Factor XIII A—expressing inflammatory monocytes promote lung squamous cancer through fibrin cross-linking

Lung cancer is the leading cause of cancer-related deaths worldwide, and lung squamous carcinomas (LUSC) represent about 30% of cases. Molecular aberrations in lung adenocarcinomas have allowed for effective targeted treatments, but corresponding therapeutic advances in LUSC have not materialized. However, immune checkpoint inhibitors in sub-populations of LUSC patients have led to exciting responses. Using computational analyses of The Cancer Genome Atlas, we identified a subset of LUSC tumors characterized by dense infiltration of inflammatory monocytes (IMs) and poor survival. With novel, immuno-competent metastasis models, we demonstrated that tumor cell derived CCL2-mediated recruitment of IMs is necessary and sufficient for LUSC metastasis. Pharmacologic inhibition of IM recruitment had substantial anti-metastatic effects. Notably, we show that IMs highly express Factor XIII A, which promotes fibrin cross-linking to create a scaffold for LUSC cell invasion and metastases. Consistently, human LUSC samples containing extensive cross-linked fibrin in the microenvironment correlated with poor survival.
For decades lung cancer has been the leading cause of cancer-related deaths in the U.S. and worldwide. Because non-small cell lung cancer (NSCLC) has a dismal (~15%) 5-year survival rate, novel therapies are desperately needed. The recent discovery of select molecular aberrations (e.g. EGFR mutations and ALK translocations) in lung adenocarcinomas (LUAD) has led to the development of highly effective targeted therapies in these subsets of lung cancer. On the contrary, such advances in the treatment of lung squamous carcinomas (LUSC), which account for about 30% of lung cancer cases, have not materialized. However, the therapeutic blockade of immune checkpoints in LUSC patients has demonstrated exciting responses. In fact, several phase III clinical trials recently led to FDA approval of anti-PD1 antibodies in the first- and second-line treatment of LUSC, suggesting that LUSC may be suitable for additional examination of immune-oncology approaches.

Molecular profiling analyses based on The Cancer Genome Atlas (TCGA) data have revealed that LUSC tumors are highly idiosyncratic and not likely driven by solitary actionable pathways. Using microarray analyses of LUSC tumors, one group previously defined four gene expression subtypes: Classical, Basal, Primitive, and Secretory. These subtypes feature distinct biological processes based on patterns of gene expression. Amongst these subtypes, the Secretory subtype was defined by an immune-response signature rich in genes associated with complement activation, immune cell recruitment, and inflammation. Building upon these observations, we computationally analyzed the TCGA TMA previously characterized by subtype. We found that compared with Classical and Basal subtypes, the Secretory subtype of LUSC is immune-rich and has poor survival.

Tumors recruit IMs (CCR2HighCD14+CD16Low in humans; CCR2HighLy6CHigh in mice) through secretion of the CCL2 chemokine. IMs differentiate into either tumor-associated macrophages (TAMs) or dendritic cells (DCs), and IM-derived TAMs have been intensely investigated for their roles in promoting cancer progression. For example, IM-derived TAMs can promote metastasis through production of VEGF. VEGF has well-recognized roles in distant metastasis formation, in part because it transiently increases vascular permeability to facilitate cancer cell extravasation. TAM secretion of epidermal growth factor (EGF) and IL-6 promote increased migration and infiltration. Inflammation.

We used Ingenuity Pathway Analysis (IPA) to investigate possible relationships between the "survival genes" and found that leukocyte migration, wound healing, and complement activation emerged as potential tumor-promoting mechanisms within the microenvironment. Gene ontology (GO) analyses showed that the Secretory subtype is highly enriched for immune-response biological processes, while Classical is characterized by genes of the reduction-oxidation responses. This is consistent with the Secretory subtype being highly associated with the KEAP1/NRF2 pathway, a tightly coupled antioxidant program, which is also enriched in KRAS/Lkb1 LUADs that lack immune-response features. Gene Set Enrichment Analysis (GSEA) revealed that top signatures in the Secretory subtype were strongly associated with monocytes and T-cell infiltration. Intriguingly, several markers of IMs, often regarded as precursors of TAMs, were leading predictors of LUSC progression, whereas standard macrophage markers displayed more modest survival associations.

Inflammatory monocytes associate with Secretory LUSC tumors. Splitting CD14 into expression quartiles confirmed a marked separation of the survival curves at the median expression level (Fig. 1a). In fact, CD14 had remarkable prognostic relevance among stage II patients following surgery (Supplementary Fig. 1c), implying its importance in disease recurrence. We found that tumors with above median CD14 expression predominantly enriched for the Basal and Secretory subtypes (Fig. 1b). Consistent with their importance in recruiting IMs and promoting metastasis, several chemokines (notably CCL2, the classic CCR2 ligand) significantly predicted poor survival and strongly correlated with CD14 expression. We found an impressive dynamic range of expression for CD14 and these chemokines by subtype, with the Basal and Secretory subtypes consistently having the highest levels (Fig. 2c, d). Recently, CD14 expression in bladder cancer cells was identified as a mechanism of tumor progression. To determine whether LUSC CD14 expression arose from tumor cells and/or tumor-infiltrating leukocytes, we performed immunohistochemistry (IHC) on a LUSC tissue microarray (TMA) previously characterized by subtype. The secretory subtype had the highest expression of CD14, while the classical subtype had the lowest expression, with the intermediate subtypes having intermediate expression. Intense membranous CD14 staining was predominantly found on tumor-infiltrating leukocytes, and...
are associated with reduced overall survival and are markers of monocytes and macrophages. Genes with log-rank p

Enrichment Analyses (GSEA) of the LUSC TCGA dataset. The GSEA is performed going from Secretory to Classical; the GSEA

details). Individual bars represent most statistically signi

processes most linked with genes differentially expressed, moving between the Classical and Secretory subtypes (see the Supplementary Methods for

details). Individual bars represent most statistically significant GO terms in either the Classical (red bars) or Secretory (green bars) subtype. g Gene Set

