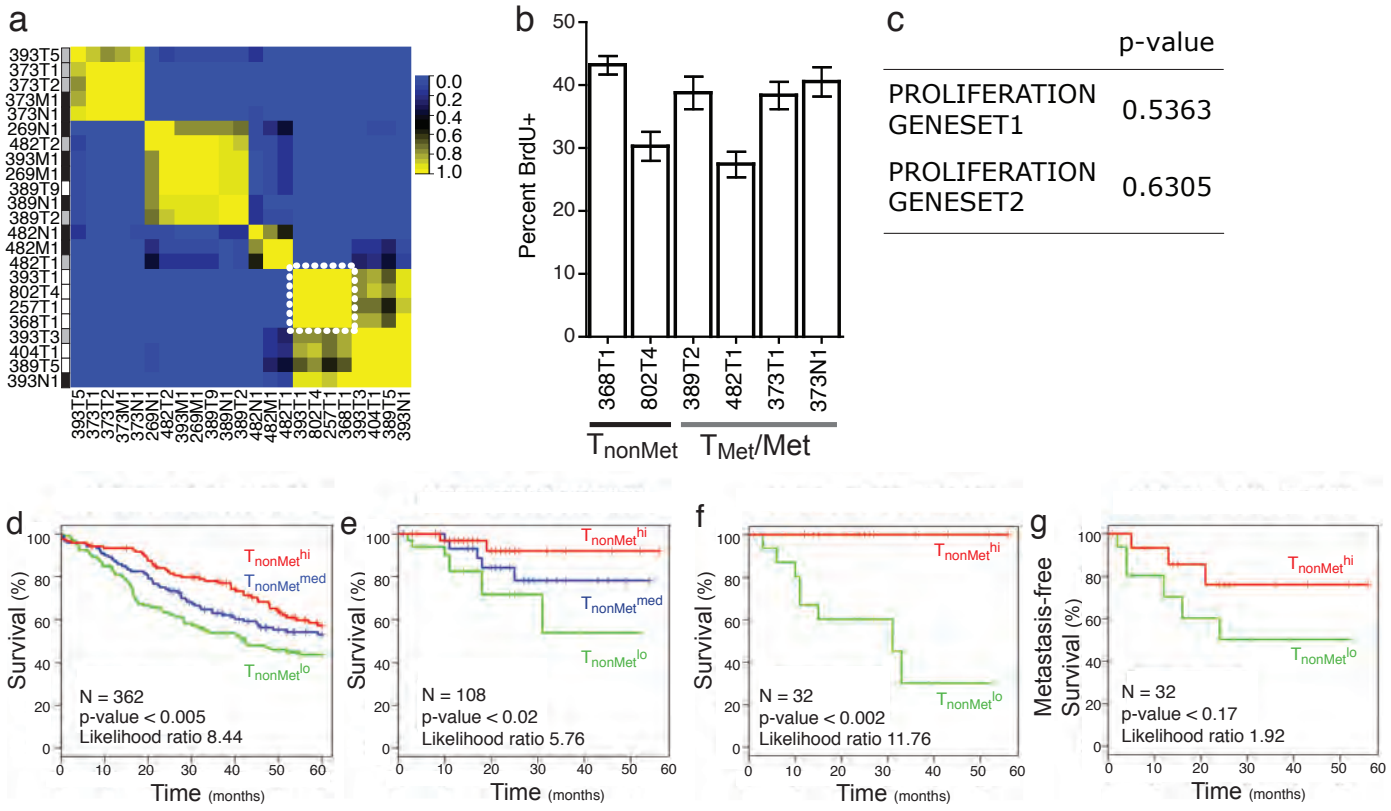


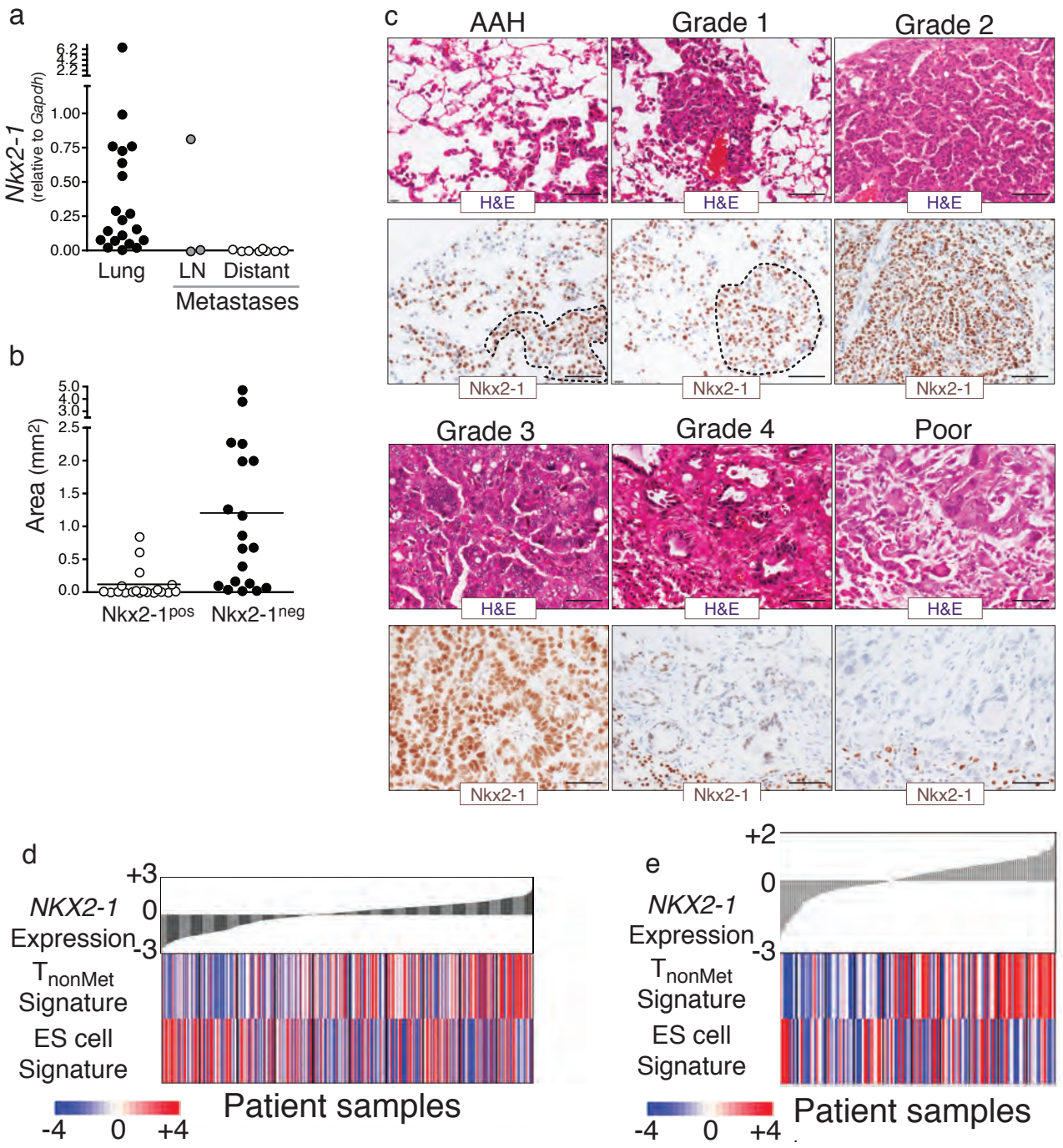
Supplementary Figure 1. Mouse model of metastatic lung adenocarcinoma based on the expression of oncogenic Kras^{G12D} and deletion of p53

a, Human lung adenocarcinomas contain a diverse set of oncogene and tumor suppressor mutations (in many cases as a consequence of tobacco use) with oncogenic KRAS mutations occuring in approximately 30% and p53 mutations occuring in approximately 50% of tumours. Expression of Cre-recombinase in *Kras$^{LSL\text{-}G12D/+}$;p53$^{flox/flox}$* cells results in the deletion of the floxed stop cassette resulting in expression of oncogenic G12D Kras from its endogenous promoter and the deletion of both floxed alleles of the p53 tumour suppressor (Jackson *et al*, 2001; Jackson *et al*, 2005; Jonkers *et al*, 2001). **b,** Intratracheal infection of *Kras$^{LSL\text{-}G12D/+}$;p53$^{flox/flox}$* mice with lentiviral vectors that express Cre-recombinase results in the development of metastatic lung adenocarcinoma. **c,** Representative lymph node (LN), pleural, adrenal gland, and liver metastases (H&E staining). T indicates tumour.



