Title

The Lineage-Defining Transcription Factors SOX2 and NKX2-1 Determine Lung Cancer Cell Fate and Shape the Tumor Immune Microenvironment

Journal

Immunity, 49(4)

ISSN

1074-7613

Authors

Mollaoglu, Gurkan
Jones, Alex
Wait, Sarah J
et al.

Publication Date

2018-10-01

DOI

10.1016/j.immuni.2018.09.020

Peer reviewed

HHS Public Access

Author manuscript

Immunity. Author manuscript; available in PMC 2019 October 16.

Published in final edited form as:

Immunity. 2018 October 16; 49(4): 764–779.e9. doi:10.1016/j.immuni.2018.09.020.

The lineage defining transcription factors SOX2 and NKX2-1 determine lung cancer cell fate and shape the tumor immune microenvironment

Gurkan Mollaoglu[1], **Alex Jones**[2], **Sarah J. Wait**[1], **Anandaroop Mukhopadhyay**[1], **Sangmin Jeong**[1], **Rahul Arya**[1], **Soledad A. Camolotto**[2], **Timothy L. Mosbruger**[3], **Chris J. Stubben**[3], **Christopher J. Conley**[3], **Arjun Bhutkar**[4], **Jeffery M. Vahrenkamp**[1], **Kristofer C. Berrett**[1], **Melissa H. Cessna**[5], **Thomas Lane**[2], **Benjamin L. Witt**[2,6], **Mohamed E. Salama**[2,6], **Jason Gertz**[1], **Kevin B. Jones**[1,7], **Eric L. Snyder**[1,2], and **Trudy G. Oliver**[1,8,*]

[1]Department of Oncological Sciences, Huntsman Cancer Institute, University of Utah, Salt Lake City, UT 84112, USA.

[2]Department of Pathology, University of Utah, Salt Lake City, UT 84112, USA.

[3]Huntsman Cancer Institute, Bioinformatics Shared Resource, Salt Lake City, UT 84112, USA.

[4]David H. Koch Institute for Integrative Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139, USA.

[5]Intermountain Biorepository, Intermountain Healthcare, Salt Lake City, UT 84111, USA.

[6]ARUP Laboratories at University of Utah, Salt Lake City, UT 84108, USA.

[7]Department of Orthopaedics, University of Utah, Salt Lake City, UT 84112, USA.

[8]Lead Contact

SUMMARY

The major types of non-small cell lung cancer (NSCLC) - squamous cell carcinoma and adenocarcinoma - have distinct immune microenvironments. We developed a genetic model of

*Correspondence: Trudy.Oliver@hci.utah.edu.

AUTHOR CONTRIBUTIONS

Conceptualization, G.M., E.L.S., T.E.L., T.G.O.; Methodology, G.M., A.J., A.M., K.B.J., E.L.S., T.G.O.; Formal Analysis, G.M., A.J., T.L.M., A.B., C.S., C.J.C., J.M.V., J.G., M.E.S., E.L.S., T.G.O.; Investigation, G.M., A.J., A.M., K.C.B., D.H., S.A.C., S.J., S.J.W., S.C., B.L.W., J.G., M.E.S., K.B.J., E.L.S., T.G.O.; Pathology, B.L.W., M.E.S., E.L.S.; Resources, M.H.C.; Data Curation, -; Writing – Original Draft, G.M. & T.G.O.; Writing – Review & Editing, G.M., E.L.S., T.G.O.; Visualization, G.M., T.G.O.; Supervision, E.L.S., T.G.O.; Project Administration, E.L.S., T.G.O.; Funding Acquisition, G.M., E.L.S., T.E.L., T.G.O.

DECLARATION OF INTERESTS

The authors have no competing interests.

Data and Software Availability

Data Resources

Tissue RNA-sequencing data, single-cell RNA-sequencing data, and SOX2 chromatin immunoprecipitation sequencing data in this study have been deposited in the NCBI GEO (GEO: GSE118252).

squamous NSCLC based on overexpression of the transcription factor Sox2, which specifies lung basal cell fate, and loss of the tumor suppressor *Lkb1* (SL mice). SL tumors recapitulated gene expression and immune infiltrate features of human squamous NSCLC, including enrichment of tumor-associated neutrophils (TANs) and decreased expression of NKX2–1, a transcriptional regulator that specifies alveolar cell fate. In *Kras*-driven adenocarcinomas, mis-expression of *Sox2* or loss of *Nkx2–1*, led to TAN recruitment. TAN recruitment involved SOX2-mediated production of the chemokine CXCL5. Deletion of *Nkx2–1* in SL mice (SNL) revealed that NKX2–1 suppresses SOX2-driven squamous tumorigenesis by repressing adeno-to-squamous transdifferentiation. Depletion of TANs in SNL mice reduced squamous tumors, suggesting that TANs foster squamous cell fate. Thus, lineage defining transcription factors determine the tumor immune microenvironment, which in turn may impact the nature of the tumor.

Keywords

lung cancer; SOX2; NKX2–1; tumor-associated neutrophils; tumor immune microenvironment; mouse models; transdifferentiation

INTRODUCTION

Non-small cell lung cancer (NSCLC) represents approximately 85% of all lung cancers and is predominantly comprised of lung squamous cell carcinoma (LSCC) and adenocarcinoma (LADC). LSCC and LADC can be distinguished by their histopathology, biomarkers, gene expression patterns, genomic alterations and response to therapy (Campbell et al., 2016; Langer et al., 2016). Adenosquamous carcinoma is a less common NSCLC variant defined by the presence of both LADC and LSCC components. Recently, it has been shown that LADC and LSCC have differences in the immune microenvironment (Kargl et al., 2017; Nagaraj et al., 2017; Xu et al., 2014). This is clinically relevant because advances in immunotherapy have led to the approval of immune checkpoint inhibitors for NSCLC (Ribas and Wolchok, 2018). Immunotherapy response is impacted by the complex and dynamic interactions among multiple immune cell types and cancer cells in the tumor immune microenvironment (TIME) (Pitt et al., 2016). It is important to decipher the function of cells in the TIME to better understand the role of the immune system in tumor initiation and progression and to fully exploit the potential of immunotherapy.

Neutrophils are among the first responders to infection and tissue damage (Powell and Huttenlocher, 2016). Neutrophils are present in the TIME of most solid tumors and can contribute to tumor progression by promoting cell growth, angiogenesis, metastasis and immune evasion, although they can play anti-tumor roles as well (Coffelt et al., 2016; Powell and Huttenlocher, 2016). A pan-cancer study of 25 malignancies including NSCLC showed that neutrophil abundance in the TIME is a leading predictor of poor outcome (Gentles et al., 2015). Similarly, a high neutrophil to lymphocyte ratio is associated with poor prognosis in many solid tumors, including NSCLC (Templeton et al., 2014). Mice bearing activating mutations in *Kras* and deficient for the tumor suppressor Lkb1 [$Kras^{LSL\text{-}G12D/+};Lkb1^{fl/fl}$ (KL mice)] harbor a spectrum of NSCLC histologies, and tumor-associated neutrophils

(TANs) in this model demonstrate pro-tumor features (Koyama et al., 2016; Nagaraj et al., 2017).

Neutrophils are more abundant in human LSCC compared to LADC, whereas macrophages are more abundant in LADC (Eruslanov et al., 2014; Kargl et al., 2017). TAN enrichment in LSCC is also observed in genetically engineered mouse models (GEMMs); LSCCs in *Lkb1*$^{fl/fl}$;*Pten*$^{fl/fl}$ (LP) and *Sox2*$^{LSL/LSL}$;*Pten*$^{fl/fl}$;*Cdkn2ab*$^{fl/fl}$ (Sox2PC) mice have more neutrophils but fewer macrophages compared to LADC GEMMs (Ferone et al., 2016; Xu et al., 2014). Similarly, adenosquamous tumors in KL mice have more TANs and fewer macrophages compared to adenocarcinomas in the same genetic model (Nagaraj et al., 2017). Based on this result, it was hypothesized that squamous histological classification (i.e. histotype) rather than genetic alterations (i.e. genotype) determine the immune contexture (Nagaraj et al., 2017). Tumor histology and genetic alterations are intimately linked in NSCLC, so it is difficult to tease apart whether histotype or genotype is the most important determinant of the TIME. A significant subset of lung tumors is difficult to classify histologically (so-called “not otherwise specified” tumors), and both tumor heterogeneity and plasticity, including changes in tumor subtype in response to targeted therapy, occur in the clinical setting. Therefore, it is important to determine whether and how histotype and/or genotype dictate the immune microenvironment. Specifically, it remains unknown which factors promote the difference in TAN levels between LSCC and LADC.

During development, the transcription factors SOX2 and NKX2–1 exhibit opposing patterns of expression within the primitive foregut and embryonic lung (Morrisey and Hogan, 2010). Likewise, in the adult lung, SOX2 levels are highest in the proximal airways while NKX2–1 levels are highest in type 2 pneumocytes of the alveoli. In the normal lung, SOX2 promotes proliferation and maintains stem and basal cell identity, whereas NKX2–1 specifies alveolar cell identity. SOX2 and NKX2–1 exhibit opposing patterns of expression in the two major subtypes of NSCLC as well. *SOX2* is amplified in ~21% and overexpressed in >80% of LSCCs, whereas it is rarely expressed in LADC (Campbell et al., 2016). In contrast, *NKX2–1* is amplified in ~10% and highly expressed in 70–80% of LADC, and is rarely expressed in LSCC (Campbell et al., 2016). LADC shares features with normal alveolar cells where NKX2–1 is normally expressed, whereas LSCC shares features of basal cells where SOX2 is normally expressed. *SOX2* is thus considered a lineage-specific oncogene because it drives multiple squamous cancers including lung, and directs tumor type toward a basal cell fate (Bass et al., 2009; Ferone et al., 2016; Lu et al., 2010; Mukhopadhyay et al., 2014). On the other hand, the impact of NKX2–1 in LADC appears to be context dependent. *NKX2–1* is considered to be a lineage-specific oncogene in adenocarcinoma when it is amplified or highly expressed (Tanaka et al., 2007). However, a subset of adenocarcinomas is NKX2–1-negative, and these tumors are associated with poor prognosis relative to NKX2–1^{+} adenocarcinomas. Furthermore, loss of NKX2–1 accelerates *Kras*-mutant lung adenocarcinoma in mice and alters their cell fate, suggesting that NKX2–1 has lineage-specific tumor suppressor functions as well. Here, we developed and employed multiple GEMMs to elucidate the role of lung cancer lineage specifiers SOX2 and NKX2–1 in tumor cell fate and neutrophil recruitment.

RESULTS

Murine SOX2-Driven Lung Squamous Tumors Recapitulate Human Pathology

Our previous work demonstrated that lentiviral *Sox2* overexpression combined with *Lkb1* deletion promotes LSCC (Mukhopadhyay et al., 2014). To create a genetic model, we generated mice with a *Lox-Stop-Lox-Sox2-IRES-GFP* cassette in the *Rosa26* locus (*Rosa26*$^{LSL\text{-}Sox2\text{-}IRES\text{-}GFP}$) and crossed them to *Lkb1*$^{fl/fl}$ mice to yield *Rosa26*$^{LSL\text{-}Sox2\text{-}IRES\text{-}GFP}$;*Lkb1*$^{fl/fl}$ (SL) mice (Figure S1A). To confirm that the *Sox2* allele functions properly, we isolated mouse embryonic fibroblasts (MEFs) from *Sox2*$^{LSL/+}$;*Lkb1*$^{fl/fl}$ and *Sox2*$^{LSL/LSL}$;*Lkb1*$^{fl/fl}$ mice (Figure S1B). GFP and SOX2 expression were induced in MEFs infected with adenoviruses that express Cre recombinase under the control of a CMV promoter (Ad5-CMV-Cre), and *Sox2* homozygous cells expressed both proteins at higher levels than heterozygous cells (Figures S1C and S1D).

Next, we infected homozygous and heterozygous SL mice, LP mice, and *Kras*$^{LSL\text{-}G12D/+}$;*Trp53*$^{fl/fl}$ (KP) mice with Ad5-CMV-Cre using intranasal inhalation. While heterozygous SL mice did not develop tumors by one year post-infection, homozygous SL mice developed tumors after an average of 11 months, comparable to the latency of other LSCC GEMMs (Ferone et al., 2016; Xu et al., 2014) (Figures S1E and S1F). Tumor penetrance was higher in SL mice than Lenti-Sox2-Cre-infected *Lkb1*$^{fl/fl}$ mice (71% vs. 40%) and slightly higher than that of LP mice (60%) (Figure S1G). SL mice developed one or a few tumors similar to the LP model (Figure S1H) and were GFP positive by whole mount microscopy (Figure S1I). The vast majority of tumors in the SL mice (> 90%) were classified as squamous by two board-certified pathologists (B.L.W., E.L.S.), while smaller LADCs were rarely detected in the same animals (Figures 1A and S1J). SL tumors exhibited characteristic features of squamous tumors with flattened and keratinized cells and desmoplastic stroma, similar to human LSCC and murine LP tumors (Figure 1A). SL tumors were positive for two established LSCC biomarkers KRT5 and DNp63 (Figures 1B–1D).

Mutations in components of the PI3K-AKT-mTOR pathway result in activation of this signaling axis in ~47% of LSCC (Campbell et al., 2016). LKB1 functions as a tumor suppressor through its negative regulation of the mTOR pathway. *LKB1* loss of function alterations are relatively infrequent in human LSCC but are found in a significant fraction of adenosquamous tumors (Zhang et al., 2017). We assessed mouse and human LSCC for phosphorylated 4EBP1 (p4EBP1), an established marker of mTOR pathway activation. SL tumors were positive for SOX2 and p4EBP1 at levels similar to LP and human LSCC (Figures 1E–1G). KP adenocarcinomas were negative for SOX2 and had low to no detectable p4EBP1. IHC analysis of 43 human LSCCs demonstrated that 58% of samples were positive for p4EBP1, and 93% were positive for SOX2 (with 58% positive for both), suggesting that murine SL tumors resemble a significant fraction of human LSCCs (Figures 1H and 1I). Together, these results demonstrate that the combination of SOX2 overexpression and mTOR pathway activation, frequent characteristics of human LSCC, promotes LSCC tumorigenesis.

Murine SOX2-Driven Lung Squamous Tumors Recapitulate Molecular Phenotype of Human LSCC

To determine whether SL tumors recapitulate the molecular features of human LSCC, we compared gene expression profiles of SL, LP and KP mouse tumors to normal adult mouse lung and human LSCC by RNA-sequencing (RNA-seq). Cross-species analysis of differentially expressed genes demonstrated similarities among SL, LP and human LSCC (Figure 2A). Mouse and human squamous tumors aligned more closely in a semi-unsupervised clustering analysis and were distinct from KP adenocarcinomas and normal lung. Expression of squamous biomarkers such as *Krt5*, *Krt14*, other cytokeratin genes, as well as *Trp63* and *Sox2* was significantly increased in SL tumors, whereas expression of adenocarcinoma markers such as *Nkx2–1*, *Sftpc* and other surfactant genes was decreased (Figure 2B). Gene set enrichment analysis (GSEA) of human LSCC compared to LADC derived from The Cancer Genome Atlas (TCGA) revealed a significant positive enrichment for genes expressed in SL or LP versus KP tumors (Figures 2C and S2A). There was a significant depletion of adenocarcinoma-enriched genes in SL and LP versus KP tumors (Figures 2C and S2A). Ingenuity pathway analysis (IPA) of differentially expressed genes in SL versus KP tumors identified human and mouse “embryonic stem cell pluripotency” pathways (Figure S2B). By GSEA, SL tumors and human LSCCs were enriched for gene sets including the DNp63 pathway, lung stem cell markers (Vaughan et al., 2015; Zuo et al., 2015) and tracheal basal cell surface markers (Van de Laar et al., 2014) (Figure S2C and Tables S2 and S3). Together, these findings suggest that SL tumors strongly recapitulate human LSCCs at the molecular level.