Enrichment Analyses (GSEA) of the LUSC TCGA dataset. The GSEA is performed going from Secretory to Classical; the GSEA ‘mountain plots’ show only

the two most divergent subtypes. Gene set names were shortened to fit this figure. h Table showing 9 genes from the upper portion of the heat map that

are associated with reduced overall survival and are markers of monocytes and macrophages. Genes with log-rank p≤0.001 are highlighted in red

rarely on cancer cells (Supplementary Fig. 3). Consistent with

RNA-seq data, CD14 protein expression correlated with subtype, and the Secretory subtype displayed the highest levels, followed

by Basal (Supplementary Fig. 3). To better characterize the

number and location of CCR2High/CD14+ IMs in these LUSC subtypes, we performed multiplex IHC for CCR2, CD14, and

pan-cytokeratin (CK) (Fig. 2g). Across all subtypes, we found that
dual positive (CD14+/CCR2+) cells were almost three times

more abundant in the stromal (pan-CK negative) regions than in

the cancer cell islet (pan-CK positive) regions (Fig. 2g, h). Consistent with mRNA expression levels, there were more dual

positive cells in the Basal and Secretory subtypes in both stromal

and cancer cell islet regions (Fig. 2i). To determine what propor
tion of CD14+/CCR2+ cells represent differentiated M2

macrophages, we performed multiplex IHC for CD14/CCR2/

CD206 on 99 lung tumors. This technique revealed that 85% of

the immune infiltrates stained exclusively for CD14 and CCR2,
suggesting that only a small subset of CD14+/CCR2+ cells

represent CD206+ TAMs (P < 10−4, Supplementary Fig. 4).

Inflammatory monocytes correspond with poor survival and CD14. Although CD14 is most highly expressed on IMs amongst

leukocytes,20,26 it is possible that the poor survival associated with
Given the recent advances in immunogenomic profiling to uncover immune infiltrates in tumors with high fidelity, we applied a modified 'Immunome' signature across all LUSC TCGA tumors based on a median CD14 expression level threshold (Supplementary Data 7). Classical tumors were predominately represented in the low ( < median) CD14 expression cohort, while Secretory tumors were predominantly in the high ( ≥ median) CD14 expression cohort.
CD14 expression cohort (Fig. 3a, top). Basal tumors more often segregated to the high cohort, and Primitive tumors significantly re-classified in the low cohort (Fig. 3a, top).

Next, we assessed 9 immune cell types for their correlation with CD14 expression and their individual contribution to overall survival (Fig. 3a, Supplementary Fig. 5). In both analyses, IMs had the strongest relationship with CD14 expression and poor overall survival (Fig. 3b, c). Intriguingly, activated DCs (aDCs) and M2 macrophages, both derivatives of IMs, had the second and third strongest relationships with poor survival, respectively (Fig. 3b, c). These findings imply that IMs have both direct and indirect roles in LUSC progression via differentiation into aDCs or M2 macrophages. Furthermore, when assessing for all adaptive and innate immune cell densities, we observed that IMs have strong correlations not only with aDCs and M2 macrophages but also with regulatory T-cells (Tregs) and immune checkpoints (Supplementary Fig. 6), strongly implicating the presence of an immune suppressive environment. We thus define this high CD14 expressing cohort as the ‘IM-rich subset’ of LUSC.

**Fig. 3 IMs have the strongest correlation with CD14 and LUSC survival.** a At top, the mRNA subtype of each sample is displayed (Classical: red, Basal: blue, Primitive: black, Secretory: green); CD14 expression levels sorting from left (lowest) to right (highest) for the corresponding samples are displayed below. Heat maps of the CD14 + populations that are associated with a statistically significant survival in LUSC are arranged by the most (IMs) to the least (MDSCs) statistically significant. CD14 is a marker both of IMs and MDSCs but is not shown in their heat maps, to avoid data redundancy. Note for the heat map figure legend: gray represents ‘null normalized values’ (NNV). b Scatter plot of overall survival results and CD14 scores for the 9 CD14 + immune cell types. Survival is represented with hazard ratio (HR) ± 95% confidence intervals. Values above 1 indicate worse survival based on cell type density. The CD14 score is the rank of the ratio between the average cell type density score of samples with high (≥ median) vs. low (< median) CD14 levels. c Log-rank p-values of overall survival for the 9 CD14 + immune cell types (high vs. low cell density score). The red bars show statistically significant cases, while black bars are used when the statistical significance threshold of 0.05 is not reached.

TNFα-dependent NFκB activation leads to CCL2-driven IM recruitment. Given our findings that IMs may promote LUSC metastasis, we sought to functionally characterize IMs in an immune-competent mouse model. To date, the field has lacked an immune-competent mouse model of LUSC that faithfully metastasizes. To address this limitation, we began by characterizing the metastatic properties of the murine LUSC cell line (KLN205) derived from bronchial carcinogen exposure by performing orthotopic injections in syngeneic DBA2 mice. The resulting tumors were poorly differentiated, exhibited central necrosis, and displayed classic IHC patterns of human LUSC (Supplementary Fig. 7a, b). Following several rounds of an in vivo passage selection technique, we developed sub-clones (LN2-2 and LN4K1) with distinct metastatic properties. Both sub-clones had significantly increased number and frequency of lymph node metastases; however, the LN4K1 sub-clone developed more distant metastases, while LN2-2-injected mice rapidly died from malignant pleural effusions (Supplementary Fig. 7c–g). Although KLN205 and LN4K1 had similar intrinsic growth rates in vitro, LN4K1 tumors grew significantly faster in vivo (Supplementary...
**Fig. 4** TNFα activation of NFκB promotes CCL2-mediated IM recruitment. a Microarray expression data (left) comparing murine bronchial epithelial cells (MBECs), parental KLN205 and the LN4K1 sub-clone. Top upstream pathways (right) from Ingenuity Pathway Analysis (IPA) are shown for the differentially regulated genes shown in brackets. b An upstream network visualization from IPA of all over-expressed genes (all nodes) in the upper portion of the heat map shown in Fig. 1c with significant log-rank (Vaisir analysis) p-value (<0.05). TNFα and NFκB (blue nodes) were amongst the top upstream regulators known to have direct roles (black lines) in promoting CCL2, CCL3 and CSF1 chemokines (the degree of redness corresponds with increasing statistical significance). e Relative expression of TNFα, d CCL2, CCL3 and CSF1 by qPCR. Data are averages ± s.e.m. **P**-values were obtained with Student’s t-test in comparison with KLN205. f Relative levels of CCL2 as measured by ELISA from secreted media of cells growing in vitro or, f, from plasma of tumor-bearing mice. Data are averages ± s.e.m. g Relative mRNA expression of CCL2 and p65 in LN4K1 cells following treatment with control or p65 siRNA with or without exogenous TNFα (100 ng/mL). h Relative expression of CCL2 mRNA (top) and phospho-p65 and p65 protein (bottom) following treatment with DMSO or an IKKβ inhibitor (Compound A, 5 μM) for 5 h. i Relative IM counts in the bone marrow, blood, spleen from healthy DBA2 mice versus those with LN4K1 tumors. j Relative IM, TAM and TReg counts from the lungs of healthy versus LN4K1-bearing mice. IMs were also assessed in age-matched DBA2 mice following HBSS ‘Mock’ injection. P-values obtained with one-sided Student’s t-test, n = 5 mice/group for f, i, and j. * P <0.05, ** P ≤0.01, *** P ≤0.001
Fig. 8a, b), suggesting important differences in the TME. While no changes were observed in angiogenesis, LN4K1 tumors had increased proliferative indices (Supplementary Fig. 8c, d). Similarly, human Secretory tumors exhibited increased proliferative indices relative to Classical tumors, while no significant differences in angiogenesis were found among the four subtypes of LUSC (Supplementary Fig. 8e, f).