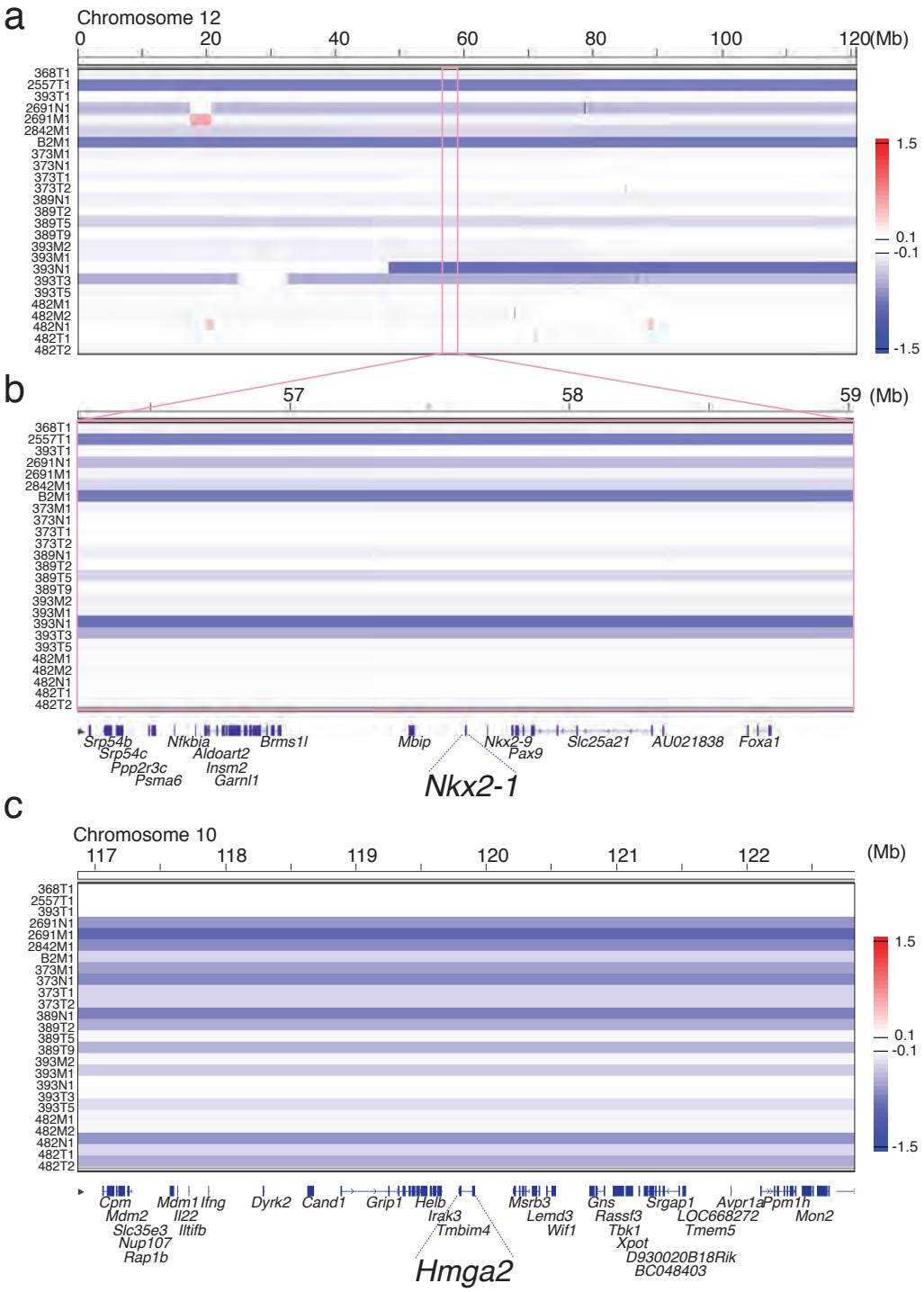
Supplementary Figure 2. A gene expression signature derived from murine lung adenocarcinoma can predict patient outcome and relapse

a, Consensus clustering of genome-wide mRNA expression across 23 samples. Samples are named as mouse number, T for tumour, N for lymph node metastasis, M for distant metastasis, then the lesion number. Bars: White=T_{un}, Grey=T_{Met}, Black=metastases. The white dashed box highlights the four samples that represent a distinct subset of lung tumour-derived cell lines. The colour scale indicates the Pearson's correlation coefficient. **b,** T_{nonMet}, T_{Met}, and Met cell lines do not have consistently different proliferation rates in vitro. Percent of BrdU-positive cells for each cell line is indicated. Mean +/- SD of triplicate wells. **c,** Comparison of T_{nonMet} to T_{Met}/Met cell lines does not show differential expression of two established proliferation gene sets. Student's t-test. Two-samples (4 T_{nonMet} versus 19 T_{Met}/Met), unpaired t-test. **d,** Survival of lung adenocarcinoma patients (Shedden *et al*, 2008). Samples are split into three equally sized tertiles representing the samples with high, medium, and low concordance with the T_{nonMet} signature. **e,** Survival of lung adenocarcinoma patients (Nyugen *et al*, 2009). Samples are split into three equally sized tertiles representing the samples with high, medium, and low concordance with the T_{nonMet} signature. To determine how well this T_{nonMet} signature predicted outcome relative to other genesets, we generated 1000 random gene sets of 250 genes each (the size of the T_{nonMet} signature) and calculated the likelihood ratio for each geneset. The T_{nonMet} signature has a higher likelihood ratio than ~90% of other genesets when considering the Shedden *et al.* dataset and ~94% when considering the Nyugen *et al.* dataset. Importantly only 15/1000 random genesets predict outcome in both datasets better than the T_{nonMet} signature. **f,** Survival of patients with KRAS-mutant lung adenocarcinoma (Nyugen *et al*, 2009) is also predicted by the T_{nonMet} signature. Samples are split into two equally sized groups representing the samples with high and low concordance with the T_{nonMet} signature.
g, Metastasis-free survival of patients with KRAS-mutant lung adenocarcinoma (Nyugen *et al*, 2009). Samples are split into two equally sized groups representing the samples with high and low concordance with the T_{nonMet} signature.