Mouse Lung Tumors Recapitulate Human Tumor Immune Microenvironment with Elevated Levels of Tumor-Associated Neutrophils

IPA identified “granulocyte adhesion and diapedesis” among the top enriched pathways in SL versus KP tumors (Figure S2B). Compared to KP tumors, IPA identified multiple immune-related pathways as significantly enriched in SL tumors, including specific immune cell types, cytokines and transcription factors as key upstream regulators (Figures S2D and S2E). The comparison of all significantly enriched pathways in human LSCC versus LADC (Campbell et al., 2016) and SL versus KP tumors demonstrated that there is a high degree of overlap of immune-related pathways between mouse and human LSCC (Figure S2F). GSEA also suggested that SL tumors are enriched for gene signatures of immune cells in lymphoid, myeloid and macrophage lineages (Figure S2G). Compared to normal lung tissue and KP tumors, SL tumors had a significant positive enrichment for T cell, neutrophil and TAN gene sets (Figures 2D and S2H). Consistent with TAN enrichment in squamous tumors, we found that multiple upstream molecules associated with neutrophil recruitment and “N2” pro-tumor function were enriched in SL tumors (Figure S2D). Together, this suggested that LSCC and LADC have significantly different TIMEs.

In order to better understand the LSCC TIME, we analyzed the abundance of T cells in murine tumors. IHC for CD3 demonstrated significantly more T cell infiltration in SL and LP tumors compared to KP adenocarcinomas (Figure 2E). High neutrophil infiltration in SL tumors was evident on H&E evaluation by a pathologist (E.L.S.) (Figure S2I) and confirmed by IHC analysis of neutrophil markers CD11B, MPO and LY6G (Figure 2E). Neutrophils

were found in extravascular clusters located throughout the TIME; these clusters were found at the periphery and within the tumor, often immediately adjacent to the carcinoma cells. Because squamous tumors were apparently growing in this inflammatory and T cell rich microenvironment, we postulated that immune evasion mechanisms may be employed by the cancer cells. We found that $FOXP3^+$ regulatory T (Treg) cells were enriched in SL tumors compared to LP and KP tumors (Figure 2E). SL tumors also displayed increased expression of multiple immunosuppressive molecules such as *Arg1*, *Vtcn1*, *Cd80*, *Btla*, *Havcr2* and *Cd274 (Pd-l1)*, as measured by RNA-seq, some of which are increased in LP squamous tumors (Xu et al., 2014) (Figure 2F). SL tumors showed reduced expression of major histocompatibility complex (MHC) genes (Figure 2F), suggesting reduced antigen presentation as a potential immune evasion mechanism active in these tumors. Collectively, SL squamous tumors display some hallmarks of an immunosuppressive TIME including enrichment of Treg cells and TANs. This does not rule out the possibility that adenocarcinoma also harbors an immunosuppressive TIME, albeit potentially through distinct mechanisms.

SOX2 Promotes Tumor-Associated Neutrophil Recruitment in the Absence of Squamous Transdifferentiation

Recent studies suggest that elevated TANs in LSCC may be due to a squamous histotype as opposed to a genetic phenomenon (Nagaraj et al., 2017). It is challenging to decouple the function of SOX2 from its role in “squamousness” since SOX2 can promote squamous differentiation (Bass et al., 2009; Lu et al., 2010; Mukhopadhyay et al., 2014). To tease apart these functions, we overexpressed SOX2 in the KP adenocarcinoma model to determine whether SOX2 can alter tumor histotype and/or TAN levels. KP mice were infected with Lenti-Sox2-Cre or control Lenti-GFP-Cre viruses (Figure 3A) and lungs were harvested four months later. Lenti-GFP-Cre-infected KP tumors (KPG) were GFP-positive as expected (Figure 3B). We analyzed serial sections by IHC for squamous and neutrophil markers in matched individual tumors. In Lenti-Sox2-Cre-infected KP animals (KPS), SOX2 overexpression was observed in many but not all tumors, indicating an occasional uncoupling of SOX2 expression from Cre-mediated recombination (Figures 3C and3D). All KPS tumors exhibited adenocarcinoma histology regardless of SOX2 protein levels (Figure 3B). Neither $SOX2^+$ nor $SOX2^-$tumors in KPS lungs expressed squamous biomarkers DNp63 or KRT5 (Figures 3C and 3E-3F), suggesting that SOX2 is not sufficient to promote squamous fate in this context. However, we observed elevated TAN levels specifically in $SOX2^+$ KPS tumors but not in control $SOX2^-$ KPS, KPG or KP tumors (Figures 3C and3G-3I). Based on analyses of individual $SOX2^+$ and $SOX2^-$ tumors within the same lungs, TAN recruitment appeared to be localized to specific tumors expressing SOX2 rather than a systemic effect of some tumors throughout the entire lung. Together, these data suggest that SOX2 is sufficient to promote TAN recruitment independent of tumor histotype.

SOX2 Suppresses NKX2–1 Activity and NKX2–1 Loss Promotes TAN Recruitment in the Absence of Squamous Histotype

Next, we aimed to understand the mechanisms of TAN recruitment by SOX2 in a histotype-independent manner. During lung development, SOX2 and NKX2–1 exhibit inverse patterns of expression at multiple stages. In the course of investigating KPS tumors, we observed that

NKX2–1 protein amounts were significantly decreased in $SOX2^+$ KPS tumors compared to $SOX2^-$ KPS and KP tumors (Figures 4A and 4B). Consistent with the reduction in NKX2–1 levels, $SOX2^+$ KPS tumors had significantly lower protein amounts of SPC (an established NKX2–1 target gene product) compared to control tumors (Figures 4A and 4C). SL and LP tumors, both of which have high SOX2 expression, also had subtly but significantly lower protein amounts of NKX2–1 compared to KP adenocarcinomas (Figures 4D and 4E). Lower levels of NKX2–1 were due to reduced intensity (evident by comparing NKX2–1 levels in tumor cells to neighboring alveolar cells) rather than a reduction in the number of tumor cells expressing NKX2–1 (Figures S3A and S3B). In squamous tumors, SPC protein amounts were significantly reduced further suggesting suppressed NKX2–1 activity in cells with high SOX2 protein (Figures 4D and 4F). Consistently, IPA identified NKX2–1 inhibition as a top upstream regulator in SL tumors compared to KP tumors (Figure 4G). Together this suggests that SOX2 promotes a reduction in NKX2–1 levels and/or activity. While the mechanism of regulation is unclear, chromatin immunoprecipitation-RNA sequencing (ChIP-seq) analyses suggested that SOX2 and NKX2–1 can each bind their own promoters, and SOX2 may weakly bind the *Nkx2–1* promoter (Figure S3C). Furthermore, inducible SOX2 expression can repress NKX2–1 protein amounts in four of five human lung cancer cell lines examined (Figure S3D).

As SOX2 promotes TAN recruitment and leads to decreased NKX2–1 activity, we tested whether NKX2–1 plays a role in TAN recruitment. We employed $Kras^{FSF\text{-}G12D/+};Trp53^{Frt/Frt};Rosa26^{FSF\text{-}CreERT2};Nkx2\text{–}1^{fl/fl}$ (KPN) mice, where *Nkx2–1* can be temporally deleted in the KP adenocarcinoma model. KPN mice were infected with Ad5-CMV-Flp to initiate tumorigenesis and six weeks post-infection, treated with vehicle or tamoxifen to cause *Nkx2–1* deletion. Four weeks after tamoxifen treatment, tumors were harvested for IHC analysis. KPN tumors exhibited a significant reduction in NKX2–1 and SPC expression compared to KP tumors, exhibited a mucinous adenocarcinoma histology (Figures 4H–4J) consistent with our prior studies (Snyder et al., 2013), and did not express SOX2 (Figures 4H and 4K). However, *Nkx2–1* deletion in KPN adenocarcinomas resulted in significantly elevated levels of TANs (Figures 4H, 4L–4N, and S3E). Together, these results demonstrate that NKX2–1 suppresses TAN recruitment independent of SOX2 induction or squamous histotype.

SOX2 and NKX2–1 Inversely Regulate the Neutrophil Chemoattractant *Cxcl5*

We sought to define the mechanisms by which SOX2 and NKX2–1 regulate neutrophil recruitment. GSEA suggested that SL tumors have a positive enrichment for cytokine and chemokine signatures compared to KP tumors (Figure S4A). Expression of genes associated with neutrophil recruitment such as *Cxcl2, Cxcl3, Cxcl5* and *Ppbp* (*Cxcl7*) was increased in SL and LP tumors compared to KP tumors (Figure 5A). We employed an unbiased comprehensive approach to determine whether SOX2 and NKX2–1 converge on the regulation of any of these chemokines. First, we identified SOX2 binding sites in the genomes of LP tumors and KPS cell lines by ChIP-seq. We chose LP tumors because they represent a neutrophil-rich squamous tumor with physiological levels of SOX2 expression, and KPS samples because they recapitulate SOX2 function in TAN recruitment independent of squamous histotype. To generate a similar dataset for NKX2–1 targets, we used published

NKX2–1 ChIP-seq data from K adenocarcinomas and gene expression data comparing K and *Kras*$^{LSL\text{-}G12D/+}$;*Nkx2–1*$^{fl/fl}$ (KN) tumors (Snyder et al., 2013). ChIP-seq analyses identified binding motifs for SOX2 and NKX2–1 that were similar to published studies (Maeda et al., 2012; Watanabe et al., 2014) (Figure S4B). SOX2 and NKX2–1 bound some of the same genes, while each factor also bound unique targets (Figure 5B). We integrated ChIP-seq and gene expression data for each dataset to identify genes that were bound by and transcriptionally activated or suppressed by SOX2 and NKX2–1. Genes that were transcriptionally regulated by both SOX2 and NKX2–1 were mostly regulated inversely (Figure 5C), such that SOX2-induced genes were repressed by NKX2–1, and NKX2–1-induced genes were suppressed by SOX2. In contrast, many fewer genes were induced (or repressed) by both SOX2 and NKX2–1. These data suggest that SOX2 and NKX2–1 have unique genomic targets and inversely regulate transcription of many genes in lung cancer.

SOX2 promotes TAN accumulation whereas NKX2–1 inhibits neutrophil accumulation in tumors, so we focused on genes that were induced by SOX2 and/or repressed by NKX2–1, or both. *Cxcl5* was the only known neutrophil chemo-attractant gene bound by either transcription factor in this category. Binding sites for both SOX2 and NKX2–1 could be identified in the promoter and first exon of *Cxcl5* (Figure 5D). *Cxcl5* mRNA expression was significantly elevated in SL and LP tumors compared to KP tumors and normal lung tissue, and in KN versus K tumors (Figure S4C). *CXCL6*, the human homolog of mouse *Cxcl5* (Figures S4D and S4E), was significantly elevated in human LSCC compared to LADC (Campbell et al., 2016) (Figure S4C). *CXCL6* was one of only two CXCR2 ligands that significantly correlated with squamous subtype and high *SOX2* and low *NKX2–1* expression in a large collection of human NSCLC samples from TCGA (Figures 5E and S4C and S4F).

Consistent with the gene expression data, SOX2^{+} KPS tumors and KPN tumors had significantly higher CXCL5 protein levels compared to control SOX2^{-} KPS, KPG and KP tumors (Figures 5F and 5G). Moreover, SL and LP squamous tumors had high levels of CXCL5, whereas KP adenocarcinomas completely lacked it. Together, these findings suggest that SOX2 and NKX2–1 inversely regulate *Cxcl5* expression. To further test this, we overexpressed SOX2 in a panel of human lung cancer cell lines. Exogenous SOX2 was sufficient to induce CXCL6 in five of six cell lines (Figure S4G). In contrast, altering NKX2–1 levels by overexpression or CRISPR-mediated loss was not sufficient to alter CXCL56 levels in vitro (Figures S4H and S4I). Thus, it appears that NKX2–1 regulation of CXCL6 differs in vitro in human cells compared to mouse tumors.

To determine whether CXCL5 expression is sufficient to promote neutrophil recruitment, we infected KP mice with Lenti-Cxcl5-Cre or Lenti-GFP-Cre viruses and analyzed lungs 3–4 months post-infection. Similar to our observations with Lenti-Sox2-Cre, there was an occasional uncoupling of Cre and CXCL5 expression as some tumors lacked high CXCL5 expression (Figure 5H). KP-CXCL5^{+} (KPC5) tumors demonstrated high expression of CXCL5 and significantly increased TAN accumulation (Figure 5H). When tumors with and without CXCL5 were detected near each other in the KPS or KPC5 models, only CXCL5^{+} tumors harbored neutrophils—suggesting that CXCL5' s impact on neutrophil recruitment is relatively local (Figure S4J). These data demonstrate that tumor-derived CXCL5 expression is sufficient to promote TAN accumulation in lung cancer.

Loss of NKX2–1 Dramatically Accelerates Squamous Lung Tumorigenesis

To better understand the functions of SOX2 and NKX2–1, we performed pathway analyses using Enrichr on genomic targets that were transcriptionally regulated in vivo (Figure 5C and Table S4). SOX2 induced expression of its own target genes, as well as genes bound by the transcriptional regulators TP63 and NFE2L2, which are known oncogenic transcription factors in LSCC (Campbell et al., 2016). SOX2-induced genes included those related to squamous cell fate (*Krt6a* and *Krt19*), ciliated cell fate (*Foxj1*), and TP63 target gene (*Perp*) (Table S4 and Figure S5A). SOX2-repressed genes overlapped with genes bound by chromatin regulators including EZH2, which is enriched in squamous lung tumors (Zhang et al., 2017). SOX2-repressed genes included regulators of gastric and liver fate such as *Hnf4a*, *Foxa2* and *Vil1* (Maeda et al., 2012). In contrast, NKX2–1 positively regulated genes involved in lung and alveolar cell fate including *Sftpa1*, *Stfpb* and *Sftpc*. NKX2–1 inhibited expression of genes involved in gastrointestinal differentiation consistent with our previous findings (Snyder et al., 2013). NKX2–1-repressed genes overlapped with known SOX2 genomic targets, as well as gastrointestinal cell fate genes including *Foxa1*, *Perp*, *Lgals2*, *Lgals4*, *Krt20*, *Vil1*, *Cdh17* and mucous metaplasia gene, *Spdef* (Figures S5A and Table S4).

SL tumors have reduced but not completely absent NKX2–1 expression, whereas human LSCCs tend to be completely negative (Figure 4E). Given that NKX2–1 repressed the expression of genes that are induced by SOX2, we speculated that complete loss of NKX2–1 may facilitate SOX2-driven tumorigenesis. To address this hypothesis, we generated $Rosa26^{LSL\text{-}Sox2\text{-}IRES\text{-}GFP};Nkx2\text{–}1^{fl/fl};Lkb1^{fl/fl}$ (SNL) mice. We infected SNL and SL mice with Ad5-CMV-Cre using intratracheal inhalation and monitored mice for tumor formation by microCT imaging. We observed tumors in SNL mice as early as 8 weeks after tumor initiation (Figure S5B). We then subjected a large cohort of mice to histopathologic review at 16 weeks post-infection. All SNL mice (n = 10) harbored multifocal neoplasia, including mucinous adenocarcinomas, adenosquamous and squamous cell carcinomas (Figures 6A–B and S5C). In contrast, only one SL mouse (n = 25) harbored a single LSCC at this time point. In SNL mice, both mucinous and squamous components expressed SOX2 and GFP and lacked NKX2–1 (Figures 6B and S5D). SOX2 intensity was heterogeneous in the squamous lesions, with basal-like cells expressing higher levels than more highly keratinized tumor cells. SOX2 levels were uniformly high in adenocarcinoma components, in concordance with a similar observation in KL mice (Zhang et al., 2017). DNp63 and KRT5 were robustly expressed in the squamous components of the SNL tumors and were absent in the mucinous adenocarcinoma cells, which instead expressed markers of gastric differentiation including HNF4A, LGALS4 and CTSE (Figures 6B and S5D). This is consistent with our previous work showing that NKX2–1 loss leads to activation of a gastric differentiation program in LADC (Snyder et al., 2013).