To explore the molecular mechanisms underlying the metastatic properties of LN4K1, we compared the expression profiles of 3 cell lines: murine bronchial epithelial cells (MBEC) isolated...
from healthy adult DBA2 mice, KLN205, and LN4K1. Several patterns emerged, the most significant of which were signatures of TNFα, IFNγ, and NFκB signaling activation (Fig. 4a, Supplementary Data 8). Intriguingly, this result matched an IPA screening of the TCGA “survival genes”, which also revealed that TNFα (P < 10^-34) and NFκB (P < 10^-11) were highly significant and interconnected upstream drivers of the IM chemokines (Fig. 4b, Supplementary Data 9). Higher TNFα expression determined poorer overall survival and was significantly observed in the Secretory LUSC subtype. TNFα expression also corresponded with the pattern of IM chemokines (Fig. 2e, f Supplementary Fig. 9). Using an analysis comparing our lung squamous carcinoma model and TCGA LUSC RNA-seq data based on the approach used by Xu et al., we observed that i) the two main biological processes (with the most genes and lowest p-values) among genes up-regulated in Secretory as well as in LN4K1 are ‘signal transduction’ (P = 2.6 x 10^-53, 23 genes) and ‘inflammatory response’ (P = 4.2 x 10^-34, 11 genes); and ii) shared inflammation-related genes included CCL2 and TNFα (Supplementary Fig. 10, Supplementary Data 10). Notably, the inflammatory response was also one of the key GO hits for the broader set of genes that are more highly expressed in Secretory (Supplementary Data 4). These results suggest that inflammation-related genes may be triggered, at least partially, by a cell-intrinsic manner in LUSC.

Consistent with these findings, compared with KLN205, LN4K1 had markedly elevated TNFα and CCL2 expression levels and modest increases in CSF1 (Fig. 4c, d). Compared with KLN205 and a metastatic LUAD line (344SQ), LN4K1 secreted abundant levels of CCL2 in vitro (Fig. 4e), which was confirmed in the plasma of tumor-bearing mice (Fig. 4f). Our group previously showed that TNFα-mediated activation of the canonical NFκB pathway directly promotes CCL2 expression. Indeed, silencing p65 significantly decreased basal and exogenous TNFα-mediated stimulation of CCL2 in LN4K1 cells (Fig. 4g). Moreover, compared with KLN205, LN4K1 displayed increased activation of NFκB, and pharmacological inhibition of IKKβ using Compound A (Cmpd A) significantly reduced the levels of phospho-p65 and CCL2 (Fig. 4h).

Next, to characterize the immunologic changes elicited by the LN4K1 model, we performed flow cytometry on different immune populations of mice injected with this cell line versus healthy, non-tumor-bearing mice. The LN4K1 model promoted substantial increases in IM generation in the bone marrow (BM), leading to a significant increase in IMs in the blood and a non-significant increase in the spleen (Fig. 4i). The lung TME was characterized by significant increases in IMs, TAMs, and Tregs (Fig. 4i), as well as neutrophils, natural killer cells, and conventional CD4 and CD8 cells and a non-significant increase in DCs.

CCL2 is necessary and sufficient for enhanced LUSC metastasis. Considering that metastasis accounts for approximately 90% of cancer-related deaths, there is a surprising paucity of scientific knowledge concerning the specific mechanisms that drive LUSC metastasis. Arguably one of the least understood steps in the metastatic cascade occurs after cancer cells successfully intravasate into the circulation. It is now well recognized that distant colonization is an extremely inefficient process, and most cancer cells that reach distant tissues rapidly undergo apoptosis.

Consistent with increased secretion of the CCL2 chemokine (Fig. 4i), compared with the parental KLN205 cell line, LN4K1 was associated with marked increases in IM recruitment and rapid development of distant metastases (Supplementary Fig. 11). Given the robust immunologic response that LN4K1 invokes in the lung TME, we addressed whether the CCL2-mediated recruitment of IMs is necessary and sufficient for distant metastasis development in LUSC using an experimental metastasis model. Indeed, the stable overexpression of CCL2 in LN4K1 was sufficient to account for the enhanced metastatic properties of LN4K1, while the silencing of CCL2 in LN4K1 with two different shRNAs had the opposite effect, substantially decreasing metastatic properties (Fig. 5a–c). Additionally, consistent with the effects of CCL2 on metastasis and IM recruitment, CCL2 overexpression in LN4K1 led to significantly shorter survival and increased IM infiltration in the lungs (Fig. 5d, e). Conversely, silencing CCL2 in LN4K1 dramatically extended survival, which corresponded with significantly decreased IM infiltration (Fig. 5f, g). To assess how robust CCL2-mediated recruitment of IMs is on LUSC progression, we developed an additional model of LUSC metastasis. Using the parental KAL cell line, which was derived from a kinase-dead IKKa genetically-engineered mouse (GEM) model of LUSC, we performed two rounds of in vivo selection as described for the LN4K1 model (Supplementary Fig. 7c). With this approach, we developed the KAL-LN2E1 metastatic sub-clone, which forms large, poorly differentiated orthotopic LUSC tumors and rapidly develops lymph node and chest wall metastases (Supplementary Fig. 12a, b). To corroborate our findings with the LN4K1 model, we assessed whether IMs contribute to LUSC metastases independent of their suppressive role on T-cells. We generated stable KAL-LN2E1 lines expressing CCL2 shRNA hairpins and orthotopically injected them into NSG mice (Supplementary Fig. 12c). While no effect was observed on primary tumor development (not shown), both groups injected with shCCL2 KAL-LN2E1 lines showed decreased numbers and incidence of distant metastases, consistent with decreased IM recruitment in the primary tumor (Supplementary Fig. 12d–f).
Development of a molecular strategy for targeting LUSC IMs. We then assessed the therapeutic efficacy of targeting IMs in our metastatic LUSC model. While anti-CCL2 antibodies have shown initial promise in breast cancer, a rebound effect that accelerates metastasis has been observed upon drug cessation and this therapy is no longer being clinically developed. Furthermore, other chemokines such as CSF1 and CCL3 can have redundant properties in recruiting IMs and TAMs. Thus, targeting IMs requires a more effective strategy. Recent studies with a CCR2 inhibitor (PF-04136309) have demonstrated effective blockade of IM recruitment in pancreatic cancers. To assess the effects of PF-04136309 on established LUSC metastases, one week following tail vein injection of luciferase-labeled LN4K1 cells, mice were treated with vehicle or PF-04136309 and imaged one week later (Fig. 5h). Significant reduction in lung metastasis was observed with PF-04136309 treatment, consistent with significant reductions in both circulating and lung TME IMs (Fig. 5h–k, Supplementary Fig. 13). However, there were no significant changes in…
TAMs, DCs, NKs, or Tregs (Fig. 5). In accordance with a prior study using CCR2 inhibition, 1 week of treatment did not significantly affect CD206<sup>High</sup> or CD206<sup>Low</sup> TAM subsets (Supplementary Fig. 14a). Additionally, we found no significant effects on recruitment of gamma-delta T-cells (Supplementary Fig. 14b), which express CCR2 and are involved in IFNγ production<sup>13</sup>. These results strongly suggest that the therapeutic effects arose from blocking IM recruitment into the TME. Additionally, CCR2 blockade with PF-04136309 significantly inhibited metastasis development when treatment was initiated on the day of cancer cell injection (Fig. 5m).