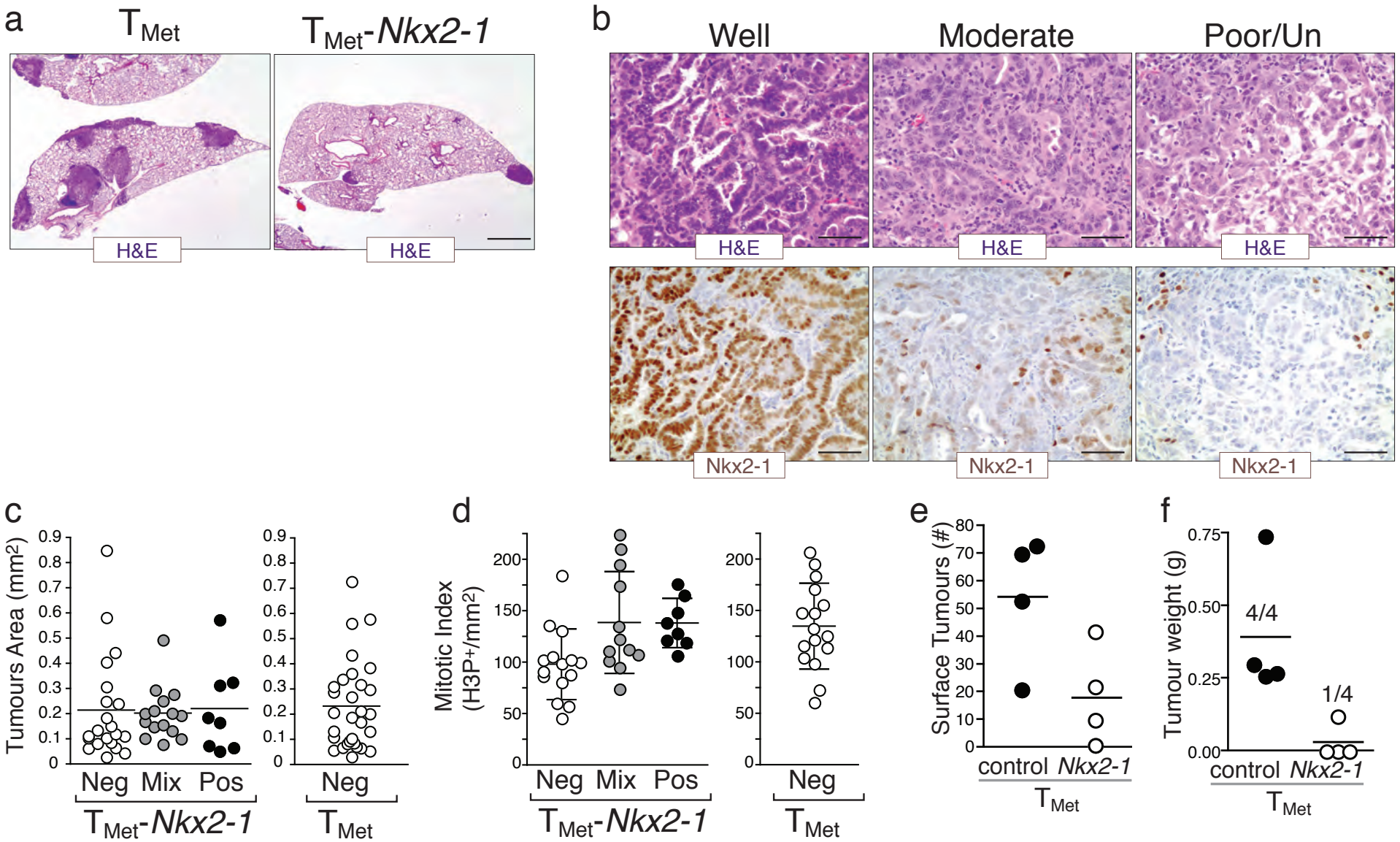


Supplementary Figure 3. Nkx2-1 expression correlates with differentiation state

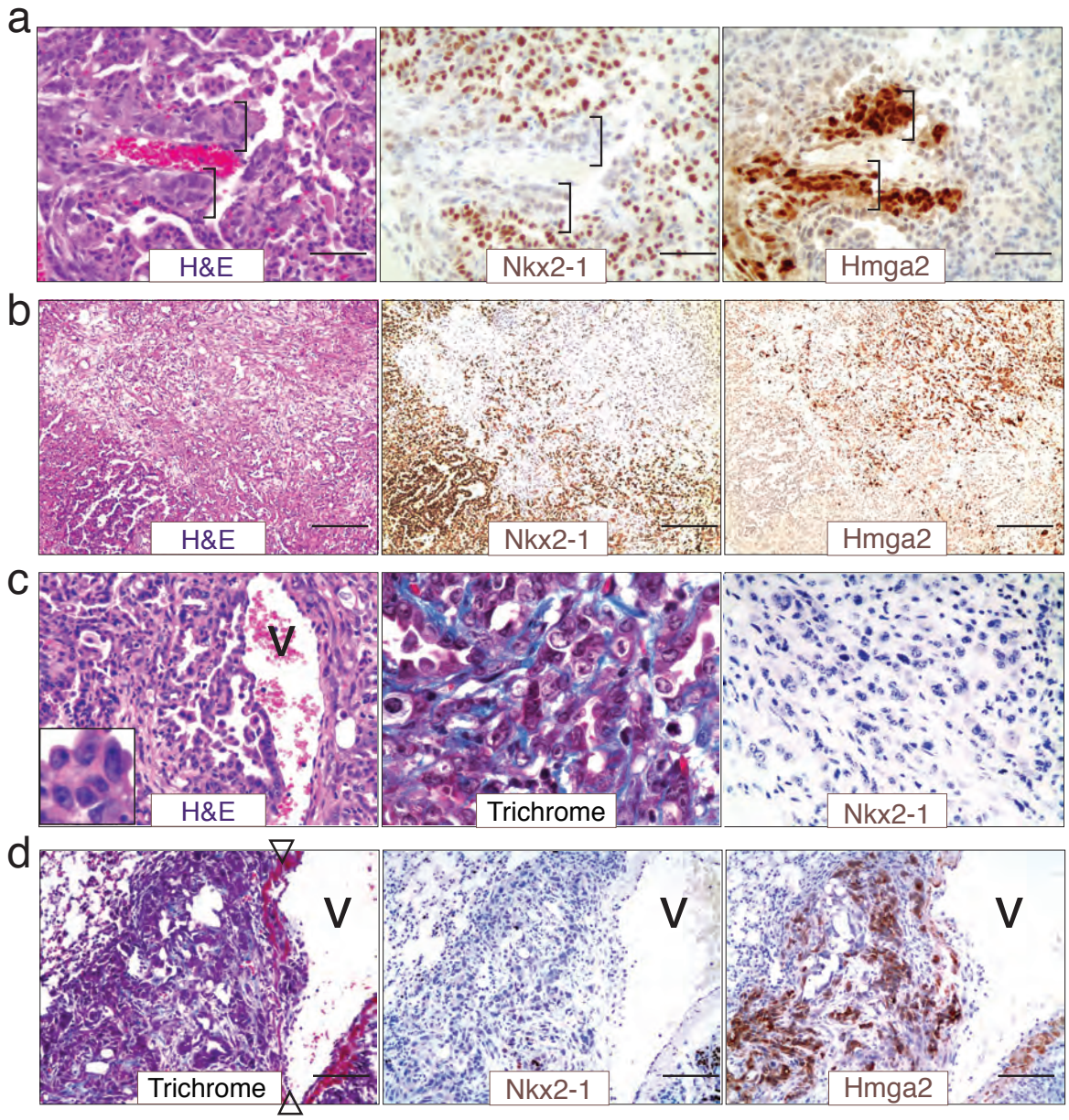
a, *Nkx2-1* gene expression in microdissected lung tumours, lymph node (LN) metastases, and distant metastases. **b,** Metastasis size and Nkx2-1 staining. (N=19/group, p-value <0.003). **c,** Nkx2-1 downregulation correlated with progression to a less differentiated state. Representative images of the progression of *Kras*$^{G12D/+}$;*p53*$^{\Delta/\Delta}$ lung tumours from early well differentiated Nkx2-1^{pos} atypical adenomous hyperplasias (AAH) and adenomas (Grades 1 and 2) to poorly differentiated Nkx2-1^{neg} adenocarcinomas (Grades 4 and 5). Nkx2-1 is diffusely expressed in AAHs and in low grade (1-3) tumours, which typically exhibit a papillary architecture. Grade 4 tumours are found at late time points and are associated with lower grade tumours (note Nkx2-1^{pos} cells at the bottom left). Grade 4 tumours are glandular and often express lower levels of Nkx2-1 in only a subset of tumour cells. Grade 5 tumours exhibit a sheet-like, poorly differentiated architecture and lack detectable Nkx2-1. Hematoxylin and Eosin (top) and anti-Nkx2-1 (bottom) staining is shown. Scale bar = 100μm. Dashed regions denote the AAH and Grade 1 lesions. **d,** Relative *NKX2-1* expression (log_2) in human lung adenocarcinomas (Shedden et al) correlates with a T_{nonMet} signature (r2 = 0.36). The T_{nonMet} signature and an ES cell signature anti-correlate (r2 = -0.44). Each vertical bar represents a lung adenocarcinoma sample. Scale bar indicates the signature score. e. Relative *NKX2-1* expression (log_2) in a second human lung adenocarcinona dataset (Nguyen *et al*, 2009) correlates with a T_{nonMet} signature (r^2 = 0.55). The T_{nonMet} signature and an ES cell signature anti-correlate (r^2 = -0.48). Each vertical bar represents a lung adenocarcinoma sample. Scale bar indicates the signature score