Adenocarcinomas arising in KL mice undergo transdifferentiation to a squamous differentiation state over time (Li et al., 2015). To determine whether a similar phenomenon occurs in SNL mice, we analyzed lung tumors at 4, 8, and 12 weeks post-infection. At four weeks, we identified multifocal mucinous adenocarcinoma in the lungs of all mice (n = 6), whereas $DNp63^{+}$ tumor cells were found only in a minority (33%) of mice (Figure 6C). The proportion of mice with $DNp63^{+}$ tumor cells increased over time, such that by 12 weeks

post-infection DNp63$^+$ tumor cells could be identified in the lungs of all SNL mice. The relative quantity of DNp63$^+$ tumors in each mouse increased over time as well (Figure 6D). Taken together, these data show that loss of NKX2–1 dramatically accelerates tumorigenesis driven by SOX2 and loss of LKB1. These data suggest that alteration of these three genes induces mostly if not entirely mucinous adenocarcinoma lesions that undergo transdifferentiation to LSCC over time.

We sought to utilize this LSCC model to further investigate TAN regulation. TANs were abundant in all SNL tumors regardless of histotype although squamous tumors had subtly but significantly higher levels of CD11B and LY6G (Figures 6E and6F). Flow cytometry from SNL and KP tumors confirmed that squamous tumors exhibit increased TANs but fewer macrophages compared to adenocarcinomas (Figures 6G–H and S5E-F). SNL tumors with squamous and adenocarcinoma histotypes both exhibited high CXCL5 protein levels (Figures 6E and 6F). Furthermore, ~80% of SNL tumor cells expressed CXCL5, whereas < 10% of CD45$^+$ immune cells from tumor-bearing or normal lungs expressed CXCL5 (Figure 6I and S5G). TANs within both SNL LADC and LSCC tumor regions expressed CXCR2, the receptor for CXCL5 (Figures 6J and S5H–I). Together, this suggests that CXCL5-expressing tumor cells recruit CXCR2$^+$ neutrophils through a paracrine mechanism. This finding further suggests that TAN recruitment to tumors is determined by transcriptional regulators that specify cell fate.

TANs Exhibit Pro-tumor Features and Promote Squamous Tumors

To further characterize TANs in this model, we sorted CD45$^+$CD11B$^+$GR1$^+$ cells and examined their cytology (Figure S6A). Many TANs exhibited a circular banding phenotype reminiscent of N2 TANs (Fridlender et al., 2009). We analyzed CD45$^+$CD11B$^+$GR1$^+$ cells from tumor-bearing lungs and the blood of SNL animals for the recently identified N2-marker SiglecF (Engblom et al., 2017). Approximately 40% of TANs, but not peripheral blood neutrophils (PBNs), expressed SiglecF (Figures 7A and S6B). SiglecF$^+$ neutrophils exhibit multiple pro-tumor functions including increased reactive oxygen species (ROS) production and ability to foster tumor progression (Engblom et al., 2017). To determine whether SNL TANs exhibit pro-tumor functions, we measured the ability of normal neutrophils (NNs) from healthy control lungs, PBNs and TANs to produce intracellular ROS. TANs exhibited significantly more ROS than both NNs and PBNs by ~3–4 fold (Figures 7B and S6C).

As TANs exhibited cytological heterogeneity and only a subset expressed SiglecF, we sought to explore TAN heterogeneity using single cell RNA-seq (scRNA-seq). We sorted CD45$^+$CD11B$^+$LY6G$^+$ cells from the peripheral blood of tumor-bearing SNL animals (i.e. PBNs) or lung tumors (i.e. TANs) from the same animals (Figure S6D). We used the Chromium drop-seq platform (10x Genomics) to obtain transcriptomes from PBNs (n = 1744 cells) or TANs (n= 519 cells). Chromium-derived transcriptional profiles were visualized using t-distributed stochastic neighbor embedding (tSNE) (van der Maaten and Hinton, 2008). Cell Ranger clustering called nine distinct clusters, seven within the PBNs and two among the TANs (Figure 7C). *Cxcr2* and house-keeping genes *Actb* and *B2m* were expressed in most clusters regardless of the source of neutrophil. Although clusters 8 and 9

were negative for *Cxcr2*, they expressed *S100a8/9*, suggesting that these cells were also neutrophils (Figure S6E). Consistent with the flow cytometry data, *SiglecF* expression was largely confined to the TAN clusters (Figure 7C). GSEA revealed enrichment of the SiglecF-high versus SiglecF-low neutrophil gene signature (Engblom et al., 2017) in the SNL TAN population (Figure S6F). Furthermore, many genes associated with N2 neutrophils such as *Arg1*, *Ccl3*, *Ccl4* and *Csf1* had significantly increased expression in TANs versus PBNs (Figures 7C and 7D). Pathway analyses using Enrichr on differentially expressed genes among TANs versus PBNs (Figure 7D and Table S5) revealed that TANs had increased expression of genes involved in ROS and reactive nitrogen species (RNS) production, extracellular matrix (ECM) degradation, and N2-related chemokine signaling (Figure 7E). Neutrophil degranulation genes were differentially expressed among PBNs and TANs, and PBNs were enriched for pathways related to “positive regulation of the immune system” (Figures S6E and S6G). Together, these data suggest that LSCC TANs are a heterogeneous population with at least some cells exhibiting gene expression and functional activities ascribed to N2 TANs.

Finally, we sought to determine the function of TANs in the SNL model using a neutrophil depletion strategy (Fridlender et al., 2009; Steele et al., 2016). SNL mice were treated with control or anti-LY6G antibody for three weeks followed by harvest of tumor-bearing lungs. TANs were significantly reduced in anti-LY6G treated animals as determined by IHC analyses (Figures 7F and7G). The lungs of neutrophil-depleted animals had fewer squamous tumors upon inspection of H&E-stained sections. Consistent with this observation, neutrophil-depleted lungs had significantly fewer $DNp63^{+}$ tumors compared to controls (Figures 7F and7H). These findings suggest that TANs promote squamous tumorigenesis and may directly or indirectly impact tumor cell fate in this transdifferentiation model. Altogether, these data suggest that lineage specifier genes impact the tumor immune microenvironment, and the immune microenvironment in turn may impact tumor phenotype.

DISCUSSION

A distinctive feature of the LSCC TIME compared to LADC is the enrichment of neutrophils. Multiple LSCC mouse models (LP, Sox2PC, and now SL and SNL) display elevated neutrophils compared to LADC GEMMs (KP and K). It was previously postulated that histotype determines the TIME in NSCLC. Abundant keratinization in squamous tumors, for example, may molecularly resemble a persistent wound accompanied by aberrant inflammation. On the other hand, a growing body of evidence suggests that cancer cell intrinsic factors also impact the TIME (Bezzi et al., 2018; Li et al., 2018). STAT3 and NF-kB signaling have roles in inflammation and are elevated in LSCC compared to LADC (Mukhopadhyay et al., 2014; Xiao et al., 2013). A recent finding that *Egfr* and *Kras*-driven LADC mouse tumors have distinct lymphocyte compositions also suggests that oncogenic drivers may dictate the TIME (Busch et al., 2016). Our findings here reveal tumor-intrinsic mechanisms whereby lineage-specific tumor drivers (SOX2 and NKX2–1) inversely regulate TAN accumulation.

Multiple chemokines can act upon the CXCR2 neutrophil receptor to induce neutrophil chemotaxis (Zlotnik and Yoshie, 2012). *Cxcl3*, *−5*, and *−7* have increased expression in SL

and LP tumors compared to KP and K tumors. Although the exact mechanism of transcriptional regulation requires further study, unbiased ChIP and RNA-seq data from GEMM tumors revealed that *Cxcl5* is a direct genomic target of SOX2 and NKX2–1. We found that SOX2 overexpression or NKX2–1 loss is sufficient to induce CXCL5 in tumor cells in vivo, and tumor-associated CXCL5 is sufficient to recruit TANs. It is likely that other neutrophil recruitment chemokines are indirectly regulated in these lung cancer subtypes and may also be sufficient for TAN recruitment. Future studies employing *Cxcl5* genetic deletion and/or antibody-mediated CXCL5 depletion will be necessary to determine whether CXCL5 is required for TAN recruitment. Our data suggest CXCL5 regulation is governed by lineage specifiers rather than tumor histotype since we observe CXCL5 expression and TAN influx in adenocarcinomas when either SOX2 or NKX2–1 is altered. These data are consistent with previous studies in KL mice where TANs are enriched in LSCCs as opposed to LADCs. In the KL model, SOX2 and CXCL5 expression levels are both increased while NKX2–1 levels are decreased during adeno-to-squamous transdifferentiation (Nagaraj et al., 2017; Zhang et al., 2017), consistent with the mechanisms proposed here. *CXCL6*, the human homolog of mouse *Cxcl5*, is positively correlated with *SOX2* and negatively correlated with *NKX2–1* expression in human NSCLC, suggesting the mechanisms described here likely function in human tumors.

Importantly, we found that NKX2–1, which has pro- and anti-oncogenic activities in LADC, is a tumor suppressor in LSCC. In SL and LP mice, the peripheral location of squamous tumors and the detection of smaller adenosquamous lesions at early time points hinted that these tumors originate from distal lung epithelium. Studies in Sox2PC mice showed that in addition to basal cells, club cells and alveolar type 2 cells can give rise to squamous tumors (Ferone et al., 2016). SOX2 overexpression (either genetically or naturally acquired during tumorigenesis) is common to all three squamous GEMMs. NKX2–1 levels are reduced in squamous GEMMs, but are not entirely absent, in contrast to complete NKX2–1 loss in human LSCC. These findings suggest that squamous tumors in GEMMs are primarily initiated in distal epithelium and transdifferentiate to squamous fate over time, perhaps during their long latency. It seems likely that all LSCCs must downregulate NKX2–1 to adopt a squamous fate, whether they arise from basal cells in the proximal airway, or in the distal lung where peripheral-type LSCCs arise. Consistent with this hypothesis, *Nkx2–1* deletion significantly accelerated LSCC development with evidence of transdifferentiation. ChIP and RNA-seq data suggested that NKX2–1 repressed known SOX2 target genes, so we speculate that this provides a mechanism by which NKX2–1 loss facilitates SOX2-driven transformation. An independent study recently found that *Sox2* expression and *Nkx2–1* deletion in the mouse lung (i.e. SN) cooperate to promote squamous lung cancer whereby NKX2–1 loss alters SOX2 occupancy in the genome (Tata et al., 2018). Similar to our observations in SNL mice, SN cells in that study may transdifferentiate in organoid cultures, including in the absence of stromal cells. If confirmed in vivo, this may suggest that neutrophils promote squamous tumorigenesis but are not required for transdifferentiation. Future studies should address how SN and SNL tumor development may differ in terms of latency, tumor cell fate and the TIME. While the Tata et al. study did not investigate the TIME, our data predict that SN tumors will have elevated CXCL5 and TANs, as in SNL mice. Tata et al. suggest that airway cells are more permissive than alveolar cells for

squamous transformation, so additional studies are warranted to determine how the cell of origin impacts tumor cell fate upon *Nkx2–1* loss.

Our scRNA-seq data suggest that TANs are fundamentally different from PBNs and acquire pro-tumor features such as increased ROS activity, and elevated expression of genes that block T cell activity and promote ECM degradation. Since the scRNA-seq data presented here are limited to small numbers of cells, additional studies will be required to fully decipher the molecular and functional heterogeneity among TANs. The neutrophil depletion experiments demonstrated that TANs preferentially promote squamous tumors. This observation is intriguing as it suggests a bidirectional cross-talk between tumor cells and immune cells. The underlying mechanism by which TANs impact tumor cell fate remains elusive as TANs may either create a favorable TIME for preexisting squamous cancer cells over adenocarcinoma cells, or may accelerate adeno-to-squamous transdifferentiation. Since ROS and hypoxia can promote adeno-to-squamous transdifferentiation (Han et al., 2014; Li et al., 2015), and both are induced by neutrophils (Campbell et al., 2014; Coffelt et al., 2016), future studies should test the role of TAN-induced ROS and hypoxia in this phenomenon.

Multiple studies have indicated that relapsed LADC following targeted therapy or chemotherapy can transition to LSCC, although the mechanisms are still poorly understood (Hou et al., 2016). Understanding the heterogeneity and plasticity of the TIME and how tumor genotype shapes the TIME is crucial for effective therapy. Here we find that the genetic mechanisms that regulate adeno-to-squamous transdifferentiation also shape the TIME, and that in turn, immune cells can impact tumor phenotype. Future studies are warranted to determine how TANs and other immune cell types evolve during tumor transdifferentiation, and how these fluctuations may affect response to immunotherapy.

STAR METHODS

Key Resources Table

Submitted as a separate file.

Contact for Reagent and Resources Sharing

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Trudy G. Oliver (trudy.oliver@hci.utah.edu).

Experimental Models and Subject Details

Mice

Mice were housed in an environmentally controlled room and all experiments were performed in accordance with University of Utah's Institutional Animal Care and Use Committee. $Lkb1^{fl/fl}$ mice were purchased from The Jackson Laboratory. Mice harboring $Kras^{LSL\text{-}G12D/+}$ (Jackson et al., 2001), $Trp53^{fl/fl}$ (Jonkers et al., 2001), $Nkx2\text{–}1^{fl/fl}$ (Kusakabe et al., 2006), $Kras^{FSF\text{-}G12D/+}$ (Young et al., 2011), $Trp53^{FRT/FRT}$ (Lee et al., 2012), $Rosa26^{FSF\text{-}CreERT2}$ (Schönhuber et al., 2014), $Pten^{fl/fl}$ (Zheng et al., 2008) alleles have been

previously described. *Sox2*$^{LSL/LSL}$ mice were generated using a standard approach: the full-length cDNA of *Sox2* was placed in the *Rosa26* targeting vector behind a floxed stop and neomycin resistance cassette followed by *eGfp* on an internal ribosomal entry site. The targeting vector for the *Rosa26-Lox-Stop-Lox-Sox2-IRES-GFP* allele was electroporated into mouse R1 embryonic stem cells and the G418-resistant clones were screened by long-range PCR for the correct 5' and 3' insertion. The correct clones were then subjected to Southern blot analysis to identify insertion number and size. Cells from one clone with a confirmed target were microinjected into C57BL/6 blastocysts to generate chimeric mice. The chimeric mice were mated to C57BL/6 females and their agouti offspring were tested by PCR to confirm germline transmission of the conditional allele. All *Rosa26* targeted mice were genotyped with the following primers using tail-tip derived DNA: Forward 5'-GTTATCAGTAAGGGAGCTGCAGTGG-3', Reverse-targeted 5'-AAGACCGCGAAGAGTTTG TCCTC-3', Reverse-wildtype 5'-GGCGGATCACAAGCAATAATAACC-3'. These primers yield a 300 base pair (bp) band for the targeted locus and a 415 bp band for wild type locus. The PCR conditions were set to 95°C for 30 seconds, 59°C for 45 seconds and 72°C for 30 seconds, for 30 cycles.

At 6–8 weeks of age, anesthetized mice were infected with 10^7–10^8 plaque-forming units of Ad5-CMV-Cre viruses (University of Iowa Cat#VVC-U of Iowa-5) or Ad5-CMV-Flp (University of Iowa Cat#VVC-U of Iowa-530) or with ~10^7 infectious units/ml of Lenti-Sox2-Cre, Lenti-Gfp-Cre, or Lenti-Cxcl5-Cre lentiviruses by intranasal or intratracheal instillation. Viruses were administered in a Biosafety Level 2+ room according to Institutional Biosafety Committee guidelines. At six weeks post-infection, KPN mice were treated with vehicle or tamoxifen for four weeks to induce *Nkx2–1* deletion. Tamoxifen (Sigma Cat#T5648) was dissolved in corn oil with a final concentration of 20 mg/ml and administered at a dose of 120 mg/kg per day for 6 days. Additionally, mice were fed ad libitum with 500 mg/kg tamoxifen-supplemented chow (Envigo Cat#TD.130858) in place of standard chow for the duration of the experiment. Both male and female mice were equally segregated for all experiments.