**IMs cause FXIII<sub>A</sub>-mediated fibrin cross-linking and LUSC progression.** We next sought to investigate the molecular mechanism by which IMs promote LUSC metastasis. Over the past decade there has been rapid growth in our understanding of how perivascular macrophages promote tumor growth and metastasis<sup>13</sup>; however, the contribution of IMs prior to differentiating into macrophages remains poorly understood. Given the recent findings that RMs have opposing, anti-metastatic functions when compared with IMs<sup>17</sup>, we hypothesized that divergent expression patterns of IMs and RMs may reveal important differences in their biology. In a previously performed transcriptome profile of IMs and RMs in the blood and spleen<sup>17</sup>, we observed that F<sub>3</sub>a<sub>1</sub> had the sharpest differential expression pattern (expression in IMs > RMs) for both tissue compartments. F<sub>3</sub>a<sub>1</sub>, which encodes for factor XIII-A subunit (FXIII<sub>A</sub>), cross-links fibrin and other substrates and has critical roles in blood clot stabilization and wound healing<sup>42</sup>. Compared with RMs, we confirmed that IMs express markedly increased levels of F<sub>13</sub>a<sub>1</sub> mRNA (Fig. 6a). Importantly, comparing the expression levels of F<sub>13</sub>a<sub>1</sub> mRNA in CD206<sup>High</sup> and CD206<sup>Low</sup> TAM subsets revealed that TAMs express F<sub>13</sub>a<sub>1</sub> levels that are similar to RMs (Supplementary Fig. 15). We also found a strong level of positive correlation of expression among F<sub>13</sub>a<sub>1</sub>, CD14, and our IM-rich subset at the transcriptional level (Supplementary Fig. 16a).

Neither VEGFa nor the well-validated myeloid-derived suppressor cell (MDSC) mediators Arg1 and NOS2 exhibited such a correlation with CD14 or the IM-rich subset (Supplementary Fig. 16a). Furthermore, compared to tumors with low CD14 expression, F<sub>13</sub>a<sub>1</sub> expression levels in CD14<sup>High</sup> tumors were more than 2.5-fold higher (Supplementary Fig. 16b). In agreement with the Secretary subtype frequently having high CD14 expression (Fig. 2b, Supplementary Fig. 16a), we found that this subtype also associated with above median expression of F<sub>13</sub>a<sub>1</sub> (Supplementary Fig. 16c). These findings implicate IMs as a rich source of F<sub>13</sub>a<sub>1</sub> in the tumor microenvironment of LUSC.

Immunofluorescent staining revealed that FXIII<sub>A</sub> protein is produced at higher levels in IMs than RMs (Fig. 6b), and FXIII<sub>A</sub> localizes with CD11b near the cell surface (Fig. 6c). Using confocal microscopy, we observed dense deposits of FXIII<sub>A</sub> near podosome-like structures in IMs (Fig. 6d, Supplementary Video).

Using IM and RM cell densities observed in the TME of our LN4K1 model, we found that IMs induced fibrin cross-linking (γ-δ formation) when added to FXIII<sub>A</sub>-depleted (Peak 1) fibrinogen, and this activity was fully inhibited with a FXIII<sub>A</sub> inhibitor, T101 (Fig. 6e). In contrast, only subtle cross-linking was seen with even high RM densities (Fig. 6e).

Previously, others have shown that cancer cells can utilize cross-linked fibrin to form invadopodia<sup>43</sup>. To test the contribution of FXIII<sub>A</sub> activity to tumor cell function, we interrogated LN4K1 cell invadopodia formation. We found that LN4K1 cells could easily form invadopodia when grown in unfractionated fibrinogen (which contains FXIII<sub>A</sub>); however, this was significantly attenuated when placed in FXIII<sub>A</sub>-depleted fibrinogen (Fig. 6f). We then hypothesized that IMs in the TME provide the necessary FXIII<sub>A</sub>-activity to cross-link fibrin and create a scaffold for cancer cell invasion. Using a co-culture model of GFP-labeled LN4K1 cells and freshly isolated IMs, we observed that both the low and high densities of IMs could rescue invadopodia formation in FXIII<sub>A</sub>-depleted fibrinogen to a degree similar to that of unfractionated fibrinogen (Fig. 6g). T101 treatment completely abolished this effect, implicating FXIII<sub>A</sub> activity in this mechanism (Fig. 6g). Next, we determined whether the increased cancer cell invadopodia formation induced by FXIII<sub>A</sub>-expressing IMs also corresponded with increased LUSC invasion. Using trans-well invasion chambers, we observed a significant reduction in the invasive capabilities of LN4K1 cells when grown in FXIII<sub>A</sub>-depleted fibrinogen compared with unfractionated fibrinogen (Fig. 6h). Similar to the invadopodia assays, this phenotype in FXIII<sub>A</sub>-depleted fibrinogen was significantly rescued in the presence of IMs and abolished in the presence of T101 (Fig. 6h). Importantly, the effects of T101 on invasion were not seen when performed in Matrigel (Supplementary Fig. 17), suggesting the importance of the fibrin cross-linking context. Finally, to assess the importance of FXIII<sub>A</sub> expression in IMs for promoting cancer cell invasion, we performed the invadopodia assay using IMs from age-matched wild-type or F<sub>13</sub>a<sub>1</sub> knock-out mice. Consistent with our prior findings, the presence of wild-type IMs led to substantial increases in LN4K1 cells with invadopodia formation, while co-culture with F<sub>13</sub>a<sub>1</sub><sup>-/-</sup> IMs completely abolished this phenotype (Fig. 6i).