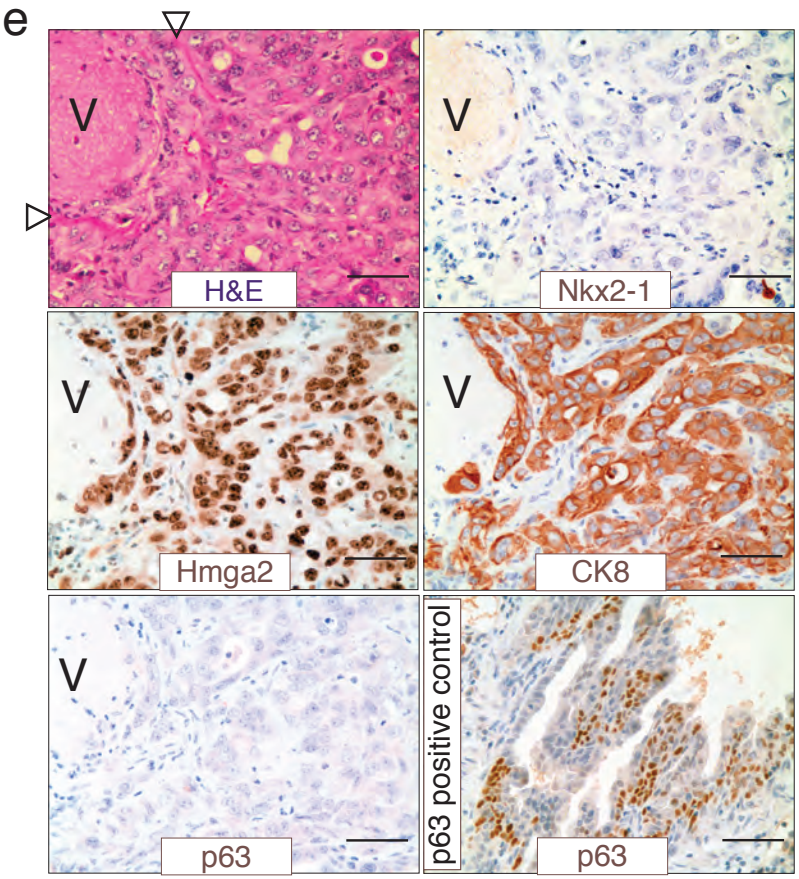


Supplementary Figure 4. No focal DNA copy number alteration around *Nkx2-1* or *Hmga2*
Illumina sequencing-based DNA copy number analysis on 25 lung adenocarcinoma cell lines from the *Kras*$^{LSL\text{-}G12D}$;*p53*$^{flox/flox}$ mouse model. **a-b,** DNA copy number data for Chromosome 12 and the region surrounding *Nkx2-1. Nkx2-1* is neither focally amplified nor lost in any samples. **c,** The *Hmga2* genomic locus does not contain any breakpoints that would indicate a copy number non-neutral translocation nor is *Hmga2* amplified in these cell lines. DNA copy number ratio is relative to a reference somatic DNA sample.



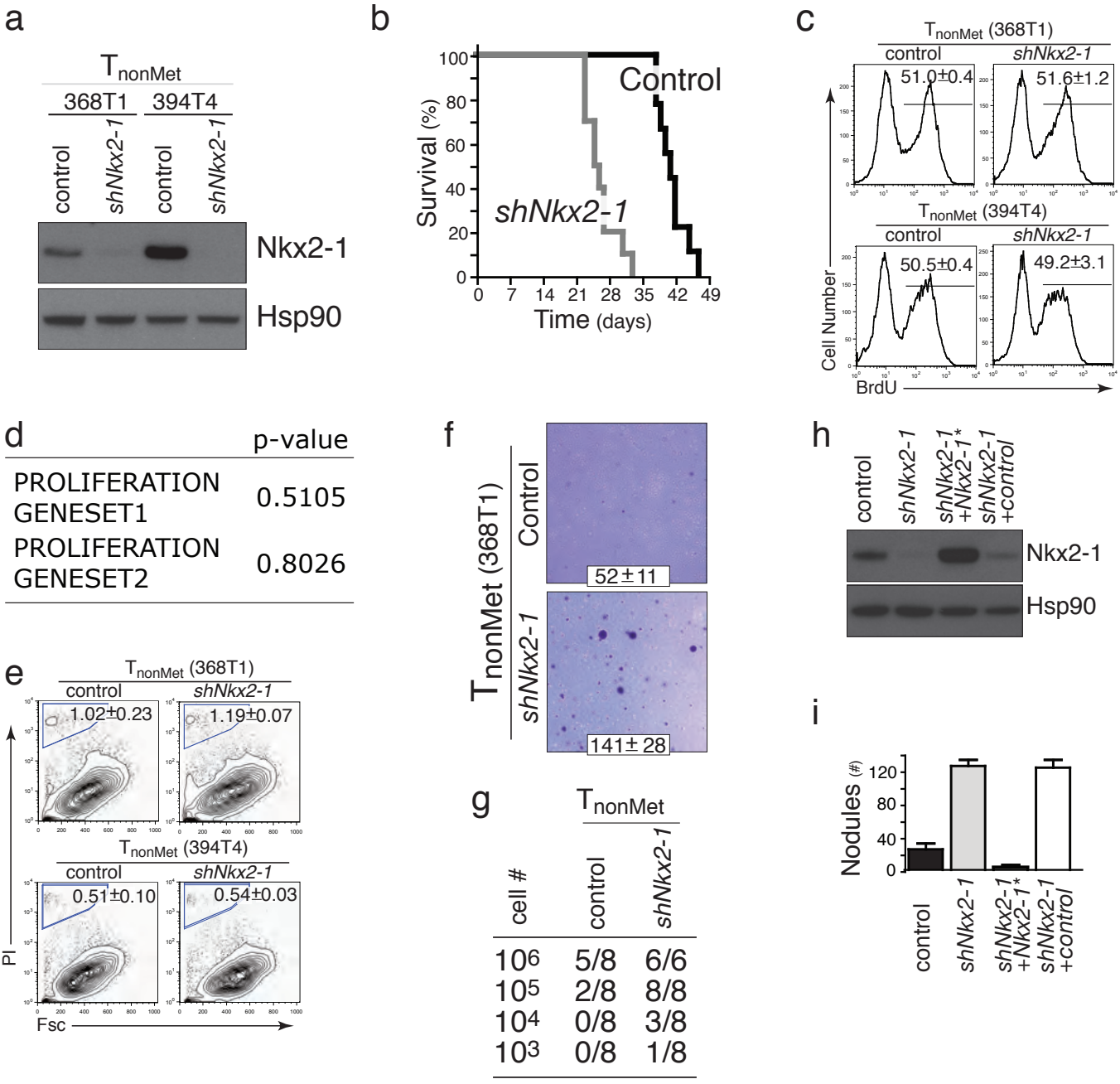
Supplementary Figure 5. Nkx2-1 reexpression inhibits tumour seeding and regulates differentiation. a, Representative H+E stained sections of lungs after intravenous transplantation of T_{Met} cells or T_{Met} cells engineered to reexpress Nkx2-1 (T_{Met}-*Nkx2-1*). Scale bar = 0.5mm. **b,** Representative images of differentiated, moderately differentiated and poor/undifferentated tumours that formed after intravenous transplantation of T_{Met}-*Nkx2-1* cells. Scale bar = 50μm. **c,** Quantification of tumour size after intravenous transplantation of T_{Met} or T_{Met}-*Nkx2-1* cells. Tumours that form after transplantation of T_{Met}-*Nkx2-1* cells were distinguished based on their Nkx2-1 expression. Each circle represents an individual tumour. Bar represents the mean. **d,** Quantification of mitotic index (number of cells with phosphorylated histone 3 per tumour area) after intravenous transplantation of T_{Met} or T_{Met}-*Nkx2-1* cells. Tumours that formed after transplantation of T_{Met}-*Nkx2-1* cells were distinguished based on their Nkx2-1 expression. Each circle represents an individual tumour. Bar represents the mean +/- SD. **e,** Nkx2-1 cDNA reexpression reduces liver nodule formation after intrasplenic transplantation. Each circle represents an individual mouse. Bar represents the mean. Student's t-test $p < 0.06$. **f,** Nkx2-1 cDNA reexpression reduces subcutaneous tumour formation. Each circle represents an individual mouse and the fraction shows the number of mice that developed tumours/number of mice injected. Bar represents the mean. Student's t-test $p < 0.03$.