MEFs

MEFs were isolated from SL embryos on day E13.5 and cultured in DMEM, 10% FBS, 1% Pen/Strep, 1% L-Glutamine, and 50 μM 2-Mercaptoethanol. Dissected tissues were incubated in 5 ml 2X trypsin at 37 °C for 5 minutes. Single cell suspension was collected and transferred into 50 ml tubes with 25 ml pre-warmed MEF medium. After centrifugation at 1500 RPM for 5 minutes, pelleted cells were resuspended in 10 ml MEF media and transferred into tissue culture plates. Media was replaced 24 hours later.

Human Cell Lines

Human lung cancer cell lines were purchased from ATCC and grown in either DMEM, 10% FBS, 1% Pen/Strep, 1% L-Glutamine (A549, H2009) or RPMI with 10% FBS, 1% Pen/Strep, 1% L-Glutamine (H1650, H2087, H3122, HCC827, H125, H157). HEK293T cells were transfected with Lenti-TetOn plasmid with a human SOX2 coding sequence, in addition to pCMV-VSV-G (Addgene Cat#8454) and pCMV-dR8.2 dvpr (Addgene Cat#8455) plasmids using TransIT-LT1 transfection reagent (Mirus Bio Cat#MIR2305).

Viruses were collected at 48 and 72 hr post-transfection. Cells were infected twice with viral supernatant and 8 μg/ml polybrene (Santa Cruz Cat#sc-134220) and selected with blastacidin (VWR Cat#71002–676) (10 μg/ml for A549 and H3122, 8 μg/ml for H2009, 6 μg/ml for H1650, 4 μg/ml for HCC827, and 2 μg/ml for H2087) until all uninfected control cells treated in parallel were killed. Cells were treated with or without 0.5 μM of doxycycline for 48 hr before collecting and analyzing. For NKX2–1 overexpression, cells were transiently transfected with pMSCVpuro plasmid with or without human NKX2–1 coding sequence using TransIT-LT1 transfection reagent and collected at 48 hr for analysis. For NKX2–1 genetic deletion, lentiCRISPR v2 plasmid (Addgene Cat#52961) with the target sequence of "AAGAAAGTGGGCATGGAGGG" or a non-targeting control sequence were generated. Cells were infected with viruses for sgNKX2–1 or sgControl and selected with puromycin (Thermo Fisher Cat#BP2956–100) (2 μg/ml for all cell lines) until all uninfected control cells treated in parallel were killed.

Human Lung Cancer Tissue

For human LSCC tumor tissue, excess de-identified fresh tissue was obtained with prior patient consent under an approved protocol by the Institutional Review Board (#10924). Intermountain Biorepository, Intermountain Healthcare (Salt Lake City, UT) provided tissue microarray (TMA) of human LSCC. Institutional guidelines regarding specimen use were followed.

Method Details

MicroCT Imaging

At indicated time points, mice were scanned for 30 s to 2 min under isoflurane anesthesia using a small animal Quantum FX microCT (PerkinElmer) at 45 μm resolution, 90 kV, and 160 μA current. Images were acquired using PerkinElmer Quantum FX software and processed with Analyze 11.0 (AnalyzeDirect).

Ex Vivo Imaging

An Olympus MVX10 in vivo imaging system with a DP72 CCD camera (Olympus Corp) was used for ex vivo imaging of mouse lungs. High-resolution images were captured with 1/500 second exposure time for brightfield images and 1/2 second exposure time for fluorescence images. Images were analyzed with Olympus cellSens Dimension 1.15 software (Olympus Corp).

Lentivirus Production

GFP, murine *Sox2*, and murine *Cxcl5* cDNAs were cloned into bicistronic lentiviral vectors with an *Actin* promoter driving *Cre* and *Pgk* promoter driving cDNA expression to generate Lenti-GFP-Cre, Lenti-Sox2-Cre (Addgene #59019), and Lenti-Cxcl5-Cre (Addgene #110278) plasmids. Plasmids were confirmed by direct sequencing. 293T cells were transfected with a three-plasmid transfection system including the lentiviral vector, pCMV-dR8.2 dvpr (Addgene Cat#8455) and pCMV-VSV-G (Addgene Cat#8454). Viruses were harvested at 48 and 72 hours post-transfection, concentrated by ultracentrifugation (24,000 x g) and titered using HEK293T reporter cells stably expressing a *Lox-DsRed2(Stop)-Lox-*

GFP cassette. The number of cells expressing GFP was measured by flow cytometry and used to calculate titer (infectious U/ml).

Immunohistochemistry

Tissues were fixed in 10% neutral buffered formalin for 24 hrs, washed in PBS and transferred to 70% ethanol. Formalin-fixed paraffin embedded (FFPE) sections at 4–5 uM were dewaxed, rehydrated and subjected to high-temperature antigen retrieval, 20 min boiling in a pressure cooker in 0.01 M citrate buffer, pH 6.0. Slides were blocked in 3% H_2O_2 for 15 min, blocked in 5% goat serum in PBS/0.1% Tween-20 (PBS-T) for 1 hr, and stained overnight in blocking buffer with primary antibodies. An HRP-conjugated secondary antibody (Vector Laboratories) was used at 1:200 dilution in PBS-T, incubated for 45 min at room temperature followed by DAB staining (Vector Laboratories). All staining was performed with Sequenza cover plate technology. Tris-buffered saline with Tween-20 (TBS-T) was used instead of PBS-T for phosphoprotein antibodies. SignalStain Boost IHC Detection Reagent (Cell Signaling Cat#8114) was used for phospho-4EBP1 staining. The primary antibodies include: SPC (Millipore Cat#AB3786) 1:2000; NKX2–1 (Abcam Cat#ab76013) 1:250; KRT5 (BioLegend Cat#PRB-160P) 1:2000; SOX2 (Cell Signaling Cat#3728) 1:250; phospho-4EBP1 (Thr37/46) (Cell Signaling Cat#2855) 1:800; DNp63 (Novus Cat#NBP2–29467) 1:200; CD3 (Abcam Cat#5690) 1:100; CD11B (Abcam Cat#ab133357) 1:2000; MPO (Novus Cat#NBP1–42591) 1:250; FOXP3 (eBioscience Cat#14–5773) 1:100; CXCL5/6 (Abcam Cat#198505) 1:200; LY6G (BioXCell Cat#BE0075–1) 1:1500; CXCR2 (R&D Systems Cat#MAB2164) 1:500; HNF4A (Cell Signaling Cat#3113) 1:500; GFP (Cell Signaling Cat#2956) 1:250; LGALS4 (R&D Systems Cat#AF2128) 1:200; CTSE (Lifespan Cat#LS-B523) 1:12000. For human SOX2 IHC, SOX2 (Millipore Cat#MAB4423) antibody was used with 1:250 dilution.

For manual quantification, digital images of stained tissues were captured by Zeiss Axio Scope.A1 microscope using AxioVision SE64 software. IHC score (a.k.a. “H score”) was calculated by multiplying the percentage of positive cells (P; 0– 100%) by the intensity (I; 0–3). Formula: S = P x I; Range = 0–300. For immune cell markers where signal intensity is not a factor of consideration, percent positive cell score was calculated as: tumors with less than 1% positive cells were score = 0; tumors with positive cells between 1% - 3% were score = 1; tumors with positive cells between 3% - 6% were score = 2; tumors with more than 6% positive cells were score = 3.

Immunoblot

Total protein lysates were prepared as previously described, separated via SDS-PAGE and transferred to a PVDF membrane. Membranes were blocked for 1 hr in 5% milk or 5% BSA, followed by overnight incubation with primary antibodies at 4°C. Membranes were washed for 3 × 15 min at room temperature in TBS-T. Mouse and rabbit HRP-conjugated secondary antibodies (Jackson ImmunoResearch, 1:10,000) were incubated for 1 hr at room temperature followed by washing 3 × 15 min at room temperature in TBS-T. For detection, membranes were exposed to Advansta WesternBright ECL HRP Substrate Kit (VWR #490005–020) and detected on Hyblot CL film (Denville Scientific Inc). Primary antibodies: SOX2 (Millipore Cat#MAB4423) 1:2000; ACTIN (Sigma Cat#A2066) 1:10,000; NKX2–1

(Abcam Cat#ab76013) 1:2000; CXCL5/6 (Abcam Cat#ab198505) 1:1000; HSP90 (Cell Signaling Cat#4877) 1:1000. ACTIN or HSP90 were used as loading controls. Densitometry analyses were performed with Image Studio Lite software. The signal value of each sample was normalized by its loading control value. Then the normalized value of a control sample was set to 1 arbitrary unit while each test sample value was presented as a relative intensity compared to control.

RNA Isolation and RNA-Sequencing

5 μm serial sections of FFPE tissue were stained with H&E and Aniline blue. H&E stained slides were used to confirm histology of tumors and guide bordering consecutive serial sections of Aniline blue stained tissue samples. Bordered tissue samples were microdissected using sterile disposable scalpels and razor blades with the assistance of a dissection scope. 10 – 14 pieces of each sample isolated from serial sections were pooled in a microcentrifuge tube containing 100% ethanol. RecoverAll™ Total Nucleic Acid Isolation Kit for FF PE (Thermo Fisher Cat#AM1975) was used to isolate RNA from FFPE tissues according to the manufacturer's protocol. RNA samples were quantified using NanoDrop 8000 and Qubit RNA HS Assay. RNA integrity numbers were calculated on Agilent HS RNA ScreenTape. RNA was subjected to library construction with the Illumina TruSeq Stranded mRNA Sample Preparation Kit (Cat#RS-122–2101, RS-122–2102) according to manufacturer's protocol. Chemically denatured sequencing libraries (25 pM) are applied to an Illumina HiSeq v4 single read flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq SR Cluster Kit v4-cBot (Cat#GD-401–4001). Following the transfer of the flowcell to an Illumina HiSeq 2500 instrument (HCSv2.2.38 and RTA v1.18.61), a 50-cycle single-read sequence run was performed using HiSeq SBS Kit v4 sequencing reagents (Cat#FC-401–4002). Mouse mm10 annotations (Ensembl build 82) were used in the RSEM (v1.2.12) utility rsem-prepare-reference to create bowtie (v1.0.1) indices. Gene expression was determined using the RSEM utility rsem-calculate-expression with the forward strand probability set to zero. Differential expression was determined using EBSeq (v1.4.0) using 'MedianNorm' function to calculate size factors andsetting 'maxround' to 10. To adjust for transcript length, fragments per kilobase per million reads (FPKM) were calculated for all genes and log2-transformed after addition of a small constant (0.01).

Chromatin Immunoprecipitation Sequencing (ChIP-seq)

20 million cells per ChIP were cross-linked in 1% formaldehyde for 10 min at room temperature. Crosslinking was stopped with 125 mM glycine and nuclei were extracted. Alternatively, flash-frozen tumors were directly used. Chromatin was sonicated using an Epishear Probe Sonicator (Active Motif) for 4 min at 40% power. SOX2 antibody (R&D Systems Cat#AF2018) was used for IP and an input sample for each cell line served as the control. Libraries were sequenced on an Illumina HiSeq 2500 as single-end 50 bp reads to a minimum depth of 35 million reads per sample. Reads were aligned to the mm10 build of the mouse genome with bowtie using the following parameters: -m 1 -t --best -q -S -l 32 -e 80 -n 2. Peaks were called with MACS2 using a p value cutoff of 10^{-10} and the mfold parameter bounded between 15 and 100. For visualization, MACS2 produced bedgraphs with the –B and –SPMR options. The Integrative Genomics Viewer (IGV) software was

used to visualize binding sites. NKX2–1 ChIP-seq data from K adenocarcinomas and exon array data comparing K and KN tumors were obtained from previously published data (Snyder et al., 2013).

Flow Cytometry

MEFs derived from *Sox2*$^{LSL/LSL}$;*Lkb1*$^{fl/fl}$ and *Sox2*$^{LSL/+}$;*Lkb1*$^{fl/fl}$ embryos were infected with 10 multiplicity of infection (MOI) of Ad5-CMV-Cre viruses (University of Iowa) at 80% confluency in reduced serum (2% FBS + 1% Penicillin/Streptomycin) DMEM. The media was replaced with complete DMEM four hours post-infection. Uninfected cells were used as a negative control. Cells were harvested 48 hours post-infection, washed 3 times with PBS, and resuspended in incubation buffer [0.5 g bovine serum albumin (BSA) (Thermo Fisher Scientific Cat#23208) in 100 ml 1X PBS]. Flow cytometric analysis was performed on an LSR model 1a (Becton Dickinson) running BD CellQuest Pro software. Quantification of data was obtained with FlowJo 10.2 (FlowJo, LLC, Ashland, OR).

Mice were euthanized to harvest tissue. Tumor or normal lung tissue was excised and kept in cold PBS with a matching tissue piece collected for histology. Tissue sample was chopped into smaller pieces mechanically and further dissociated with 1 ml digestion media at 37 C for 30 min. Digestion reaction was stopped by adding 0.5 ml quench media. Digested tissue was passed through 20-gauge needle syringe multiple times to aid dissociation. Digested samples were passed through a 100 μm cell strainer and single cells were collected into a round-bottom polystyrene tube. Cells were pelleted by centrifugation at 1500 rpm at 4 C for 5 min. Cells were resuspended in 0.5 ml ACK (Ammonium-Chloride-Potassium) lysing buffer and incubated at 37 C for 3 min to lyse erythrocytes. The lysis reaction was stopped by addition of 4.5 ml cold PBS. Blood was collected via cardiac puncture with a 1 ml syringe with a 23-gauge needle that was pre-coated with PBS + EDTA (5 mM final concentration). Cells were pelleted by centrifugation at 1500 rpm at 4 C for 5 min and processed as described above for tumor cells. Cell number and viability were measured by Countess II FL Automated Cell Counter (Thermo Fisher Cat#AMQAX1000).

For cell surface staining, single cell suspensions were stained with fluorescent conjugated antibodies for 30 min at 4 C in dark. Cells were then washed with and resuspended in FACS buffer. DAPI was used for dead cell exclusion. For intracellular staining, cells were stained with a fixable viability dye simultaneously with cell surface staining as explained above. After washing with FACS buffer, cells were resuspended in residual buffer (100 μl) and fixed with the addition of 100 μl fixation buffer for 30 min at 4 C in dark. Cells were then washed in 2 ml of 1X permeabilization buffer two times and resuspended in 200 μl permeabilization buffer. Cells were stained with unconjugated primary antibody for intracellular antigen for 30 min at 4 C in dark. Cells were then washed with permeabilization buffer and stained with fluorescent conjugated secondary antibody for 30 min at 4 C in dark. Cells were washed in permeabilization buffer and FACS buffer, followed by resuspension in FACS buffer. Data acquisition was performed with BD FACSAria Cell Sorter (BD Biosciences) running BD FACSDiva 8.0.2 software and analyzed with FlowJo 10.2 (FlowJo, LLC, Ashland, OR).

Buffers and reagents include: Digestion media: 4200 μl HBSS-free (Thermo Fisher Cat#14175), 600 μl trypsin-EDTA (0.25%) (Thermo Fisher Cat#25200–072), 600 μl

collagenase type 4 (Worthington Biochemical Cat#LS004186) from 10 mg/ml stock prepared in HBSS with calcium and magnesium (Thermo Fisher Cat#14025), 600 µl dispase (Worthington Biochemical Cat#LS02100). Quench solution: 7.2 ml Leibovitz's L15 media (Thermo Fisher Cat#11415–064), 800 µl FBS (Sigma Cat#12303C), 30 µl DNase (Sigma Cat#D4527) at 5 mg/ml in HBSS-free media. ACK (Ammonium-Chloride-Potassium) lysing buffer: (Thermo Fisher Cat#A10492). FACS buffer: 20 ml PBS + 400 µl FBS + 2mM EDTA. Fixation and permeabilization buffers: Intracellular Fixation & Permeabilization Buffer Set (eBioscience Cat#88–8824-00).