To assess FXIII<sub>A</sub> activity within the LUSC TME and its association with disease progression and metastasis, we developed
an IHC protocol with a novel monoclonal antibody that specifically detects the cross-linked fibrin neo-epitope (Supplementary Fig. 18a). Using a TMA of 96 surgically-resected LUSC tumors, we found that compared with low or intermediate staining, high staining of intra-tumoral fibrin cross-linking was associated with significantly worse recurrence-free survival (Fig. 7a, Supplementary Fig. 18b). Consistent with this observation, LUSC tumors from the TCGA dataset expressing high F13a1 had significantly worse survival (Fig. 7b). Finally, to assess whether FXIIIA over-expressing monocytes are sufficient to enhance LUSC metastasis in vivo, we stably over-expressed silenced FXIIIA in a THP1 monocyte model (Fig. 7c). Genetically modified THP1 monocytes were infused into NSG mice daily for a total of 4 days following intravenous injection of LN4K1, and at 1 week the lungs were dissociated and micro-metastases were enumerated using FACS for EpCAM. Compared with LN4K1 alone injected mice, the only group with significantly increased metastases were the mice treated with THP1-F13 ORFs, while neither F13 shRNA-expressing groups showed an increase in metastases (Fig. 7d).

Taken together, we have uncovered a previously unappreciated ‘IM-rich subset’ of LUSC that is driven by a TNFα-NFκB-CCL2 axis of IM recruitment. These IMs express high levels of FXIIIA, which facilitates LUSC cell invasion and disease progression by promoting fibrin cross-linking (Fig. 8).

**Discussion**

PD1/PD-L1 immune checkpoint inhibitors, while effective in several cancer types, provide only about 20% response rates in unselected LUSC patients. Thus, there is an urgent need to extensively characterize other immunologic mediators of LUSC progression, which may unveil logical, non-overlapping combination approaches to treat this disease.
and aDCs. This finding was surprising given that IMs are traditionally thought of as inactive precursors that develop into macrophages, DCs, and MDSGs in response to external cues. TAMs have been studied in many syngeneic cancer models and have been associated with the promotion of metastasis. However, how IMs mechanistically contribute to the establishment of metastasis remains poorly understood. Reasonable hypotheses are that the CCL2-elicited IMs may play similar roles to TAMs by providing both a pro-tumor, pro-metastatic microenvironment through the secretion of growth factors as well as by being highly immunosuppressive, thus preventing an anti-tumor immune response. The surface flow markers used to identify IMs in this and other studies may also identify monocyte myeloid-derived suppressor cells (m-MDSCs), a subset of bone marrow elicited myelo-monocytic cells that are closely related if not identical to IMs; the field is still debating if there are subtle differences in gene expression or suppressive capabilities between IMs and m-MDSCs. When studied functionally, tumor-localized m-MDSCs are highly immunosuppressive, particularly against cytotoxic T-cells, largely through their secretion of iNOS and Arginase 1. Along these lines, a recent review discussing MDSC nomenclature and characterization suggests that IMs can be attributed to m-MDSCs. However, our results suggest that IMs actively participate in the LUSC metastatic process and may have unique functions in the TME independent of this immunosuppressive phenotype. In support of this possibility, we found no significant association between iNOS or Arginase 1 and our ‘IM-rich subset’ of LUSC tumors.

Utilizing immunogenomic approaches, we found that the ‘IM-rich subset’ can be used to re-classify the four LUSC subtypes using median CD14 expression as the threshold. To further evaluate these findings, we developed a novel immune-competent metastasis model of LUSC. This model is characterized by IMs recruited by LUSC tumor cells through TNFa-mediated activation of NFkB signaling, which promotes the secretion of the monocyte chemokine CCL2. Integrated network analyses of genes linked to worse survival for LUSC also revealed enrichment for activation of a TNFa-NFkB-CCL2 signaling axis. Although TNFa within the TME is often derived from cellular constituents of the TME itself, we observed marked elevations of TNFa in our metastatic LN4K1 sub-clone developed through serial in vivo passages. This finding suggested that metastatic LUSC sub-clones may secrete elevated levels of TNFa, and at least during the initial stages of metastasis, LUSC cells may play an autocrine role in the heightened secretion of CCL2 with subsequent recruitment of IMs. Because IMs and IM-derived TAMs also secrete copious amounts of TNFa, our data imply a feed-forward loop whereby LUSC tumors promote their own secretion of CCL2 followed by the recruitment of TNFa-secreting IMs.

After determining that CCL2 is necessary and sufficient to promote distant LUSC metastasis, we evaluated a therapeutic strategy for targeting IMs. Specifically, we targeted the main CCL2 cell surface receptor, CCR2, using a potent, clinic-ready CCR2 inhibitor (PF-0436309). This strategy significantly reduced both the blood and tumor levels of IMs, which inhibited the seeding and initial growth of LN4K1 metastasis and prevented the progression of established LUSC metastases. This approach unambiguously shows the potential of targeting this immune cell type for the treatment of LUSC patients.

Our study of an "IM-rich subset" of LUSC is reminiscent of the paradigm that tumors resemble "wounds that do not heal." The main tenets of how tumor-induced "wounds" form are as follows: (1) VEGFa promotes angiogenesis and increased vascular permeability, (2) this in turn leads to extravasation of fibrinogen and several classes of lymphocytes, (3) activation of the coagulation cascade occurs, (4) fibrin deposition takes place, and (5) an irregular collagen matrix forms. Although this process is analogous to physiologic wound healing, tumor-induced "wounding" is usually irreversible and leveraged by cancer cells to effectively paralyze the host organ. Along these lines, we found that IMs express high amounts of Factor XIIA, which rapidly and potently leads to fibrin cross-linking in the TME, and this evidence points to a novel and important mechanism of LUSC progression. Consistent with these findings, plasma levels of fibrin degradation products (D-dimers) have been linked with higher stage lung cancers and markedly worse prognosis. Moreover, plasma FXIIA has been shown to promote cancer by inhibiting the intravascular clearance of natural killer cells. Importantly, none of these prior studies have suggested a role for cellular FXIIA in IMs for tumor progression. In agreement with observations that FXIIA has pleiotropic roles in mediating wound healing, our findings point to a previously unappreciated mechanism by which tumors represent "non-healing wounds".