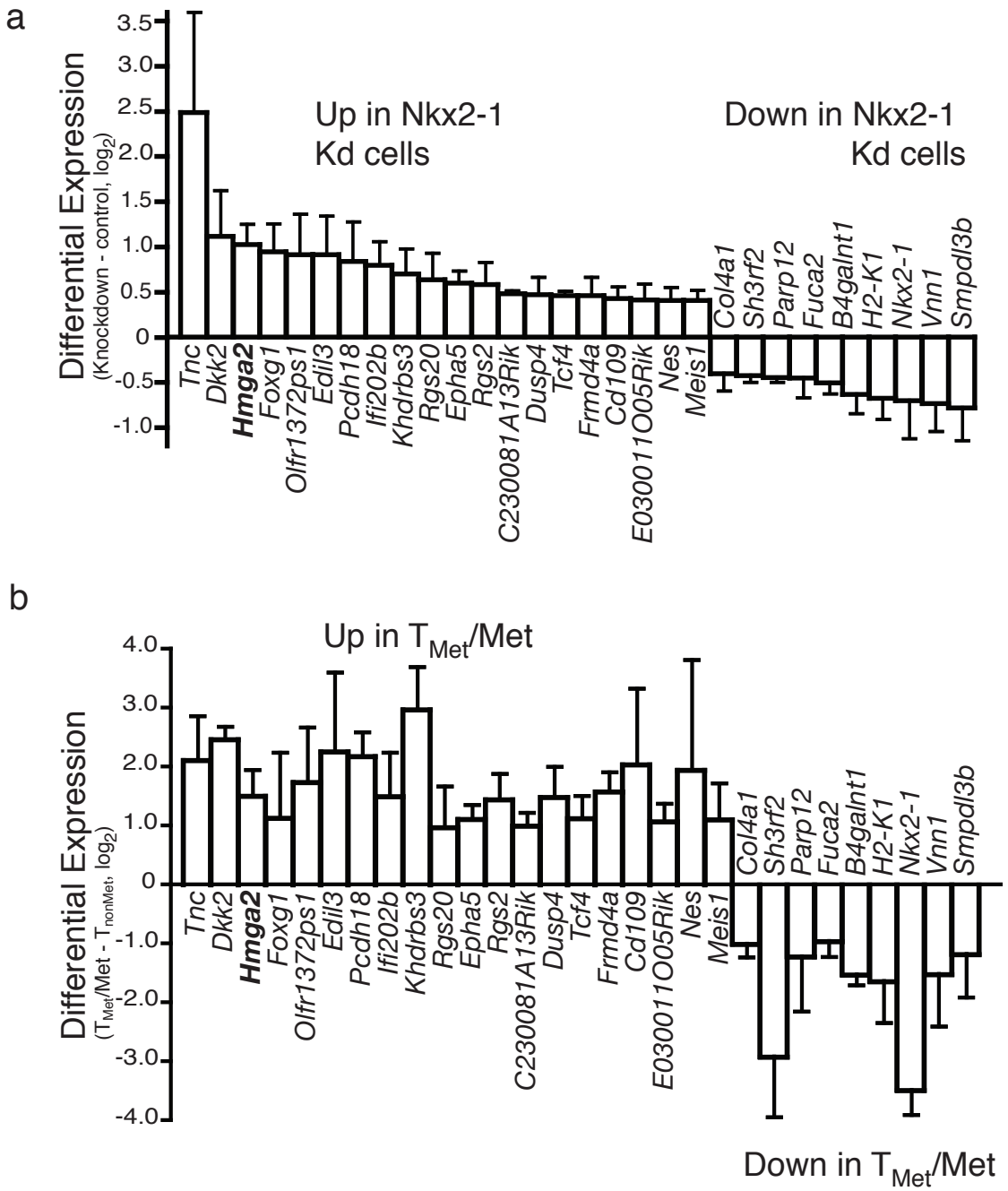
Supplementary Figure 6. Downregulation of Nkx2-1 and expression of Hmga2 in poorly differentiated invasive lung adenocarcinoma

a, Small Nkx2-1^{neg} Hmga2^{pos} areas can be found within lower grade tumours beginning around 12 weeks after tumour initiation. **b,** High grade tumour areas often exist within lower grade tumours. **c,** Verified metastatic lung adenocarcinoma (T3 from Figure 1b) invading a blood vessel (v). Inset shows nuclear atypia. Trichrome staining highlight the collagen rich matrix of this Nkx2-1^{neg} tumour highlights the collagen-rich extracellular matrix present in advanced tumors. Nkx2-1 staining shows the Nkx2-1 negativity of this advanced tumour. **d,** An advanced poorly differentiated tumour invading a blood large blood vessel, triangles mark the smooth muscle layer through which this tumour is invading. This tumour is Nkx2-1^{neg} Hmga2^{pos}. **e,** Advanced Nkx2-1^{neg} Hmga2^{pos} tumours are cytokeratin 8 (CK8)^{pos} and p63^{neg} consistent with adenocarcinomatous origin. p63 positive control staining of the squamous epithelium within the bronchioles is shown.

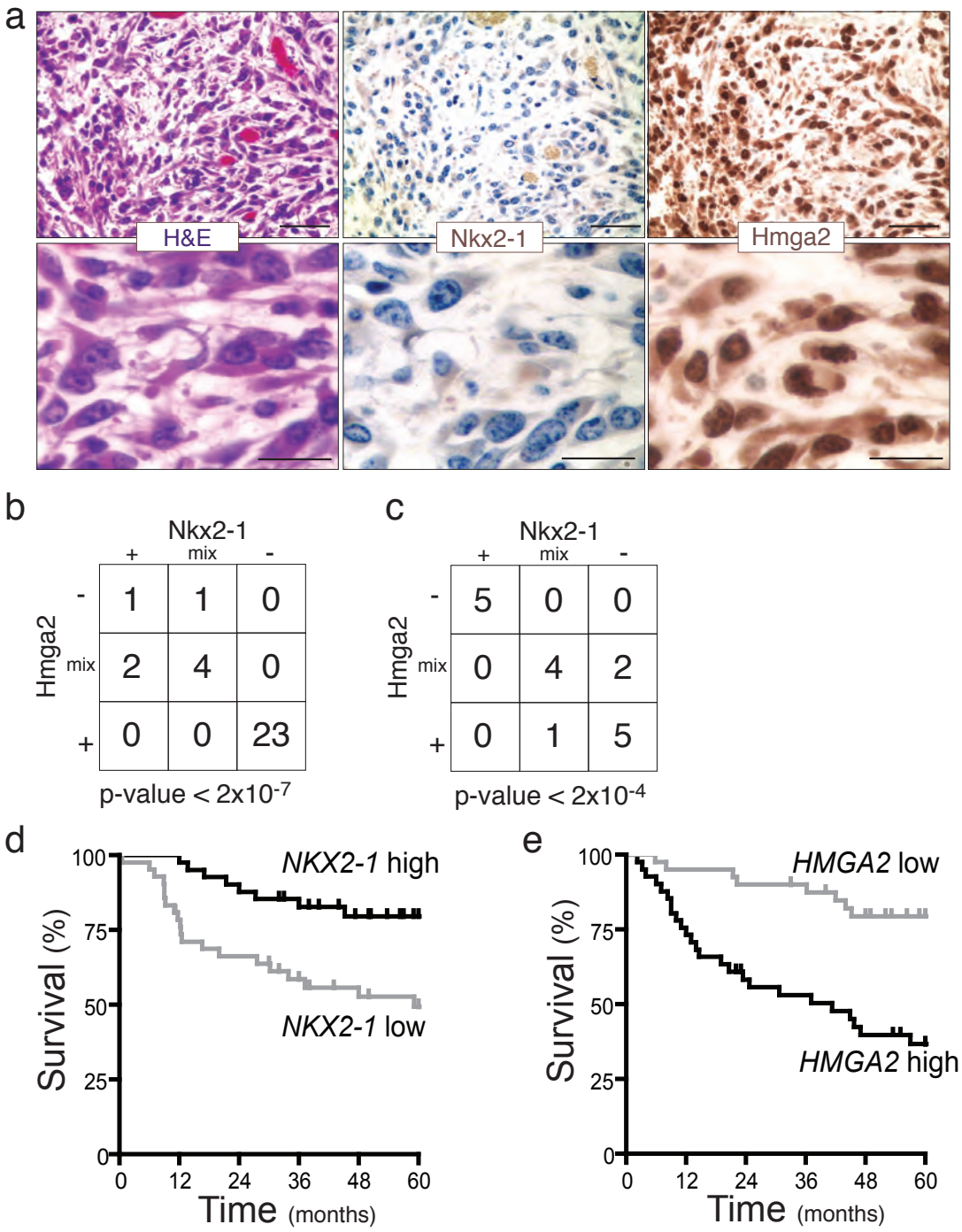


Supplementary Figure 7. Nkx2-1 knockdown enhances tumour formation and anchorage-independent growth but does not alter proliferation