Antibodies for flow cytometry include: Fixable Viability Dye eFluor 506 (eBioscience Cat#65–0866); CD45 (30-F11), PE (eBioscience Cat#12–0451-82); CD11B (M1/70), PE-Cy7 (BD Pharmingen Cat#552850); GR1 (RB6–8C5), APC (BioLegend Cat#108412); LY6G (1A8), BV786 (BD Horizon Cat#740953); SiglecF (E50–2440), BB515 (BD Horizon Cat#564514); CXCR2 (SA044G4), PerCP/Cy5.5 (BioLegend Cat#149307); CXCL5/6 (EP13083), (Abcam Cat#ab198505); F4/80 (T45–2342), BV711 (BD Horizon Cat#565612); Goat anti-Rabbit IgG (H+L) Cross-Adsorbed Secondary Antibody, Alexa Fluor 350 (Thermo Fisher Cat#A-11046).

Single Cell RNA-Sequencing

$CD45^+CD11B^+LY6G^+$ neutrophils were flow sorted and pooled to prepare TAN single cell population (n = 9 tumors from 2 SNL mice) and PBN population (n = blood from 2 SNL mice). Without any freezing or fixation steps, the single cell populations were immediately processed. Cell suspensions were partitioned into an emulsion of nanoliter-sized droplets using a 10X Genomics Chromium Single Cell Controller and RNA sequencing libraries were constructed using the Chromium Single Cell 3' Reagent Kit v2 (10X Genomics Cat#PN-120237). Briefly, droplets contained individual cells, reverse transcription reagents and a gel bead loaded with poly(dT) primers that include a 16 base cell barcode and a 10 base unique molecular index (UMI). Lysis of the cells and gel bead enables priming and reverse transcription of poly-A RNA to generate barcoded cDNA molecules. Libraries were constructed by End Repair, A-Tailing, Adapter Ligation and PCR amplification of the cDNA molecules. Purified cDNA libraries were qualified on an Agilent Technologies 2200 TapeStation using a D1000 ScreenTape assay (Agilent Cat#5067–5582 and Cat#5067–5583). The molarity of adapter-modified molecules was defined by quantitative PCR using the Kapa Biosystems Kapa Library Quant Kit (Kapa Biosystems Cat#KK4824). Individual libraries were normalized to 10 nM and equal volumes were pooled in preparation for Illumina sequence analysis. Sequencing libraries (25 pM) were chemically denatured and applied to an Illumina HiSeq v4 paired end flow cell using an Illumina cBot. Hybridized molecules were clonally amplified and annealed to sequencing primers with reagents from an Illumina HiSeq PE Cluster Kit v4-cBot (Illumina Cat#PE-401–4001). Following transfer of the flowcell to an Illumina HiSeq 2500 instrument (HCS v2.2.38 and RTA v1.18.61), a 125 × 125 cycle paired-end sequence run was performed using HiSeq SBS Kit v4 sequencing reagents (Illumina Cat#FC-401–4003).

Evaluation of Neutrophil Morphology

Lung tumors from AdCre-infected SNL and normal lung tissues from uninfected SNL control mice were processed for single cell suspension and single cell samples were stained for neutrophil markers as described in the “Flow Cytometry” methods section. A BD FACSAria III was used for FACS. Dead cells and debris were excluded by gating for FSC, SSC and DAPI. CD45 was used as a leukocyte marker and CD11B was used as a myeloid marker. Neutrophils were defined as $CD45^{+}CD11B^{+}GR1^{+/hi}$. Isolated cells were pelleted on slides with centrifugation at 1,500 rpm for 30 min in a Shandon Cytospin 3 (Shandon Lipshaw, Inc., Pittsburgh, PA). The cells were then stained with Diff-Quik stain and analyzed by light microscopy at 40X.

Ex Vivo ROS Activity

Lung tumors, healthy lung tissues, and peripheral blood samples were processed to generate single cell suspensions and cells were stained for cell surface markers with fluorescent conjugated antibodies as detailed in the “Flow Cytometry” methods. Following antibody incubation, cells were washed and resuspended in PBS-EGG buffer. Dihydrorhodamine 123 (Thermo Fisher Cat#D23806) was added to cells in PBS-EGG buffer with a final concentration of 0.5 μM and cells were incubated at 37 C for 30 min. Cells were then washed and resuspended in FACS buffer and were immediately analyzed in the FITC channel for ROS indicator. Cells treated with 200 μM hydrogen peroxide served as a positive control. Buffers include: FACS buffer: HBSS without Ca^{2+} or Mg^{2+} (Thermo Fisher Cat#14175095), 0.1% BSA (VWR Cat#97061–416). PBS-EGG buffer: PBS without Ca^{2+} or Mg^{2+} (Thermo Fisher Cat#10010–031), 1 mM EDTA (Thermo Fisher Cat#15575020), 0.05% gelatin and 0.09% glucose.

Neutrophil Depletion

SNL mice were scanned by microCT imaging regularly beginning 8 weeks post-infection. Mice with confirmed tumor growth were randomly assigned to cohorts and treated with 200 μg of InVivoMAb anti-mouse LY6G Clone 1A8 (BioXCell Cat#BE0075–1) or IgG1 control antibody diluted in 1X sterile PBS, pH 7.4 (Gibco Cat#10–010-031) via intraperitoneal injections three times a week for 3–4 weeks.

Bioinformatic Analyses

For cross-species signature analyses, Illumina HiSeq 2500 50-cycle single-end reads from mouse and human samples were mapped to the corresponding UCSC mm9 mouse or hg19 human genome build (http://genome.ucsc.edu) using RSEM v1.2.12. Raw estimated expression counts were upper-quartile normalized to a count of 1000. Non-coding RNA transcripts were dropped from consideration. Within the human and mouse datasets, genes for which the upper quartile of normalized expression counts across all samples fell below a threshold of 20 were considered to be lowly expressed and were filtered out. Mouse gene names were translated to human equivalents using mouse-human orthology assignments from the Mouse Genome Informatics Consortium (http://www.informatics.jax.org), and per-sample centered and scaled normalized counts were aggregated across species for the intersection of expressed genes (n=13,166 genes). Given the complexity of the dataset in

terms of a mixture of species, tissue-types, and genotypes, a high-resolution signature discovery approach (Independent Component Analysis, ICA) was employed to characterize global gene expression profiles as described previously (Papagiannakopoulos et al., 2016). This unsupervised blind source separation technique was used on the combined human-mouse discrete count-based expression dataset to elucidate statistically independent and biologically relevant signatures. ICA is a signal processing and multivariate data analysis technique in the category of unsupervised matrix factorization methods. Conceptually, ICA decomposes the overall expression dataset into independent signals (gene expression patterns) that represent distinct signatures. High-ranking positively and negatively correlated genes within each signature represent gene sets that drive the corresponding expression pattern (in either direction). Each signature is two-sided, allowing for identification of up-regulated and down-regulated genes across sample sets. Formally, utilizing the cross-species genes-samples matrix, ICA uses higher order moments to characterize the dataset as a linear combination of statistically independent latent variables. These latent variables represent independent components based on maximizing non-gaussianity and can be interpreted as independent source signals that, combined together, form the dataset under consideration. Each component includes a weight assignment to each gene that quantifies its contribution to that component. Additionally, ICA derives a mixing matrix that describes the contribution of each sample towards the signal embodied in each component. This mixing matrix can be used to select biologically relevant signatures from components with distinct gene expression profiles across the set of samples. The R implementation of the core JADE algorithm (Joint Approximate Diagonalization of Eigenmatrices) was used along with other R packages and custom utilities. A biologically relevant signature was derived from this analysis, identifying genes with a differential expression pattern between GEMM KP tumors plus normal mouse lung tissue (group-1) and human LSCCs plus non-adenocarcinoma GEMM tumors (group-2). Genes from this signature with |z-score| > 3 were selected as driver genes and were further filtered to retain those genes where |delta-expr| > 1 (where delta-expr is the difference between the average scaled expression values of group-1 and group-2 defined above). The resulting genes formed the “differential expression signature” and comprised 144 genes up-regulated and 10 genes down-regulated in group-2 (Table S1). These genes were used for semi-supervised clustering of the samples using a Pearson correlation-based distance metric (dendrogram and heatmap in Figure 2A). All analyses were conducted in the R Statistical Programming language (http://www.r-project.org/). Heatmaps were generated using the Heatplus package in R.

For unclustered expression data, transcripts with a Log2FC > 1 and an adjusted p-value < 0.01 were visualized as a heatmap using Heatmapper web server. For TCGA gene expression correlation, UCSC Xena Browser (https://xenabrowser.net) was used to access TCGA Lung Cancer cohort for which expression levels (polyA+IlluminaHiSeq RNA sequencing data) of selected genes were visualized as a heatmap. The downloaded data set was analyzed for Pearson correlation using GraphPad Prism 7 software.

Gene set enrichment analysis (GSEA) was performed using GSEA v2.2.4 software with default parameters, inclusion gene set size between 15 and 500, and the phenotype permutation at 1,000 times. Gene sets that met the false discovery rate 0.25 criterion were considered. Normalized enrichment scores (NES) and p values are shown below each

respective GSEA plots in the figures. A catalog of functional gene sets from Molecular Signature Database (MSigDB, version 6.0, April 2017, www.broad.mit.edu/gsea/msigdb/msigdb_index.html) was used for “DNp63 Pathway”, “KEGG Cytokine”, “KEGG Chemokine” and “Reactome Chemokine “ gene sets. Gene sets for immune cell subsets (Palmer et al., 2006), lung stem cell markers (Asselin-Labat and Filby, 2012; Kim et al., 2005; Vaughan et al., 2015; Zuo et al., 2015), tracheal basal cell surface markers (Van de Laar et al., 2014), TAN vs normal neutrophils (NN) (Fridlender et al., 2012), macrophages (Segura et al., 2013), and lung tumor T-SiglecF-high vs T-SiglecF-low (Engblom et al., 2017) were generated from published data. Details of the curated gene sets are presented in Tables S2 and S3.

For transcription factor motif analyses, motif finding was performed on 100 bp surrounding the top 500 peaks based on their integer score. Motifs were discovered using the MEME suite, searching for motifs between 6 – 50 bases in length with zero or one occurrence per sequence. Flags used with the program include: -dna -mod zoops -nmotifs 5 -minw6 -maxw 50 -revcomp.

For single cell RNA-seq analysis, 10x Genomics’ Cell Ranger software (v2.1.1) executed the primary data analysis, including but not limited to: (I) demultiplexing cDNA libraries into FASTQ files; aligning reads to the mouse genome (Cell Ranger mm10 custom reference v2.1.0); (III) barcode processing for estimation of gene-cell UMI counts; (IV) QC reporting; (V) clustering, dimension reduction, and differential gene expression analysis using default parameter inputs (see https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/using/reanalyze). For further details of the Cell Ranger data processing, see https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/algorithms/overview. Secondary data analysis was evaluated with 10x Genomics’ Cell Loupe Browser. For TANs (n = 519 cells), the estimated mean reads per cell were 99,194 with 86.3% saturation density. For PBNs (n = 1,744 cells), the estimated mean reads per cell were 33,786 with 61.8% saturation density. For analysis, genes with low average UMI count were filtered out.

Quantification and Statistical Analysis

GraphPad Prism 7 was used to perform statistical analyses. Tumor-free survival analyses were analyzed using log-rank (Mantel-Cox) test. Error bars represent mean ± SEM unless otherwise indicated. For the statistical analysis of the IHC scores and tumor penetrance, column analysis was performed by Student’s unpaired t-test with p value <0.05 considered significant. For box-and-whisker plots, boxes show 25th, median and 75th quartile; whiskers extend to 1.5x interquartile range above/below the highest/lowest quartiles. All statistical details are further described in respective figure legends. Additional statistical methods related to bioinformatic analyses can be found in Bioinformatic Analyses under Method Details. No statistical methods were used to predetermine sample sizes. Please see details of IHC quantifications in Immunohistochemistry section and Western blot quantifications in Immunoblot section of Method Details.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We thank the T. Jacks laboratory for lentiviral constructs and KP mice, and M. Van Brocklin for HEK293T reporter cells. Thanks to excellent core facilities including J. O'Shea, B. Anderson, and K. Gligorich for histological services, J. Marvin for flow cytometry and B. Dalley for bioinformatics support. We are grateful to members of Oliver, Snyder and McMahon laboratories for technical assistance and feedback including P. Ballieu, D. Hansen, R. Olsen and R. Dahlgren. We acknowledge the support of grant P30CA042014 awarded to Huntsman Cancer Institute for the use of core facilities including the Biorepository and Molecular Pathology, High-Throughput Genomics and Bioinformatics Analysis, and Flow Cytometry Shared Resources and internal funding from the Immunology, Inflammation and Infectious Diseases Initiative. G.M. is supported by the NCI (F99CA223015). E.L.S. holds a Career Award for Medical Scientists from the Burroughs Wellcome Fund AND NIH (R01CA212415). T.G.O. was supported in part by the Damon Runyon Cancer Research Foundation (no. DRR-26–13), American Cancer Society (Research Scholar Award no. RSG-13–300-01-TBG) and NIH (R01CA187457).

REFERENCES

References

Asselin-Labat M-L, and Filby CE (2012). Adult lung stem cells and their contribution to lung tumourigenesis. Open Biol 2, 120094. [PubMed: 22977734]

Bass AJ, Watanabe H, Mermel CH, Yu S, Perner S, Verhaak RG, Kim SY, Wardwell L, Tamayo P, Gat-Viks I, et al. (2009). SOX2 is an amplified lineage-survival oncogene in lung and esophageal squamous cell carcinomas. Nat. Genet 41, 1238–1242. [PubMed: 19801978]

Bezzi M, Seitzer N, Ishikawa T, Reschke M, Chen M, Wang G, Mitchell C, Ng C, Katon J, Lunardi A, et al. (2018). Diverse genetic-driven immune landscapes dictate tumor progression through distinct mechanisms. Nat. Med 24, 165–175. [PubMed: 29309058]

Busch SE, Hanke ML, Kargl J, Metz HE, MacPherson D, and Houghton AM (2016). Lung Cancer Subtypes Generate Unique Immune Responses. J. Immunol 197, 4493–4503. [PubMed: 27799309]

Campbell EL, Bruyninckx WJ, Kelly CJ, Glover LE, McNamee EN, Bowers BE, Bayless AJ, Scully M, Saeedi BJ, Golden-Mason L, et al. (2014). Transmigrating neutrophils shape the mucosal microenvironment through localized oxygen depletion to influence resolution of inflammation. Immunity 40, 66–77. [PubMed: 24412613]

Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat. Genet 48, 607–616. [PubMed: 27158780]

Coffelt SB, Wellenstein MD, and de Visser KE (2016). Neutrophils in cancer: neutral no more. Nat. Rev. Cancer 16, 431–446. [PubMed: 27282249]

Engblom C, Pfirschke C, Zilionis R, Da Silva Martins J, Bos SA, Courties G, Rickelt S, Severe N, Baryawno N, Faget J, et al. (2017). Osteoblasts remotely supply lung tumors with cancer-promoting SiglecFhigh neutrophils. Science 358, eaal5081.