IM-derived FXIIIA-mediated fibrin cross-linking creates an important scaffold for cancer cell invasion. Targeting this IM function in an immunocompetent LUSC model had substantial effects on blocking metastases. Moreover, dense intra-tumor deposits of cross-linked fibrin in resected LUSC tumors were associated with poor survival. Given the rapidly evolving landscape of precision immune-oncology, these findings identify IMs as a novel context-specific vulnerability of LUSC and provide an important insight into the mechanisms through which this immune cell type determines a poor prognosis.

Methods
Cell lines and maintenance. All cell lines were maintained in 5% CO2/95% air at 37 °C. KLN205 lung squamous cell carcinoma cells were obtained from the ATCC, parental KLN cells were kindly provided by Dr. Yinling Hu (National Cancer Institute, Frederick, MD) and 344SQ lung adenocarcinoma cells were kindly provided by Dr. John Kurie (M.D. Anderson Cancer Center, Houston, TX). THP1 monocytes were kindly provided by Dr. Gianpietro Dotti (University of North Carolina, NC). KLN205 cells and derived sub-clones (LN2-2 and LN4K1) were maintained in MEM and 344SQ cells were maintained in RPMI 1640, both supplemented with 10% fetal bovine serum (FBS) and 1% Penicillin Streptomycin. KLN cells and their derivatives were maintained in DMEM supplemented with 10% FBS. Mouse bronchial epithelial cells (MBECs) were isolated from three healthy adult C57Bl/6 mice. All cell lines were tested to confirm the absence of Mycoplasma, and all in vitro experiments were conducted with 60–80% confluent cultures.

Establishment of sub-clone cell lines. Metastatic lesions were mechanically minced in RPMI 1640 containing 0.125% collagenase III and 0.1% hyaluronidase under laminar flow within a hood and using a sterile scalpel. Cells were then pelleted, resuspended in 0.25% trypsin for 20 min at 37 °C with vortexing every 5–7 min, and placed in a T75 flask with complete MEM medium.

Lentivirus Packaging and Infection. Lentiviral particles for CCL2 overexpression in KLN205 cells were purchased from GeneCopeia: Scr ORF (pReceiver-Lv152 Negative Control Lentifect Purified LV Particles) and CCL2 ORF (CCL2 NM_011333.3 Lentifect Purified LV Particles), Lentiviral vectors for CCL2 knockdown in LN4K1 and KLN-LN2E1 cells were also purchased from GeneCopeia: Cntil shR (SHCHR0001-1-LVRU6H), CCL2 shR#1 (MHS030124-1-LVRU6H), CCL2 shR#2 (MHS030124-2-LVRU6H), CCL2 shR#3 (MHS030124-3-LVRU6H), and CCL2 shR#4 (MHS030124-4-LVRU6H). Lentiviral particles for Factor 13 overexpression and silencing in THP1 cells were purchased from GeneCopeia: Factor 13 ORF (NM_020784.3), Factor 13 shR#1 (HS050569-1-LVRU6H) and Factor 13 shR#2 (HS050569-2-LVRU6H). OgNLuc vector was a kind gift from Dr. Antonio Amelio (Lineberger Comprehensive Cancer Center; UNC Chapel Hill, NC). Lentivirus was produced by transfecting human embryonic kidney cells (293 T) with lentiviral vector, packaging plasmid (pSPAX2) and envelope plasmid (pMD2.G). Media was changed the next day, and 2 days later viral supernatant was collected and filtered to remove cellular debris. Cells were infected with lentiviral particles overnight using Polybrene and were then selected with growth medium containing 200 µg/mL hygromycin (for shR and ORF lentivirus for each respective cell line) and 1 µg/mL puromycin (for OgNLuc lentivirus).

Animals, in vivo models and tissue processing. Adult C57Bl/6, NGS and FVBn mice were purchased from Jackson Labs. Factor XIIa knock-out and wild-type C57Bl/6 mice were obtained from Dr. Alisa Wolberg. These animals were cared for...
according to guidelines set forth by the American Association for Accreditation of Laboratory Animal Care and the U.S. Public Health Service policy on Human Care and Use of Laboratory Animals. All mouse studies were approved and supervised by the University of North Carolina at Chapel Hill Institutional Animal Care and Use Committee. All animals used were between 6–10 weeks of age at the time of injection. For all animal experiments, cells were trypsinized, washed and resuspended in Hank’s balanced salt solution (HBSS; Gibco) prior to injection. Lung squamous cell carcinoma (LUSC) samples (380 total) (i) for which clinical information had been published by TCGA, GDAC (Genome Data Analysis Centers) data hub (https://gdac.broadinstitute.org/) and the Firehouse Broad GDAC (Genome Data Analysis Centers) data hub (https://gdac.broadinstitute.org/) and checked for mutual consistency. For all the analyses based on RNA-Seq data, the only samples used were those having numerical data available (on January 2014) as for messenger RNA (mRNA), human methylation (HuMe), copy number variation (CNV) and micro-RNAs (miRNAs) (348 total), in order to exclude samples not sufficiently characterized by TCGA. For maximizing the statistical power of our subtype-based (see section “Assignment of tumor subtypes”) survival analyses of LUSC patients (Fig. 1), we broadened this number to include all samples (380 total) (ii) for which clinical information had been published by TCGA, with available mRNA sequencing (see section “Assignment of tumor subtypes”) and (iii) having usable survival data (see the next section).

Survival data pre-processing. We extracted from the clinical annotation files of LUSC samples (downloaded on January 2014) four types of survival data: (1) "Days to death", (2) "Days to last follow-up", (3) "Days to last known alive", and (4) "Vital status". The processing was the following: (i) disregarding patients (5 total) having a negative value for their "Days to last follow-up" (since negative values are incompatible with survival analyses and this data conflict could not be solved differently at a dose of 200 mg/kg/mouse, twice daily by oral gavage; (ii) the two subtypes with the smallest and largest median, the relative variation between CB and PS is

Assignment of tumor subtypes. Normalized RZM (abbreviation for RNA-Seq by Expectation Maximization) gene expression values of lung squamous cell carcinomas (n = 491 tumors) were obtained from TCGA for each tumor and for the four squamous subtypes (Classical, Basal, Primitive, Secretory). For every sample, the subtype assignment was defined as a tumor’s largest correlation value (Supplementary Data 11).