a, shRNA-mediated knockdown of Nkx2-1 in two T_{nonMet} cell lines (368T1 and 394T4). Hsp90 shows equal loading **b,** Kaplan-Meier survival curve after i.v injection of control or Nkx2-1 knockdown cells (n>8 for each group, p-value<0.0001). **c,** Nkx2-1 knockdown does not alter proliferation of T_{nonMet} cells. Representative BrdU incorporation plots are shown. Numbers represent mean +/- SD of triplicate wells. **d,** Comparison of T_{nonMet}-control to T_{nonMet}-*shNkx2-1* cell lines does not show differential expression of two established proliferation gene sets. Paired t-test (N=3). **e,** Nkx2-1 knockdown does not alter T_{nonMet} cell death. Representative Propidium Iodide (PI) staining is shown. Numbers represent mean +/- SD of triplicate wells. Cell death was also comparable when cells were cultured in 0.5% serum (data not shown). **f,** Enhanced anchorage-independent growth of T_{nonMet}-*shNkx2-1* cells. Mean +/- SD of triplicate plates is shown. p-value <0.007. **g,** Subcutaneous tranplantation of T_{nonMet} and T_{nonMet}-*shNkx2-1* cells. Table shows the number of injection sites that formed tumours for each cell number. **h,** Restoration of Nkx2-1 expression with an shRNA-insensitive Nkx2-1 (Nkx2-1*) in T_{nonMet}-*shNkx2-1* cells. Hsp90 shows equal loading. **i,** Nkx2-1* reexperssion reverts the enhanced lung nodule formation after intravenous transplantation.



Supplementary Figure 8. Nkx2-1 controls the expression of Hmga2 in advanced lung adenocarcinoma
A pairwise comparison between the three control and their corresponding shNkx2-1 datasets was used to determine potential Nkx2-1 regulated genes (see Supplementary Table 2). We overlapped the gene list generated by the control versus *shNkx2-1* comparisons (Log_2 >0.4, Paired t-test <0.08) with the gene list generated from the T_{nonMet} versus T_{Met}/Met samples (Log_2 >0.9, unpaired t-test <0.02) to identify high priority targets. **a,** Differential expression of genes in Nkx2-1 knockdown cells. Mean +/- SD. Note that *Nkx2-1* is appropriately reduced in the Nkx2-1 Kd cells. **b,** Differential expression of the same selected genes in T_{nonMet} versus T_{Met}/Met cells. Mean +/- SD.



Supplementary Figure 9. Expression of Nkx2-1 and Hmga2 in murine adenocarcinoma metastasis and the abiity of these genes to predict patient outcome

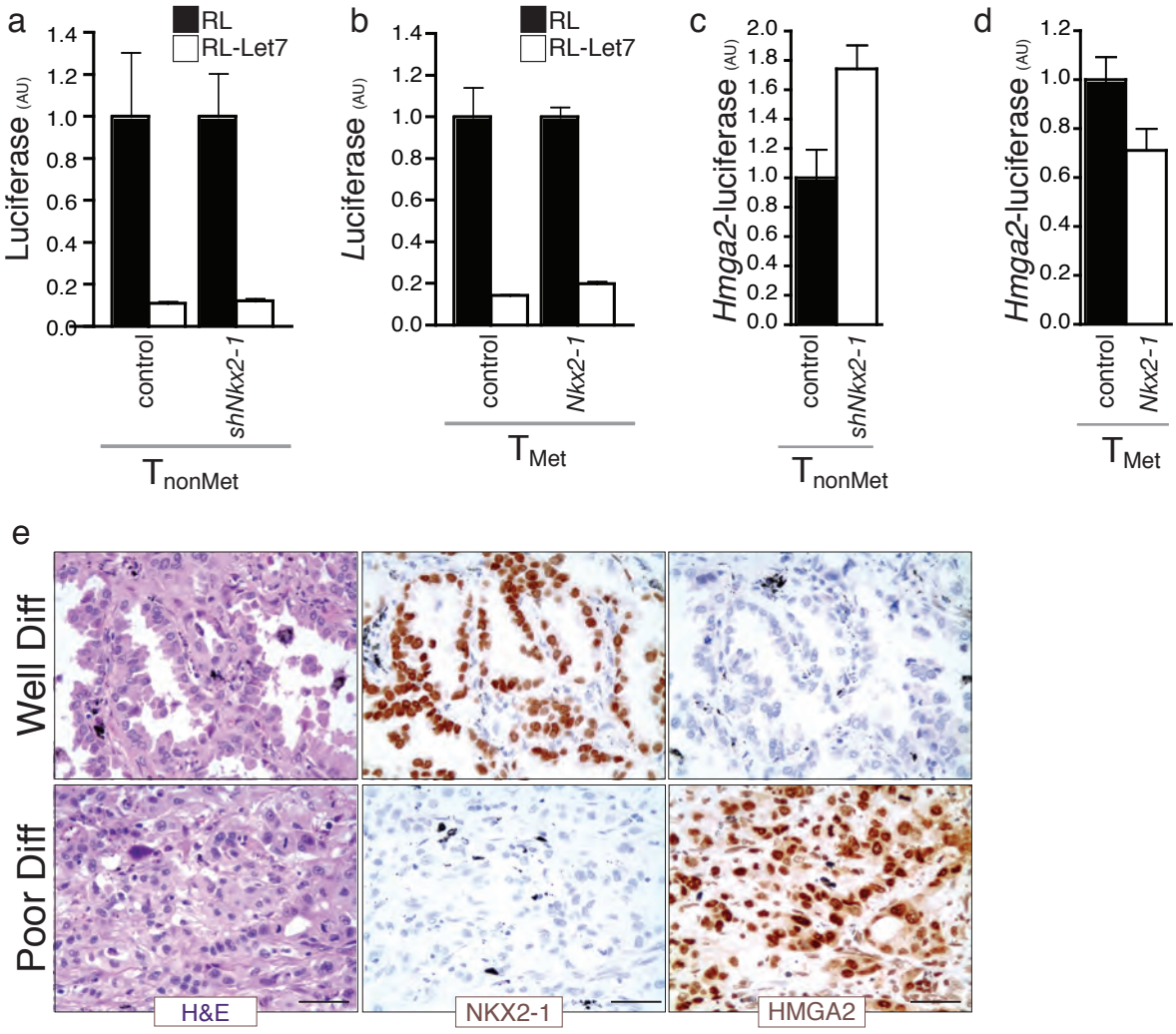
a, Representative image of a large lymph node metastasis. Top row scale bar = 50μm. Bottom row scale bar = 20μm.

b, Contingency tables of Nkx2-1 and Hmga2 expression in lymph node metastases. Fisher's exact test p-value is shown.

c, Contingency tables of Nkx2-1 and Hmga2 expression in distant metastases. Fisher's exact test p-value is shown.

d, Kaplan Meier survival curve comparing patients whose tumors express the highest or lowest levels of *NKX2-1* (top and bottom ten percent (n=36) in each group, from the Shedden *et al.* dataset). This gene expression analysis, and that shown in "e" below agrees with previous immunohistochemical staining for NKX2-1(TTF-1) and HMGA2 in human lung adenocarcinoma patients.

e, Kaplan Meier survival curve comparing patients whose tumors express the highest or lowest levels of *HMGA2* (top and bottom ten percent (n=36) in each group, from the Shedden *et al.* dataset).