Eruslanov EB, Bhojnagarwala PS, Quatromoni JG, Stephen TL, Ranganathan A, Deshpande C, Akimova T, Vachani A, Litzky L, Hancock WW, et al. (2014). Tumor-associated neutrophils stimulate T cell responses in early-stage human lung cancer. J. Clin. Invest 124, 5466– 5480. [PubMed: 25384214]

Ferone G, Song J-Y, Sutherland KD, Bhaskaran R, Monkhorst K, Lambooij J-P, Proost N, Gargiulo G, and Berns A (2016). SOX2 Is the Determining Oncogenic Switch in Promoting Lung Squamous Cell Carcinoma from Different Cells of Origin. Cancer Cell 30, 519–532. [PubMed: 27728803]

Fridlender ZG, Sun J, Kim S, Kapoor V, Cheng G, Ling L, Worthen GS, and Albelda SM (2009). Polarization of Tumor-Associated Neutrophil Phenotype by TGF-B: "N1" versus "N2" TAN. Cancer Cell 16, 183–194. [PubMed: 19732719]

Fridlender ZG, Sun J, Mishalian I, Singhal S, Cheng G, Kapoor V, Horng W, Fridlender G, Bayuh R, Worthen GS, et al. (2012). Transcriptomic Analysis Comparing Tumor-Associated Neutrophils

with Granulocytic Myeloid-Derived Suppressor Cells and Normal Neutrophils. PLoS One 7, e31524. [PubMed: 22348096]

Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, Nair VS, Xu Y, Khuong A, Hoang CD, et al. (2015). The prognostic landscape of genes and infiltrating immune cells across human cancers. Nat. Med 21, 1–12. [PubMed: 25569538]

Han X, Li F, Fang Z, Gao Y, Li F, Fang R, Yao S, Sun Y, Li L, Zhang W, et al. (2014). Transdifferentiation of lung adenocarcinoma in mice with Lkb1 deficiency to squamous cell carcinoma. Nat. Commun 5, 3261. [PubMed: 24531128]

Hou S, Han X, and Ji H (2016). Squamous Transition of Lung Adenocarcinoma and Drug Resistance. Trends in Cancer 2, 463–466. [PubMed: 28741476]

Jackson EL, Willis N, Mercer K, Bronson RT, Crowley D, Montoya R, Jacks T, and Tuveson DA (2001). Analysis of lung tumor initiation and progression using conditional expression of oncogenic K-ras. Genes Dev 15, 3243–3248. [PubMed: 11751630]

Jonkers J, Meuwissen R, van der Gulden H, Peterse H, van der Valk M, and Berns A (2001). Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. Nat. Genet 29, 418–425. [PubMed: 11694875]

Kargl J, Busch SE, Yang GHY, Kim K-H, Hanke ML, Metz HE, Hubbard JJ, Lee SM, Madtes DK, McIntosh MW, et al. (2017). Neutrophils dominate the immune cell composition in non-small cell lung cancer. Nat. Commun 8, 14381. [PubMed: 28146145]

Kim CFB, Jackson EL, Woolfenden AE, Lawrence S, Babar I, Vogel S, Crowley D, Bronson RT, and Jacks T (2005). Identification of bronchioalveolar stem cells in normal lung and lung cancer. Cell 121, 823–835. [PubMed: 15960971]

Koyama S, Akbay EA, Li YY, Aref AR, Skoulidis F, Herter-Sprie GS, Buczkowski KA, Liu Y, Awad MM, Denning WL, et al. (2016). STK11/LKB1 deficiency promotes neutrophil recruitment and proinflammatory cytokine production to suppress T cell activity in the lung tumor microenvironment. Cancer Res 76, 999–1008. [PubMed: 26833127]

Kusakabe T, Kawaguchi A, Hoshi N, Kawaguchi R, Hoshi S, and Kimura S (2006). Thyroid-specific enhancer-binding protein/NKX2.1 is required for the maintenance of ordered architecture and function of the differentiated thyroid. Mol. Endocrinol 20, 1796–1809. [PubMed: 16601074]

Van de Laar E, Clifford M, Hasenoeder S, Kim BR, Wang D, Lee S, Paterson J, Vu NM, Waddell TK, Keshavjee S, et al. (2014). Cell surface marker profiling of human tracheal basal cells reveals distinct subpopulations, identifies MST1/MSP as a mitogenic signal, and identifies new biomarkers for lung squamous cell carcinomas. Respir. Res 15, 160. [PubMed: 25551685]

Langer CJ, Obasaju C, Bunn P, Bonomi P, Gandara D, Hirsch FR, Kim ES, Natale RB, Novello S, Paz-Ares L, et al. (2016). Incremental Innovation and Progress in Advanced Squamous Cell Lung Cancer: Current Status and Future Impact of Treatment. J. Thorac. Oncol 11, 2066–2081. [PubMed: 27575423]

Lee C-L, Moding EJ, Huang X, Li Y, Woodlief LZ, Rodrigues RC, Ma Y, and Kirsch DG (2012). Generation of primary tumors with Flp recombinase in FRT-flanked p53 mice. Dis. Model. Mech 5, 397–402. [PubMed: 22228755]

Li F, Han X, Li F, Wang R, Wang H, Gao Y, Wang X, Fang Z, Zhang W, Yao S, et al. (2015). LKB1 Inactivation Elicits a Redox Imbalance to Modulate Non-small Cell Lung Cancer Plasticity and Therapeutic Response. Cancer Cell 27, 698–711. [PubMed: 25936644]

Li J, Byrne KT, Yan F, Yamazoe T, Chen Z, Baslan T, Richman LP, Lin JH, Sun YH, Rech AJ, et al. (2018). Tumor Cell-Intrinsic Factors Underlie Heterogeneity of Immune Cell Infiltration and Response to Immunotherapy. Immunity 49, 178–193.e7. [PubMed: 29958801]

Lu Y, Futtner C, Rock JR, Xu X, Whitworth W, Hogan BLM, and Onaitis MW (2010). Evidence that SOX2 overexpression is oncogenic in the lung. PLoS One 5, e11022. [PubMed: 20548776]

van der Maaten L, and Hinton G (2008). Visualizing Data using t-SNE. J. Mach. Learn. Res 9, 2579–2605.

Maeda Y, Tsuchiya T, Hao H, Tompkins DH, Xu Y, Mucenski ML, Du L, Keiser AR, Fukazawa T, Naomoto Y, et al. (2012). Kras(G12D) and Nkx2–1 haploinsufficiency induce mucinous adenocarcinoma of the lung. J. Clin. Invest 122, 4388–4400. [PubMed: 23143308]

Morrisey EE, and Hogan BLM (2010). Preparing for the First Breath: Genetic and Cellular Mechanisms in Lung Development. Dev. Cell 18, 8–23. [PubMed: 20152174]

Mukhopadhyay A, Berrett KC, Kc U, Clair PM, Pop SM, Carr SR, Witt BL, and Oliver TG (2014). Sox2 cooperates with Lkb1 loss in a mouse model of squamous cell lung cancer. Cell Rep 8, 40–49. [PubMed: 24953650]

Nagaraj AS, Lahtela J, Hemmes A, Pellinen T, Blom S, Devlin JR, Salmenkivi K, Kallioniemi O, Mäyränpää MI, Närhi K, et al.(2017). Cell of Origin Links Histotype Spectrum to Immune Microenvironment Diversity in Non-small-Cell Lung Cancer Driven by Mutant Kras and Loss of Lkb1. Cell Rep 18, 673–684. [PubMed: 28099846]

Palmer C, Diehn M, Alizadeh AA, and Brown PO (2006). Cell-type specific gene expression profiles of leukocytes in human peripheral blood. BMC Genomics 7, 115. [PubMed: 16704732]

Papagiannakopoulos T, Bauer MR, Davidson SM, Heimann M, Subbaraj L, Bhutkar A, Bartlebaugh J, Vander Heiden MG, and Jacks T (2016). Circadian Rhythm Disruption Promotes Lung Tumorigenesis. Cell Metab 24, 324–331. [PubMed: 27476975]

Pitt JM, Vétizou M, Daillère R, Roberti MP, Yamazaki T, Routy B, Lepage P, Boneca IG, Chamaillard M, Kroemer G, et al. (2016). Resistance Mechanisms to Immune-Checkpoint Blockade in Cancer: Tumor-Intrinsic and -Extrinsic Factors. Immunity 44, 1255–1269. [PubMed: 27332730]

Powell DR, and Huttenlocher A (2016). Neutrophils in the Tumor Microenvironment. Trends Immunol 37, 41–52. [PubMed: 26700397]

Ribas A, and Wolchok JD (2018). Cancer immunotherapy using checkpoint blockade. Science 359, 1350–1355. [PubMed: 29567705]

Schönhuber N, Seidler B, Schuck K, Veltkamp C, Schachtler C, Zukowska M, Eser S, Feyerabend TB, Paul MC, Eser P, et al. (2014). A next-generation dual-recombinase system for time- and host-specific targeting of pancreatic cancer. Nat. Med 20, 1340–1347. [PubMed: 25326799]

Segura E, Touzot M, Bohineust A, Cappuccio A, Chiocchia G, Hosmalin A, Dalod M, Soumelis V, and Amigorena S (2013). Human inflammatory dendritic cells induce Th17 cell differentiation. Immunity 38, 336–348. [PubMed: 23352235]

Snyder EL, Watanabe H, Magendantz M, Hoersch S, Chen T.a, Wang DG, Crowley D, Whittaker C. a, Meyerson M, Kimura S, et al. (2013). Nkx2–1 represses a latent gastric differentiation program in lung adenocarcinoma. Mol. Cell 50, 185–199. [PubMed: 23523371]

Steele CW, Karim SA, Leach JDG, Bailey P, Upstill-Goddard R, Rishi L, Foth M, Bryson S, McDaid K, Wilson Z, et al. (2016). CXCR2 Inhibition Profoundly Suppresses Metastases and Augments Immunotherapy in Pancreatic Ductal Adenocarcinoma. Cancer Cell 29, 832–845. [PubMed: 27265504]

Tanaka H, Yanagisawa K, Shinjo K, Taguchi A, Maeno K, Tomida S, Shimada Y, Osada H, Kosaka T, Matsubara H, et al. (2007). Lineage-specific dependency of lung adenocarcinomas on the lung development regulator TTF-1. Cancer Res 67, 6007–6011. [PubMed: 17616654]

Tata PR, Chow RD, Saladi SV, Tata A, Konkimalla A, Bara A, Montoro D, Hariri LP, Shih AR, Mino-Kenudson M, et al. (2018). Developmental History Provides a Roadmap for the Emergence of Tumor Plasticity. Dev. Cell 44, 679–693.e5. [PubMed: 29587142]

Templeton AJ, McNamara MG, Šeruga B, Vera-Badillo FE, Aneja P, Ocaña A, Leibowitz-Amit R, Sonpavde G, Knox JJ, Tran B, et al. (2014). Prognostic role of neutrophil-to-lymphocyte ratio in solid tumors: a systematic review and meta-analysis. J. Natl. Cancer Inst 106, dju124.

Vaughan AE, Brumwell AN, Xi Y, Gotts JE, Brownfield DG, Treutlein B, Tan K, Tan V, Liu FC, Looney MR, et al. (2015). Lineage-negative progenitors mobilize to regenerate lung epithelium after major injury. Nature 517, 621–625. [PubMed: 25533958]

Watanabe H, Ma Q, Peng S, Adelmant G, Swain D, Song W, Fox C, Francis JM, Pedamallu CS, DeLuca DS, et al. (2014). SOX2 and p63 colocalize at genetic loci in squamous cell carcinomas. J. Clin. Invest 124, 1636–1645. [PubMed: 24590290]

Xiao Z, Jiang Q, Willette-Brown J, Xi S, Zhu F, Burkett S, Back T, Song NY, Datla M, Sun Z, et al. (2013). The pivotal role of IKKα in the development of spontaneous lung squamous cell carcinomas. Cancer Cell 23, 527–540. [PubMed: 23597566]

Xu C, Fillmore CM, Koyama S, Wu H, Zhao Y, Chen Z, Herter-Sprie GS, Akbay E.a, Tchaicha JH, Altabef A, et al. (2014). Loss of Lkb1 and Pten Leads to Lung Squamous Cell Carcinoma with Elevated PD-L1 Expression. Cancer Cell 25, 590–604. [PubMed: 24794706]

Young NP, Crowley D, and Jacks T (2011). Uncoupling cancer mutations reveals critical timing of p53 loss in sarcomagenesis. Cancer Res 71, 4040–4047. [PubMed: 21512139]

Zhang H, Fillmore Brainson C, Koyama S, Redig AJ, Chen T, Li S, Gupta M, Garcia-de-Alba C, Paschini M, Herter-Sprie GS, et al. (2017). Lkb1 inactivation drives lung cancer lineage switching governed by Polycomb Repressive Complex 2. Nat. Commun 8, 14922. [PubMed: 28387316]

Zheng H, Ying H, Yan H, Kimmelman AC, Hiller DJ, Chen A-JJ, Perry SR, Tonon G, Chu GC, Ding Z, et al. (2008). p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. Nature 455, 1129–1133. [PubMed: 18948956]

Zlotnik A, and Yoshie O (2012). The chemokine superfamily revisited. Immunity 36, 705–716. [PubMed: 22633458]

Zuo W, Zhang T, Wu DZ, Guan SP, Liew A-A, Yamamoto Y, Wang X, Lim SJ, Vincent M, Lessard M, et al. (2015). p63(+)Krt5(+) distal airway stem cells are essential for lung regeneration. Nature 517, 616–620. [PubMed: 25383540]

HIGHLIGHTS

- Mouse models of lung cancer recapitulate human NSLC immune microenvironment
- SOX2 suppresses NKX2–1 activity, and NKX2–1 represses TAN recruitment
- NKX2–1 loss accelerates adeno-to-squamous transdifferentiation
- TANs possess tumor-promoting features and impact squamous tumorigenesis

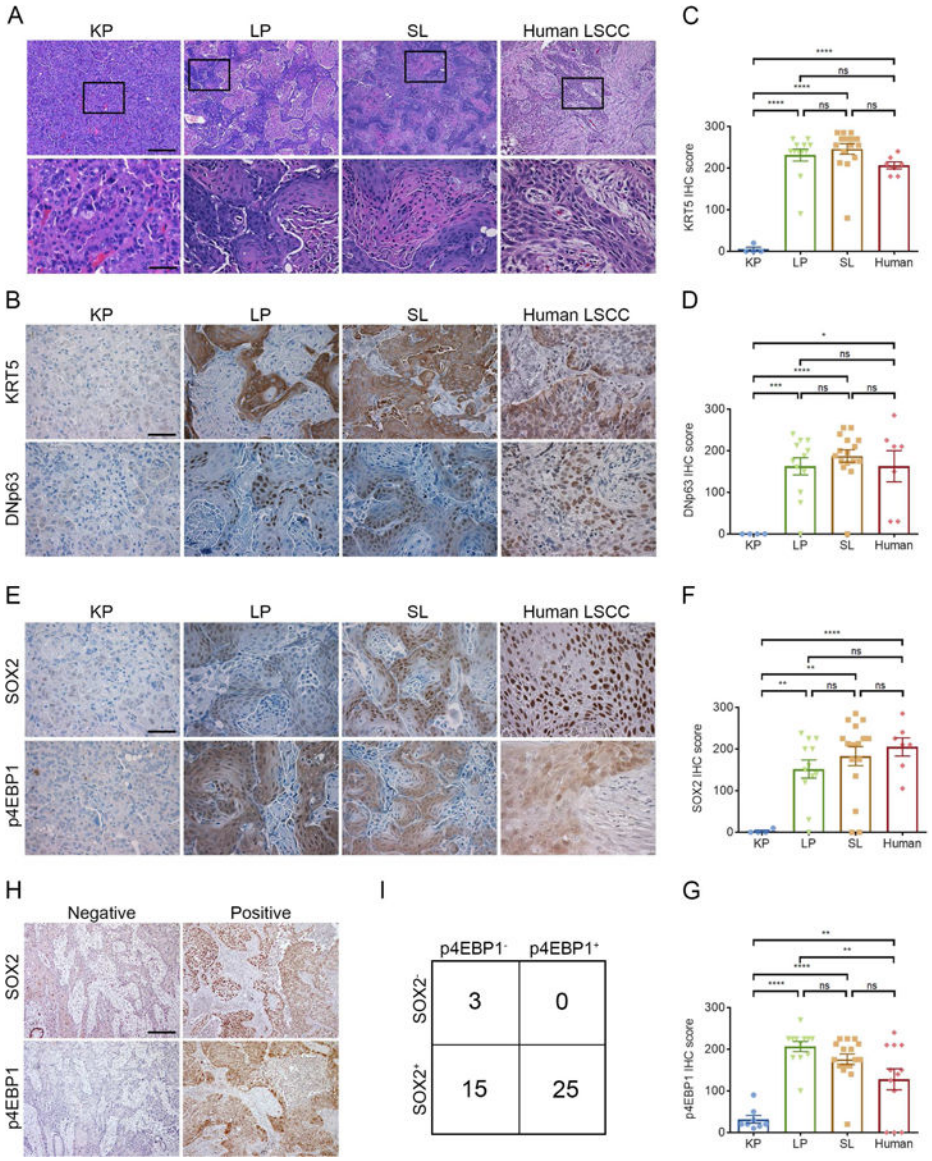


Figure 1. Murine SOX2-Driven Lung Squamous Tumors Recapitulate Human Pathology

(A) Representative H&E images from indicated GEMM tumors and human LSCC. Scale bars, 200 μm top row, 50 μm bottom row. Boxes on upper panel indicate areas shown in higher magnification on lower panel.