Analysis of RNA-Seq data: definition of a spectrum of gene expression among the subtypes. We adopted the normalization method that was chosen by TCGA Consortium for RNA-Seq data (RNA-Seq V2 pipeline), based on data quantified with the reads per kilobase of transcript per million mapped reads (RPKM) method. These files are 280 of the RNA-Seq data published by TCGA Data Portal). Then, subtypes were aggregated based on two criteria: (i) correlation among the expression values of their samples and (ii) number of samples available. Specifically, the correlation was evaluated as follows: (a) the expression values of all genes of this platform are averaged inside each subtype. For each of this pair, the vectors obtained are \( V_p, V_C, V_B, \) and \( V_S \), for the subtypes Basal, Classical, Primitive and Secretory, respectively, each having 20,531 elements; (b) Pearson’s correlation coefficients are calculated among all possible (i.e., six) couples of these four vectors; (c) the highest and mutually exclusive Pearson’s correlation coefficients are used for each couple; and (d) the closest similarity with the Classical subtype and the same happens between the Primitive and Secretory subtypes. Then, the information about the number of samples per subtype was used for deciding in which order they would be displayed in each heat map (HM), differently specified, so that the two subtypes with the highest number of samples were plotted at the two HMs sides and the two subtypes in the middle: at this stage, it is defined that the first potential aggregate of subtypes is Classical-Basal (CB, left side of the HM) and the second is Primitive-Secretory (PS, right side of the HM).

Analysis of RNA-Seq data: identification of a set of representative genes. In this hybrid analysis, genes were considered (i) differentially expressed between CB and PS, (ii) heterogeneously expressed across the available samples and (iii) quantitatively satisfactory when: (a) their median has an appreciable change between CB and PS samples. Specifically, after determining which, between CB and PS samples, has the largest and smallest median, the relative variation between CB and PS is calculated with respect to the smallest between these two values (precisely, (maximum-minimum)/(minimum + MATLAB epsilon (eps function))). This ratio expresses the relative variation of a gene and the median for a gene and thus, when \( \frac{\text{maximum}}{\text{minimum}} \) is chosen that variation strictly belongs to the top 50% across all genes; (b) they have a "calculable" p-value according to the MATLAB implementation of the Wilcoxon rank sum test (i.e., a p-value different from ‘Not a Number’ (NaN), which is the output of the MATLAB function performing this statistical test when all values of the first and second group are identical) < 0.01 and, at the same time, whose p-value is considered 'relevant' for multiple hypothesis testing (the only genes included are those having a number of values ≥ 0.5, across the four subtypes, > 60% of the total sample number (therefore ≥ 209, since there are 348 samples in this dataset); (c) their associated q-value (according to Storey’s method6), for "corrected" and "relevant" p-value < 0.01; (d) their fold-change greater than the first octile value of the range of set median values of the full gene list; (e) their standard deviation strictly belongs to the top 50% across all genes.

Analysis of RNA-Seq data: genomic dissimilarity among subtypes. After performing the gene selection above described, it was assessed that the "genomic dissimilarity" for RNA-Seq data between the two subtypes located at the two sides of this heat map (i.e., Classical and Secretory) is the maximum among the six possible cases (one-to-one comparisons among four subtypes). This genomic dissimilarity was measured as follows: (1) for each gene g, and each subtype \( i \), we calculated the mean, named \( m_g(i) = m_{gi} \). So, for g that belongs to (Classical (C), Basal (B), Primitive (P), Secretory (S)), the four vectors considered are: \( \{ m_{gi}, m_{gj}, m_{gk}, m_{gl} \} \) with \( 1 \leq i \leq 4291 \), since the heat map of hybrid differential expression has 4291 genes; (2) the six possible city-block distances between the two subtypes were plotted at the HM. For each of the other two, if the distance between \( \{ m_{gi}, m_{gj}, m_{gk}, m_{gl} \} \) and \( \{ m_{k1}, m_{k2}, m_{k3}, m_{k4} \} \) is: \( d([m_{gi}, m_{gj}, m_{gk}, m_{gl}]) = \Sigma_{i = 1}^{4} s_{i + s_{j} + s_{k} + s_{l}} \) the maximum among these six city-block distances, which corresponds to the couple Classical-Secretory, was used to establish that these two subtypes have the largest genomic dissimilarity. This analysis confirmed that, also in this data subset: (a) the closest subtype to Classical, using this subtype-to-subtype distance, is Basal, and (b) the closest subtype to Secretory is Primitive.

Analysis of RNA-Seq data: visualization of the clustered expression matrix. Selected genes were log2 transformed, mean centered, hierarchically clustered (similarity metric: correlation (uncentered), clustering method: average linkage) using the version 3.0 of Cluster (open source clustering software)25,26 and visualized with the Java-based program TreeView24, in order to assess similarities and differences among their expression patterns. Here and in other heat maps utilized, genes with a null normalized expression value are log-transformed and, therefore, are assigned to a Not-a-Number (NaN) value, which TreeView displays...
as a gray rectangle. The hierarchically clustered expression matrix, with respect to the genes (matrix rows) has 4,291 rows and 348 columns.

Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix. Keeping the four LUSC subtype in this order (i.e., (1) Classical, (2)Primitive, (3) Squamous, (4) Nontypical), from left to right, there is an interesting computational feature: namely, after splitting the matrix of gene expression into two halves (each with 2145 genes, so excluding the gene # 2146, which is exactly in the median position) and (ii) calculating the average of these means for the first half and the second half, separately, for each subtype, it turns out that in the upper half this value grows from left to right, while in the lower half it drops from left to right. This defines a clear gradient of gene expression and, for this reason, as well as for the genomic dissimilarity assessments, the above-mentioned subtype order was permanently selected. For this reason, these data for computational tasks, each independent of the other, the sample order inside each subtype was defined at the beginning of our analyses according to the sequence found in the original TCGA files (that relies on the TCGA sample barcodes) and to the availability of sample subtypes. Notably, this sample order was changed only in our analyses concerning the immune signature (see section “Analysis of the immune cell type: scoring of different infiltrates”). In order to perform a batch analysis of survival data (see section “Survival analysis based on tumor subtypes, clinical stage and expression levels for selected genes”) for selected genes associated with hyper-expression (see below) in PS, we focused on the intermediate portion of the HM (precisely, from gene 1900 to 2400) and looked for a gene that could be algorithmically defined as the last of the upper HM portion. Specifically, for every gene we (1) assessed the percentage of samples hyper-expressing that gene (i.e., having an expression, for that gene, ≥ mean across the 348 samples) per subtype, (2) calculated the average of these two percentages for CB (Average(%C, %B)) and PS (Average(%P, %S)), (3) determined the difference between these two values (Δ(CB, PS) = Average(%C, %B) - Average(%P, %S)). Since when 1900 ≤ (gene order # in HM) ≤ 2273 it follows that Δ(CB, PS) ≤ 0 and when 2274 ≤ (gene order # in HM) ≤ 2400 it follows that Δ(CB, PS) > 0, except for patterns of maximum three consecutive genes that do not meet the inequality requirement for Δ(CB, PS) in the defined upper portion of the HM (see section “Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix”) in the Secretary subtype (more precisely, of the genes shown in the portion of the RNA-Seq HM defined as “upper” in section “Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix”) and also have statistically significant p-values (< 0.05) for overall survival analysis, we considered “genes of interest” as genes which are significantly up-regulated in the HM Top portion as genes characterizing the Secretary subtype, and to genes significantly up-regulated defined in the HM Bottom portion as genes of the Classical subtype. The GO analysis that was selected relies on the Expression Analysis Systematic Explorer (EASE) score (a p-value obtained through an adjusted Fisher’s exact test) and was performed using DAVID Bioinformatics Resources; the selected background was “Hom sapiens”. Each GO category was considered for further analyses only when fulfilled these three criteria: (1) it is referred to GO biological processes (BP); (2) it has two or more gene members inside the list of genes of the Top HM portion; (3) has a p-value < 0.001.