Supplementary Figure 10. Nkx2-1 regulates Hmga2

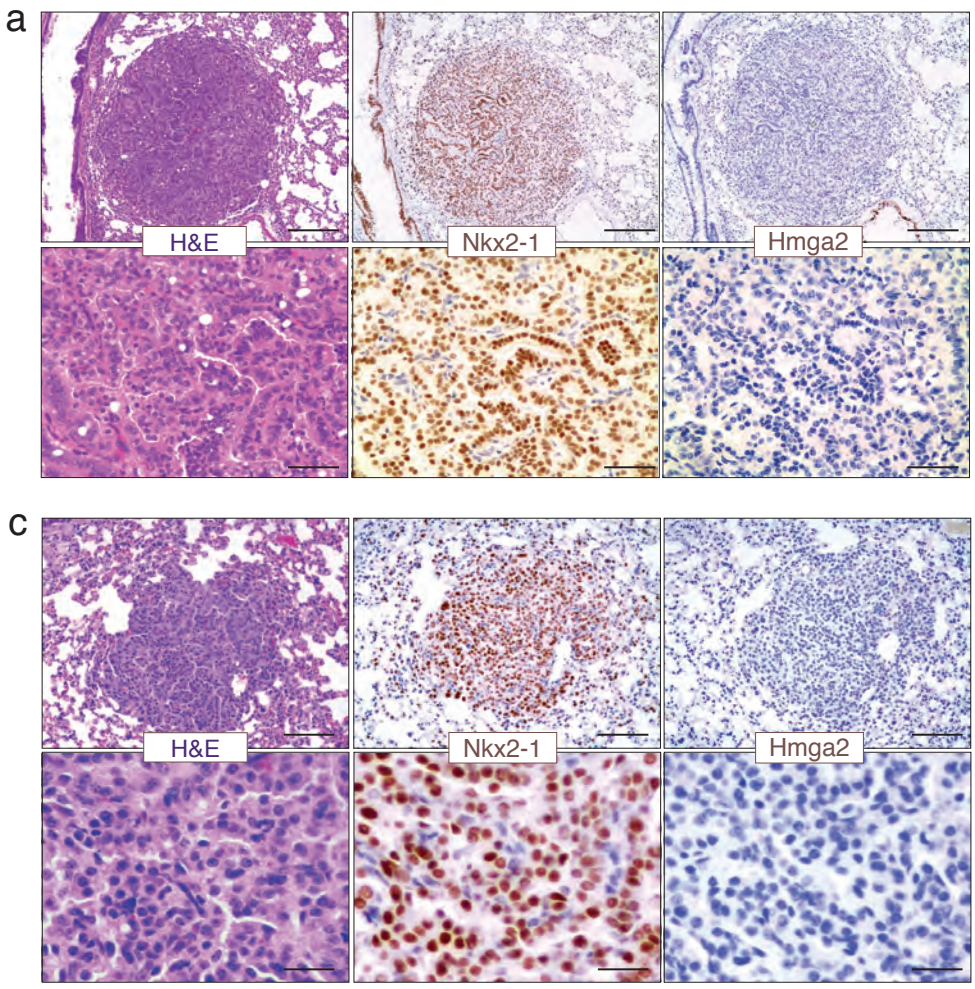
a. Let7 mediated repression is not altered in T_{nonMet} cells with Nkx2-1 knockeddown. Transfection of T_{nonMet} and T_{nonMet}-shNkx2-1 with a control renilla luciferase plasmid (RL) or a renilla luciferase plasmid with Let7 binding sites in the 3' UTR of the luciferase transcript (RL-Let7). Renilla luciferase activity is normalized to firefly luciferase activity from a control cotransfected firefly luciferase plasmid. For each cell line the renilla luciferase activity is normalized to the level in the RL transfected cells.

b. Let7 mediated repression is not altered in T_{Met} cells reexpressing Nkx2-1. Normalization was performed as described in a.

c. Hmga2 promoter activity is derepressed when Nkx2-1 is knocked down in a T_{nonMet} cell line. Hmga2-luciferase activity is the mean +/- SD of triplicate wells, normalized to a control Renilla luciferase plasmid and to the control cell line.

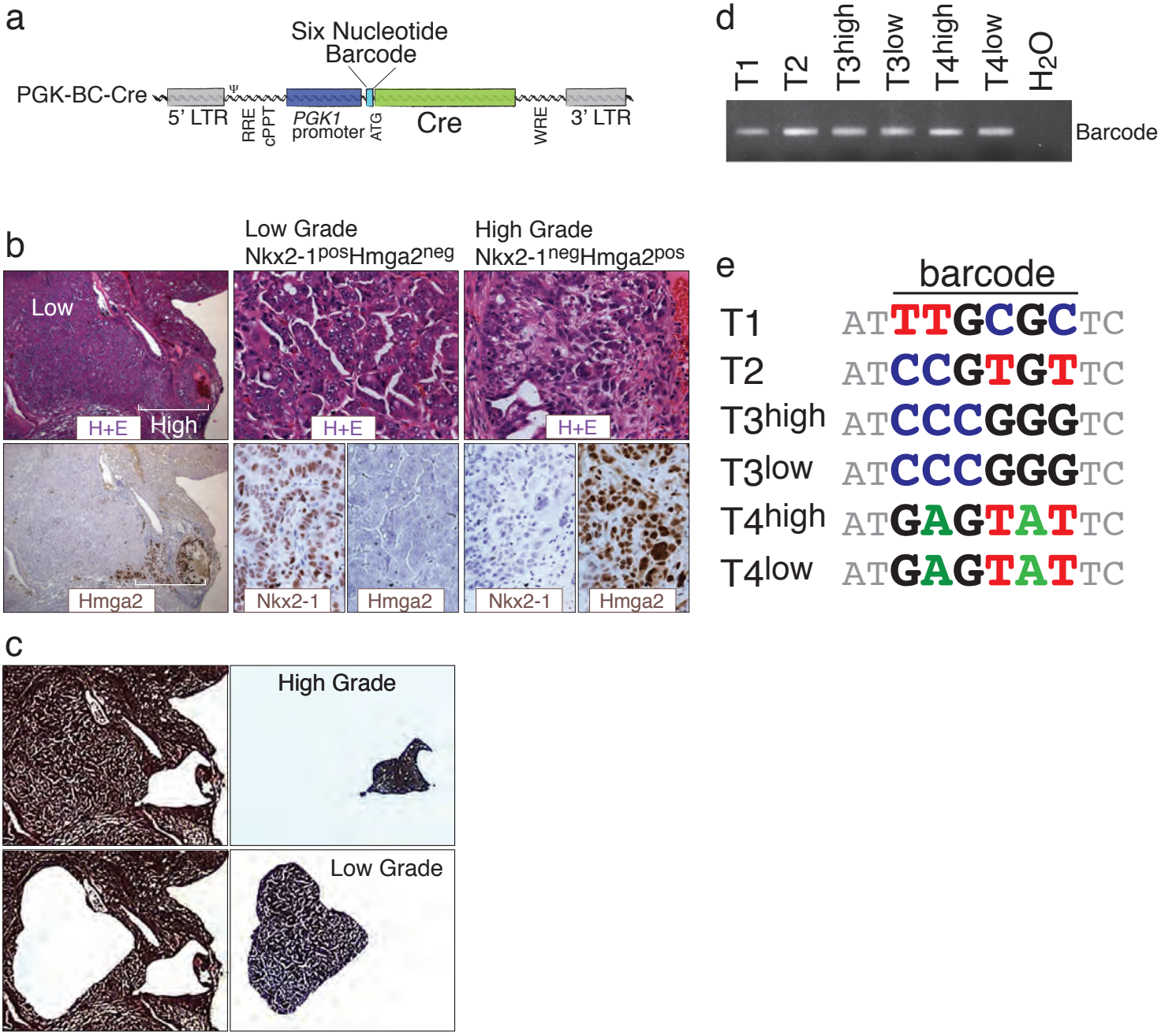
d. Hmga2 promoter activity is repressed when Nkx2-1 is expressed in a T_{Met} cell line. Hmga2-luciferase activity is the mean +/- SD of triplicate wells, normalized to the control cell line.

e. Top panels show a representative NKX2-1 positive, well-differentiated human lung adenocarcinoma. Bottom panels show a representative NKX2-1 negative, poorly-differentiated human lung adenocarcinoma. Scale bar = 50μm.