(B-D) IHC for KRT5 and DNp63 in indicated mouse and human tumor types (B) and IHC quantification of KRT5 (C) and DNp63 (D). Scale bar, 50 μm.

(E-G) IHC for SOX2 and p4EBP1 in indicated mouse and human tumor types (E) and IHC quantification of SOX2 (F) and p4EBP1 (G). Scale bar, 50 μm.

(H and I) IHC for SOX2 and p4EBP1 in a panel of 43 human LSCCs (H), and contingency table for binary staining results (I). Scale bar, 50 μm.

Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. In C, D, F, G, each dot represents one tumor from 3–8 mice per genotype and 6 patient tumors. See also Figure S1.

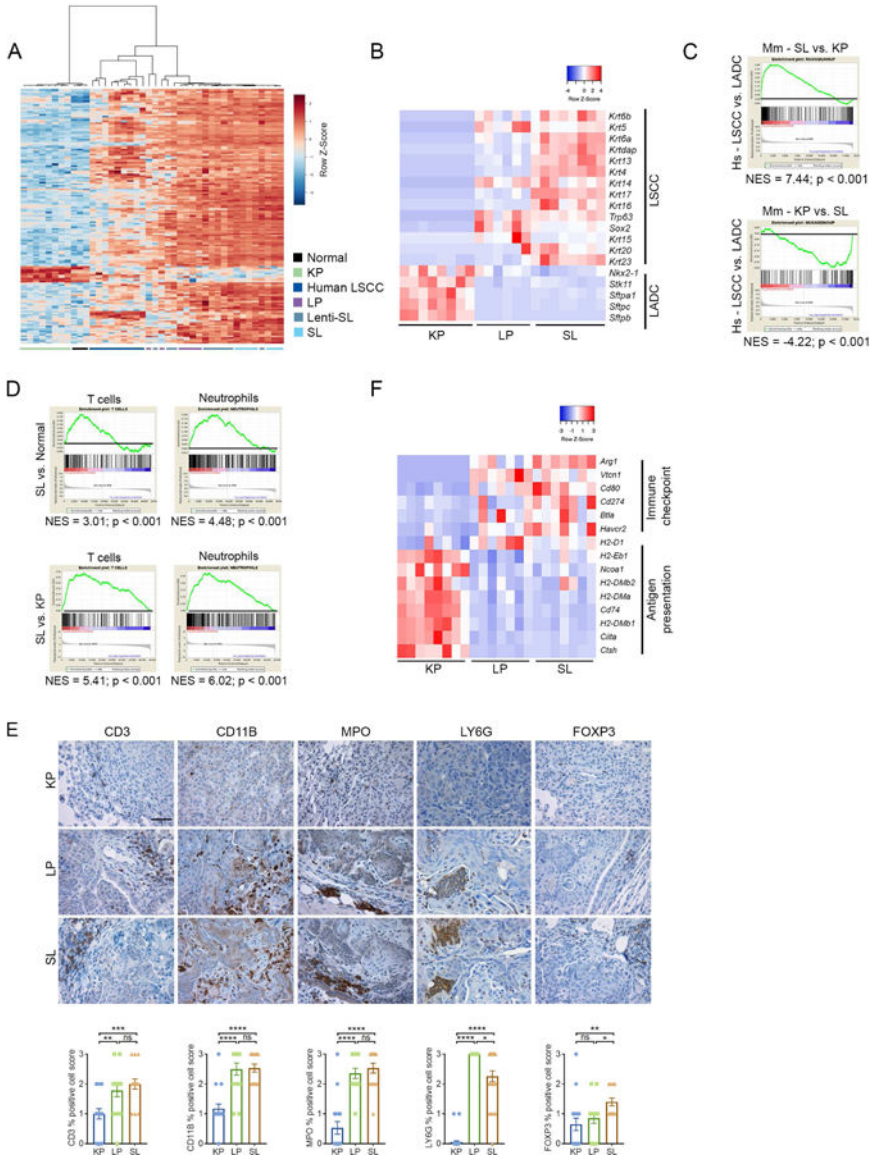


Figure 2. Mouse Lung Tumors Recapitulate Molecular Phenotype and Tumor Immune Microenvironment of Human Tumors

(A) Heatmap and hierarchical clustering of GEMM lung tumors, normal mouse lung tissue, and human LSCCs based on signature of differentially expressed transcripts from RNA-seq analysis (see Methods).

(B) Gene expression heatmap for lung squamous and adenocarcinoma marker genes comparing SL, LP and KP tumors. $p < 0.01$ Log2FC > 1 as a cutoff.

(C) GSEA from mouse (Mm) SL versus KP (top) and KP versus SL (bottom) tumors with normalized enrichment scores (NES) and p values for human (Hs) LSCC versus LADC gene signatures generated from TCGA data.

(D) GSEA from mouse SL versus normal lung tissue (top) and SL versus KP tumors (bottom) with NES and p values for T cell and neutrophil gene signatures.

(E) IHC for immune cell markers (CD3, T cells; FOXP3, Tregs; CD11B, MPO, LY6G, neutrophils) in indicated GEMM tumors (top), and IHC quantification (bottom). Scale bar,

50 μm. Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. Each dot represents one tumor from 3–7 mice per genotype.

(F) Heatmap representing selected immune-related genes differentially expressed in SL, LP and KP tumors with $p < 0.01$ Log2FC > 1 as a cutoff. See also Figure S2 and Table S1.

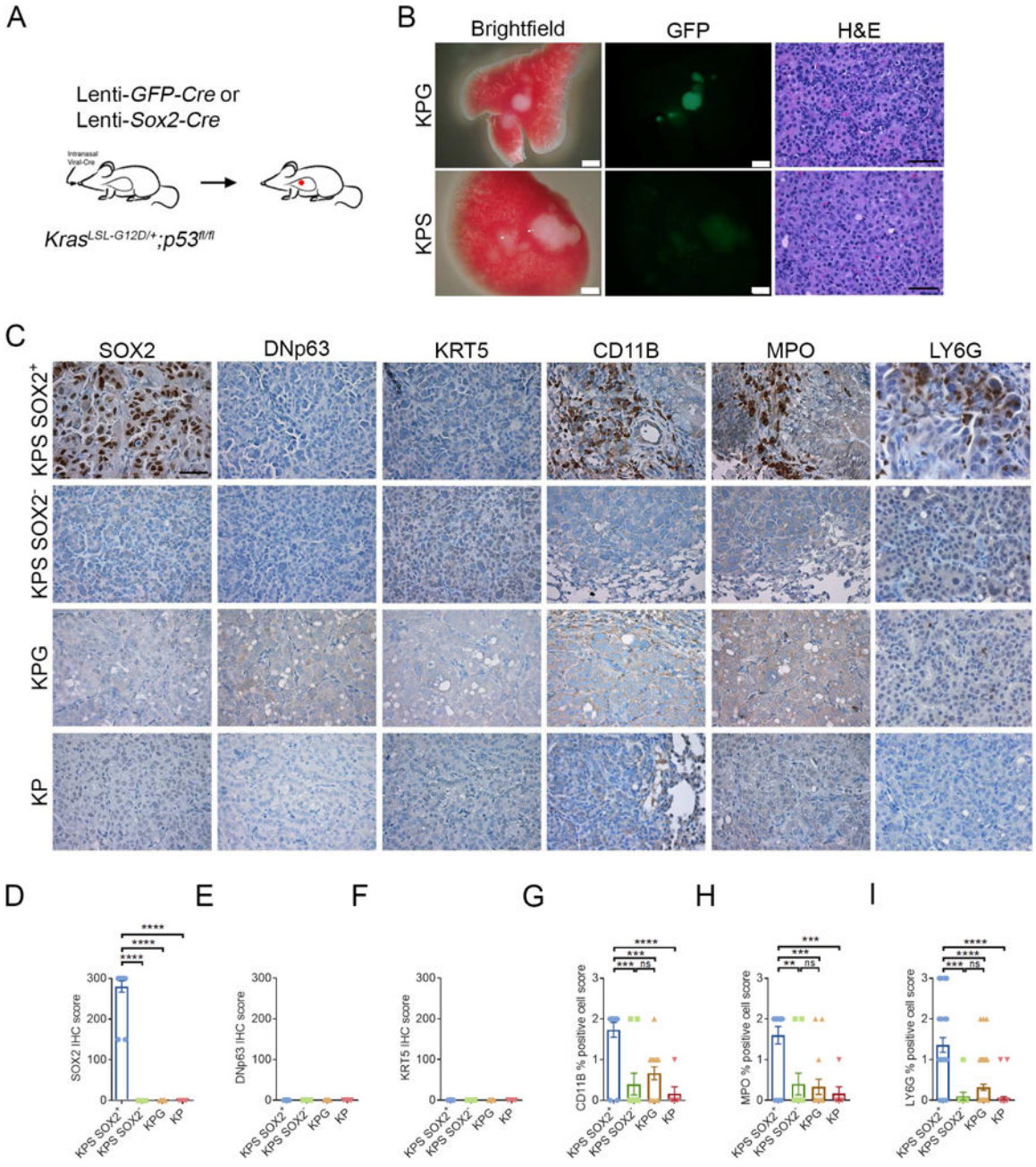
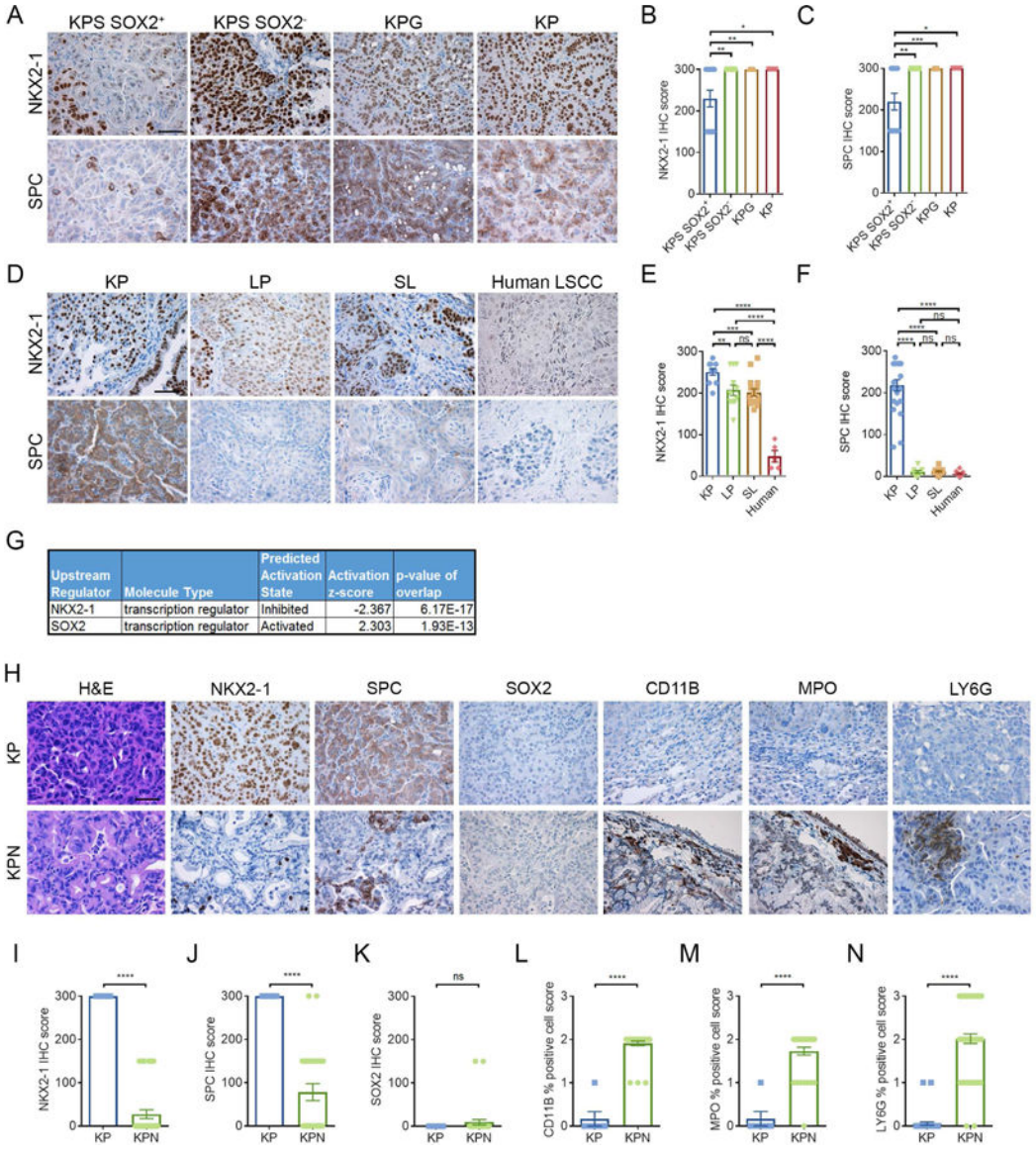


Figure 3. SOX2 Promotes Tumor-Associated Neutrophil Recruitment in the Absence of Squamous Transdifferentiation

(A) Schematic representation of lentiviral *Sox2* or *Gfp* overexpression in KP mice.

(B) Representative whole-mount brightfield (left), fluorescence (middle) and H&E stained histology (right) images from indicated GEMM tumors. Scale bars, 200 μm for brightfield and fluorescence; 50 μm for H&E.

(C-I) Representative IHC (C) and IHC quantification for SOX2 (D), DNp63 (E), KRT5 (F), CD11B (G), MPO (H), or LY6G (I) in indicated tumor models. Scale bar, 50 μm. Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{**}p < 0.01$, ns = not significant. In D-I, each dot represents one tumor from 3–5 mice per genotype.



Upstream Regulator	Molecule Type	Predicted Activation State	Activation z-score	p-value of overlap
NKX2-1	transcription regulator	Inhibited	-2.367	6.17E-17
SOX2	transcription regulator	Activated	2.303	1.93E-13

Figure 4. SOX2 Suppresses NKX2–1 Activity and NKX2–1 Loss Promotes TAN Recruitment in the Absence of Squamous Histotype

(A-C) IHC for NKX2–1 and SPC in indicated mouse tumor genotypes (A) and IHC quantification of NKX2–1 (B) and SPC (C).

(D-F) IHC for NKX2–1 and SPC in indicated mouse and human tumors (D) and IHC quantification of NKX2–1 (E) and SPC (F).

(G) IPA upstream regulator analysis of RNA-seq data identify SOX2 and NKX2–1 with activation z-scores and p values for SL versus KP tumors.

(H-N) Representative IHC (H) and IHC quantification for NKX2–1 (I), SPC (J), SOX2 (K), CD11B (L), MPO (M) and LY6G (N) in indicated tumor models.

Scale bars, 50 μm. Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. In B, C, E, F, I-N, each dot represents one tumor from 3–8 mice per genotype and 6 patient LSCC tumors. See also Figure S3.

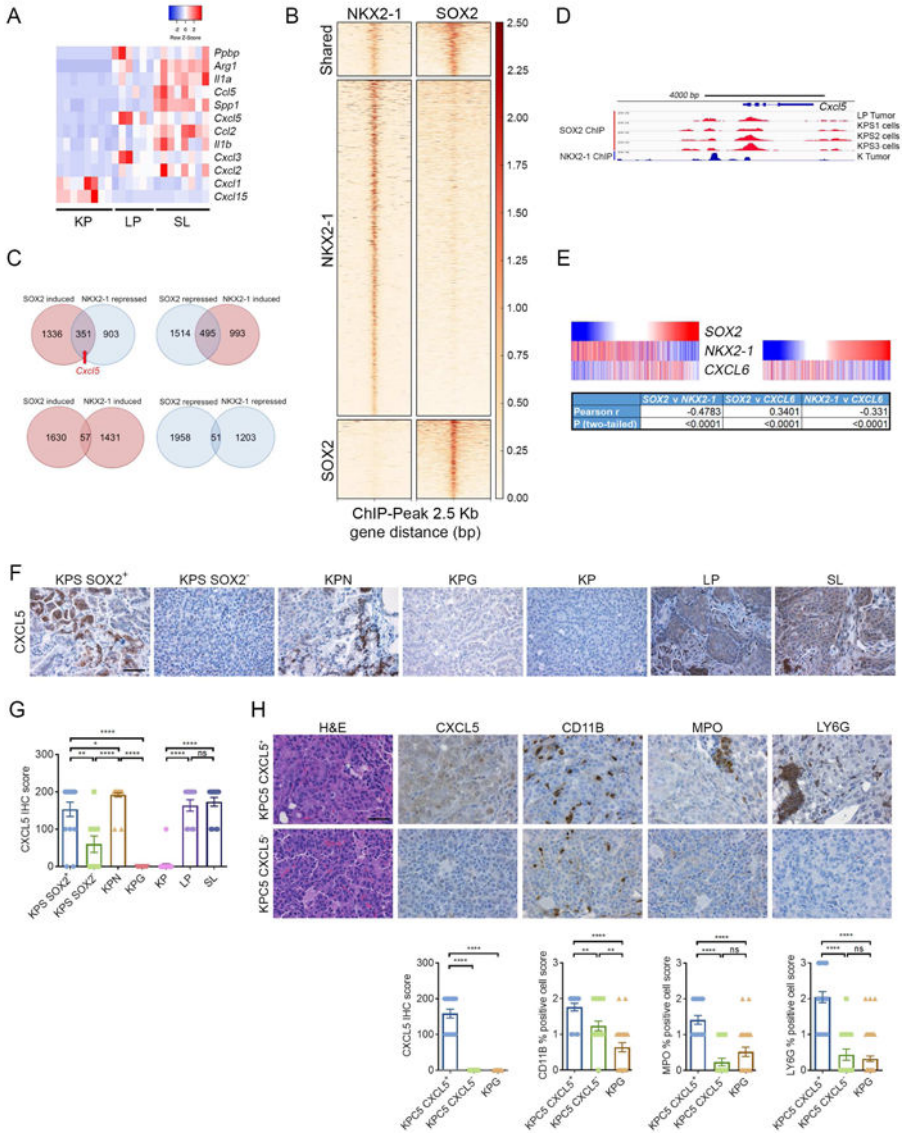


Figure 5. SOX2 and NKX2–1 Inversely Regulate Neutrophil Chemoattractant *Cxcl5*

(A) Gene expression heatmap for genes implicated in neutrophil recruitment in SL, LP and KP tumors. $p < 0.01$ Log2FC > 1 as a cutoff.

(B) ChIP-seq heatmap view of genome-wide binding sites of SOX2 (LP tumors and KPS cells) and NKX2–1 [K tumors (Snyder et al., 2013)].

(C) Venn diagrams indicating the total number of genes that are direct genomic targets of SOX2 and NKX2–1 with the directionality of transcriptional regulation. ChIP-seq data were integrated with RNA-seq data (SL versus KP tumors) and exon array data (KN versus K tumors) to define directionality of transcriptional regulation. $p < 0.05$ Log2FC > 1 as a cutoff.

(D) ChIP analysis of SOX2 and NKX2–1 genomic binding at the *Cxcl5* locus in indicated samples.

(E) Gene expression heatmaps for *SOX2*, *NKX2–1* and *CXCL6* in TCGA Lung Cancer dataset (n = 1,129). Patient samples are sorted based on SOX2 (left) or NKX2–1 (right) expression levels. Pearson correlation coefficient and two-tailed p values for each correlation gene pair is listed as a table (bottom). Data visualized by UCSC Xena Browser.
(F and G) IHC for CXCL5 in indicated GEMM lung tumors (F) and IHC quantification (G).
(H) Representative H&E images and IHC for CXCL5, CD11B, MPO and LY6G in indicated mouse tumors (top) and IHC quantification (bottom).Scale bars, 50 μm. Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. In G and H, each dot represents one tumor from 3–6 mice per genotype. See also Figure S4 and Table S4.

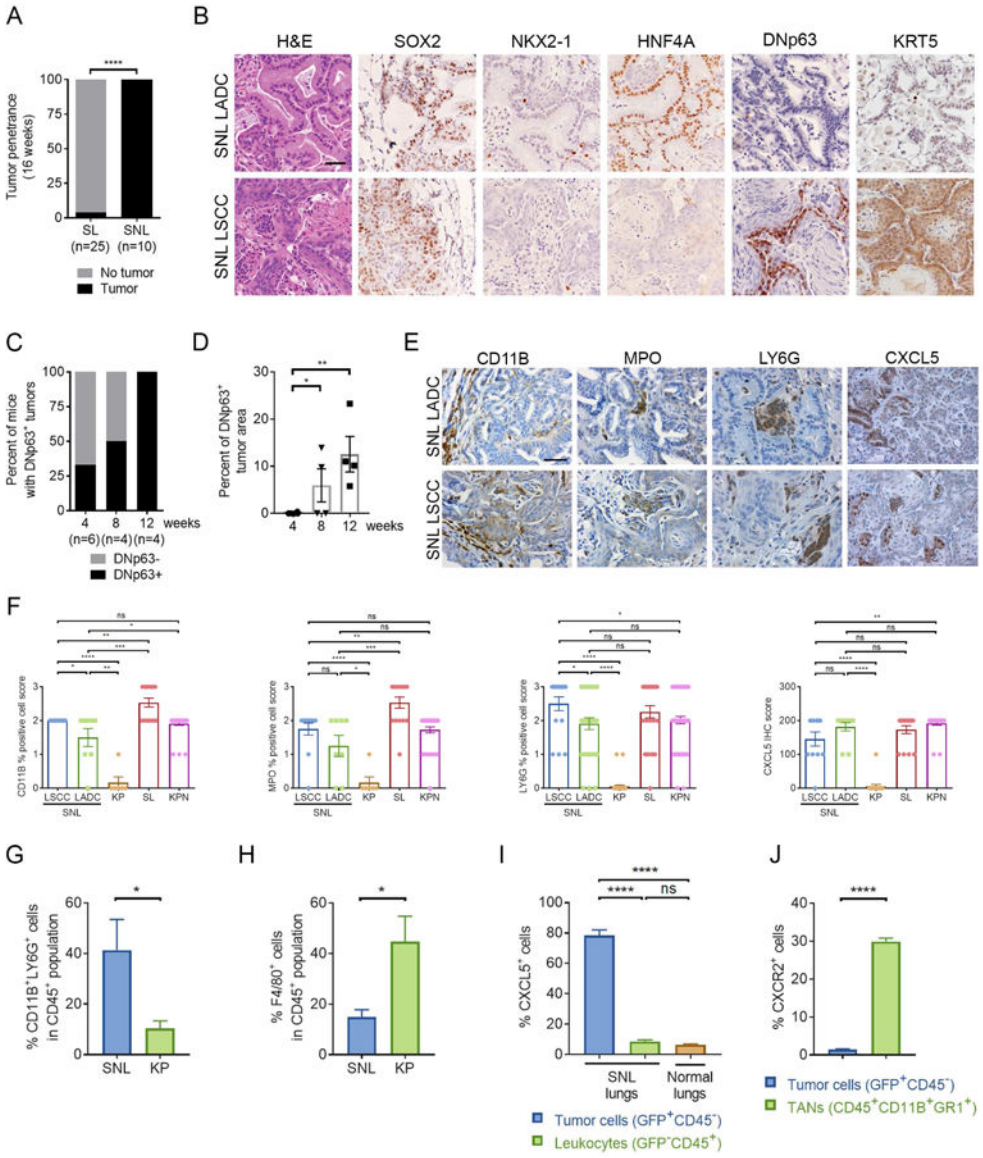


Figure 6. Loss of NKX2–1 Dramatically Accelerates Squamous Lung Tumorigenesis

(A) Proportion of SL (n = 25) versus SNL (n = 10) mice with squamous carcinoma at 16 weeks post-infection with Ad5-CMV-Cre. Fisher's exact test (two-tailed), $^{****}p < 0.0001$.

(B) Representative images of H&E and IHC for SOX2, NKX2–1, HNF4A, DNp63 and KRT5 in LADC and LSCC tumors in SNL model at 16 weeks post-infection.

(C) Percent of mice with DNp63^{+} tumors in SNL mice at 4 (n = 6 mice), 8 (n = 4), and 12 (n = 4) weeks post-infection.

(D) Percent of DNp63^{+} tumor area over total tumor area in SNL mice at 4 (n = 6 mice), 8 (n = 4), and 12 (n = 4) weeks post-infection. Mann Whitney tests, $^{**}p < 0.01$, $^{*}p < 0.05$. Each dot represents one mouse.

(E and F) Representative IHC for CD11B, MPO, LY6G and CXCL5 in LADC and LSCC in SNL modeland IHC quantification (F). KP, SL and KPN quantification data is replicated

from various other figures for ease of comparison. Each dot represents one tumor from 4 mice per group.

(G) Quantification of flow cytometry data for $CD11B^{+}LY6G^{+}$ neutrophils as percentage of leukocytes ($CD45^{+}$) in SNL versus KP lungs (n = 4–6 samples per group).

(H) Quantification of flow cytometry data for $F4/80^{+}$ macrophages as percentage of leukocytes ($CD45^{+}$) in SNL versus KP lungs (n = 6–7 samples per group).

(I) Quantification of flow cytometry data for $CXCL5^{+}$ cancer cells ($GFP^{+}CD45^{-}$) and leukocytes ($GFP^{-}CD45^{+}$) in SNL versus normal lungs (n = 4–5 samples per group).

(J) Quantification of flow cytometry data for $CXCR2^{+}$ cancer cells ($GFP^{+}CD45^{-}$) and TANs ($CD45^{+}CD11B^{+}GR1^{+}$) in SNL lungs (n = 8 samples per group).

Scale bars, 50 μm. Error bars indicate mean ± SEM. Two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$, ns = not significant. See also Figure S5.

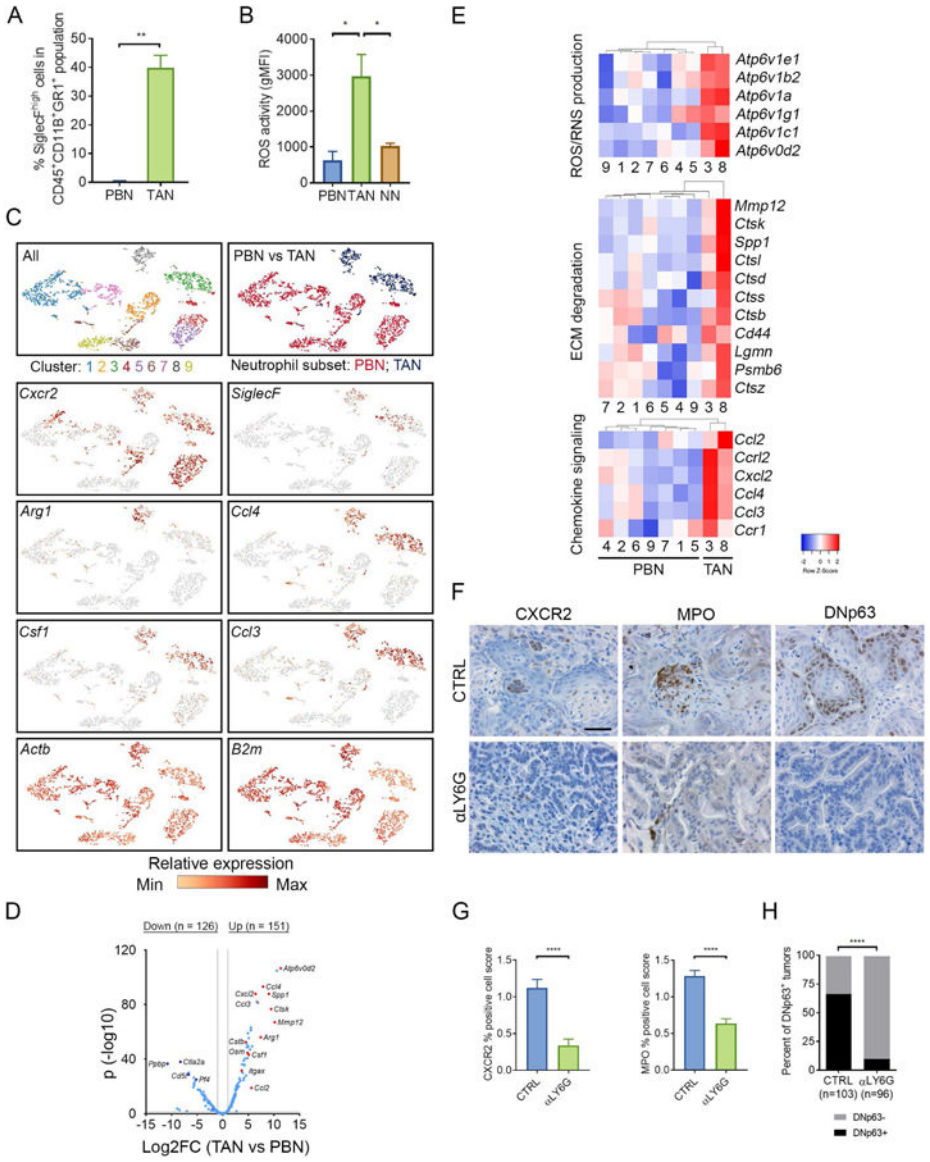


Figure 7. TANs are Distinct from Peripheral Blood Neutrophils with Pro-tumor Features

(A) Quantification of flow cytometry data for SiglecF^{high} cells as percentage of neutrophils (CD45^{+}CD11B^{+}GR1^{+}) in SNL lung tumors (n = 8) versus peripheral blood neutrophils (PBNs) from SNL tumor-bearing mice (n = 2).

(B) Quantification of flow cytometry data as geometric mean fluorescence intensity (gMFI) for ROS activity in PBNs, normal lung neutrophils (NNs), and tumor-associated neutrophils (TANs) (n = 4–8 samples per group from n = 2 mice each). Error bars indicate mean ± SEM. Two-tailed unpaired t tests, * $p < 0.05$.

(C) tSNE plots of scRNA-seq data demonstrating all cell clusters (top left), PBN versus TAN cells (top right), and relative expression levels of selected genes (other panels). Flow-sorted samples were derived and pooled from blood or lung tumor of SNL mice (n = 2 each).

(D) Volcano plot showing differential gene expression of scRNA-seq data comparing TANs versus PBNs. $p < 0.05$ Log2FC > 1 as a cutoff (denoted by gray lines parallel to X and Y axes). Selected genes are highlighted (red).

(E) Gene expression heatmaps for genes implicated in ROS/RNS production, extracellular matrix (ECM) degradation/cysteine endopeptidase activity, and chemokine signaling based on cell clusters. Gene sets derived from Enrichr analyses. Cell cluster numbers are labeled below each column identified in Figure 7C. $p < 0.05$ Log2FC > 1 as a cutoff.

(F-H) Representative images of IHC for CXCR2, MPO and DNp63 in SNL mice treated with anti-LY6G antibody versus anti-IgG1 control antibody thrice weekly for 3–4 weeks (n = 8 mice per group) (F) and IHC quantification for CXCR2 and MPO (G) and DNp63 (H). Scale bar, 50 μm. In A, B and G, error bars indicate mean ± SEM and two-tailed unpaired t tests, $^{****}p < 0.0001$, $^{**}p < 0.005$, $^{*}p < 0.05$. In H, Fisher's exact test (two-tailed), $^{****}p < 0.0001$. See also Figure S6 and Table S5.