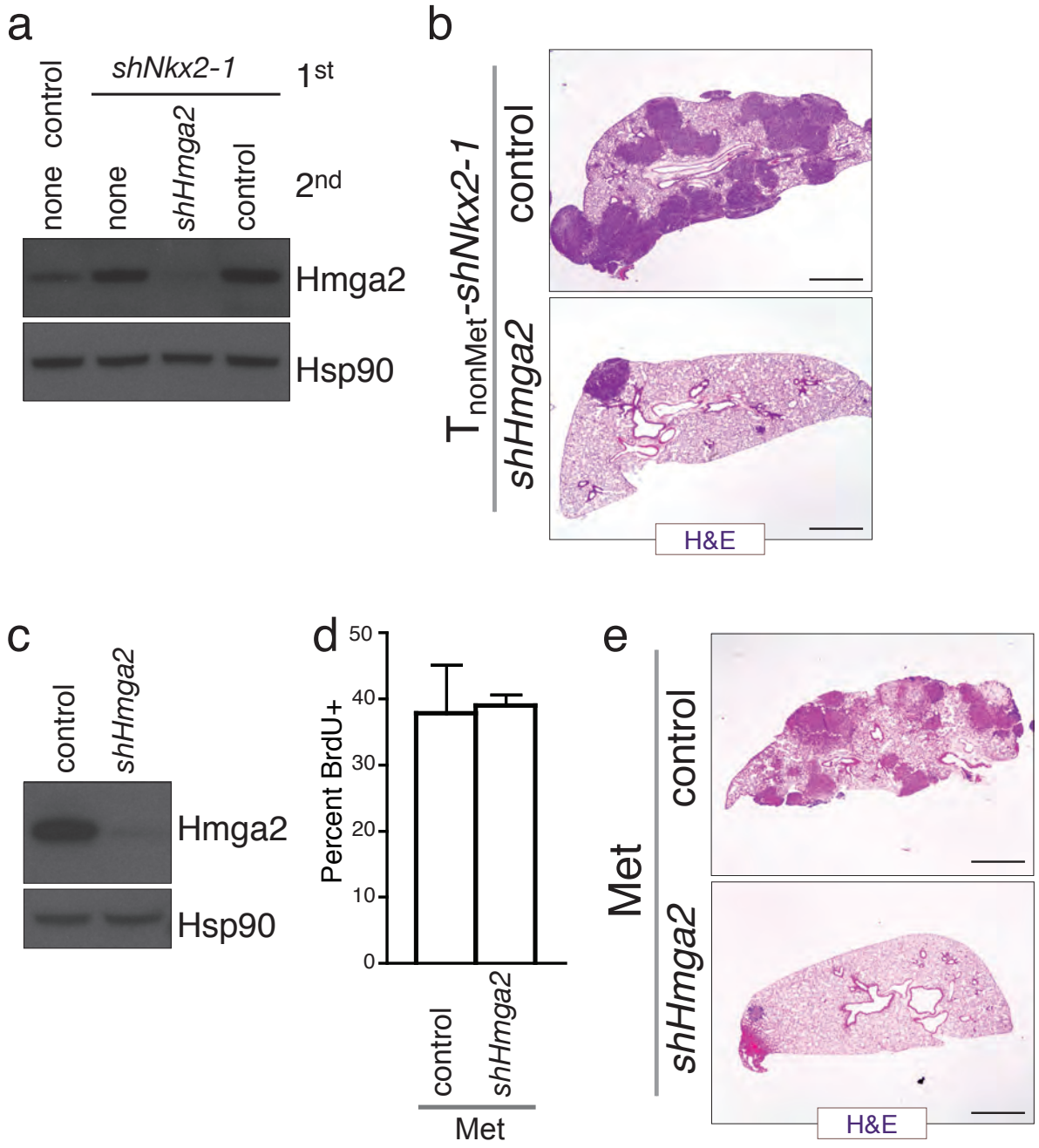


Supplementary Figure 11. Early Adenomas are homogenously Nkx2-1-positive and p53-deficiency is required for progression to the Nkx2-1 negative state.

a, Representative images of an Nkx2-1^{pos}Hmga2^{neg} *Kras*^{G12D}-only lung adenoma. Top row scale bar = 200μm. Bottom row scale bar = 50μm. Note that the positive Hmga2 staining in the lower right corner is cytoplasmic staining in smooth muscle cells **b,** Contingency table shows that *Kras*^{G12D}-onlytumours are almost universally Nkx2-1^{pos}Hmga2^{neg} **. c,** Representative image of an early Nkx2-1^{pos}Hmga2^{neg} *Kras*$^{G12D/+}$;*p53*$^{\Delta/\Delta}$ lung adenoma. Top row scale bar = 100μm. Bottom row scale bar = 25μm. Note that in each of these tumor classes the Nkx2-1^{neg} cells within the tumours are either normal Type I pneumocytes or other stromal cell and that there are no Hmga2^{pos} cells. These observation show that early during tumor development all cells within the tumour are Hmga2^{neg} and support a model where these Nkx2-1^{pos}Hmga2^{neg} cell undergo a transistion to a more agressive Nkx2-1^{neg}Hmga2^{pos} phenotype. Conversely, if the eventual transition to an Nkx2-1^{neg}Hmga2^{pos} phenotype is due to the expansion of a rare Nkx2-1^{neg} tumor cell population, the phenotype of these expanded Nkx2-1neg tumour cells would be distinct based on the gain of Hmga2 expression which likely contributes to the dedifferentiation and gain of embryonic character.



Supplementary Figure 12. Low-grade and high-grade tumour areas are clonally related
a, Diagram of the barcoded lentiviral vector with which tumours were induced in *Kras*$^{LSL\text{-}G12D/+}$;*p53*$^{f/f}$ mice. These vectors express Cre to initiate tumours and also contain a random 6-nucleotide barcode. b, A section with adjacent low-grade and high-grade tumour areas is shown. H+E images are shown at top. Hmga2 and Nkx2-1 IHC are shown on bottom. The Hmga2^{neg} cells in the high-grade areas are stromal cells including those of a major blood vessel through which the tumour has grown. High magnification Nkx2-1 and Hmga2 IHC shows the phenotype of these areas. **c,** Laser-capture microscopy (LCM) images show remaining areas after LCM and the isolated tumor (these areas are referred to as T4^{high} and T4^{low} in subsequent panels). **d,** The region of the integrated lentivirus containing the nucleotide barcode was PCR amplified from each LCM tumor area. Cases where adjacent low-grade and high-grade areas were isolated are indicated as high (high grade) and low (low grade). **e,** Barcode sequencing results for the LCM captured areas showing the barcode sequence in color and the adjacent vector sequence in grey.



Supplementary Figure 13. Nkx2-1-regulated Hmga2 contributed to tumor progression

a. Hmga2 knockdown reduces the tumourigenic potential of T_{nonMet}-*shNkx2-1* after intravenous transplantation. Western blot shows reduction in Hmga2. Hsp90 shows equal loading. **b,** Representative H+E stained lung sections after transplantation of T_{nonMet}-*shNkx2-1* with an additional control or Hmga2 targeting shRNA. Scale bar = 0.5mm. The tumours that did form after transplantation of Hmga2 knockdown cells were $Hmga2^{pos}$ (data not shown) further suggesting that Hmga2 is a critical component of the gene expression program elicited by Nkx2-1 downregulation during lung tumour progression. **c,** Knockdown of Hmga2 in a lymph node metastasis-derived cell line (482N1). Hsp90 shows equal loading. **d,** Hmga2 knockdown does not alter proliferation (as measured by BrdU incorporation) under standard culture conditions. Mean +/- SD of triplicate wells. **e,** Representative H+E stained lung sections after transplantation of a metastasis-derived cell line (Met; 482N1) with a control or Hmga2 targeting shRNA (*shHmga2*).