Ingenuity pathway analysis of RNA-Seq data. Using ingenuity pathway analysis (IPA) (http://www.ingenuity.com/products/ipa) of the genes (n = 403) that are generally hyper-expressed (see the deviation of hyper-expression given in section “Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix”) in the Secretary subtype (more precisely, of the genes shown in the portion of the RNA-Seq HM defined as “upper” in section “Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix”) and also have statistically significant p-values (< 0.05) for overall survival analysis (see section “Batch survival data analysis based on gene levels”), we assessed for the most significant biological themes in terms of “Disease or Function”. The displayed network is focused on the most significant biological function (“Leukocyte Migration”) and shows all genes belonging to it. The levels of statistical significance for these genes (in terms of survival analysis) are displayed on a color scale where 0 least significant, red: most significant; genes are clustered (using colored ovals) according to specific leukocyte categories (e.g. IMs, Macrophages) based on previously described immune subset markers. Based on the same gene set (n = 403), we also utilized IPA to perform the “upstream regulator analysis” (URA) to identify which were the most significant upstream regulators inside this gene network. We utilized the CytoTools web interface for the CytoHubba results analysis.

Survival data analysis based on tumor subtypes, clinical stage and expression levels for selected genes. The survival of distinct experimental groups according to subtype, clinical stage and expression levels for selected genes was assessed using the log-rank test (a.k.a. Mantel–Cox test) (24). The ANOVA test of point 2 was used assuming its robustness as for the normality requirement and checking the data homogeneity. For this computational step, we exploited the relatively large number of subjects of the TCGA LUSC dataset. These results were confirmed also using the Spearman’s rank correlation coefficient. Finally, the p-values of point 1 (two sets of p-values) and of the genes CD14, CCL2, CCL3, CSF1, and TNFa were assessed using non-parametric tests (Mann-Whitney test) and were included in the batch analysis of survival (see section “Analysis of RNA-Seq data: definition of the upper and lower portion of the clustered expression matrix”). The displayed network is focused on the most significant biological function (“Leukocyte Migration”) and shows all genes belonging to it. The levels of statistical significance for these genes (in terms of survival analysis) are displayed on a color scale where 1 least significant, red: most significant; genes are clustered (using colored ovals) according to specific leukocyte categories (e.g. IMs, Macrophages) based on previously described immune subset markers. Based on the same gene set (n = 403), we also utilized IPA to perform the “upstream regulator analysis” (URA) to identify which were the most significant upstream regulators inside this gene network. We utilized the CytoTools web interface for the CytoHubba results analysis.

Batch survival data analysis based on gene levels. A broader screening of survival values for differentially expressed genes in the above-mentioned LUSC subtypes (see section “Analysis of RNA-Seq data: identification of a set of representative genes”) was achieved through an in-house MATLAB (https://www.mathworks.com/products/matlab) script that calculates the hazard ratios (HR) and log rank test p-values using the “coxphfit” function for each gene selected (see the script contained in the file survival_analysis.docx). This analysis, on a gene by gene basis, splits samples depending on their being above or below a specific threshold; then, HR are calculated for two sample populations, one defined as {Samples whose gene expression is ≥ median} and one defined as {Samples whose gene expression is < median}.

Modified immuneonome signature. Based on the paper of Bindea et al. (27), which describes “a compendium of mRNA transcripts” of genes whose expression is strongly associated with specific immune cell types, we put together a list of genes for which the following variables are known: (1) type of immunity (innate or adaptive), (2) immune cell type(s) (cell type(s) of the immune system characterized by gene expression), (3) gene alias (any alternative name used by adding selected genes of known function of the immune system and that were not already present in this file (IMs: CD14, CCL2, CCR2, CCL3, CSF1R, CSF1; M2
macrophages: TGFB, VEGFA, IL10, CD206, VCAM1, CD163, ICAMI, IL1RA, CSE3R, M1 macrophages: TL2R, TL4R, CD80, CD86, CCR7, CCL5, CXCL9, CXCL10. For each gene, we completed a list by adding other immune-related genes described by Charoentong et al.28, who build upon the work of Binde et al.27, for the following immune cell types: regulatory T-cells, activated DCs, myeloid-derived suppressor cells, neutrophils, and plasmacytoid DCs. The matching between gene identifiers of the immunome signature and of RNA-Seq data was based on Entrez gene identifier (see also https://www.ncbi.nlm.nih.gov/entrez-query); immune-related genes without a match based on these identifiers were discarded. Notably, at the end of this procedure, each gene of the modified immunome signature (598 total) belongs either to one or two immune cell types (572 genes have a unique and 26 have also a second immune cell type).

**Analysis of the immune cell types: scoring of different infiltrates.** Due to the importance of CD14 for the survival of LUSC patients (as an unfavorable prognostic factor), we generated a heatmap of the 29 immune markers, using a single-score method (see Materials and Methods section). This produced a scoring system, which determines the immune cell type density of each sample. For calculating immune gene signatures, but is preceded by the non-parametric semiparametric procedure (steps 1–4). For every sample, these ranks are averaged across the immune cell type markers (i.e., by columns of the expression matrix) of the chosen immune cell type, consistently with the previous point: (a) a cell type density score is computed for each sample by using these rank averages, (b) this procedure (a) assigns to each sample only median immune cell density scores as the analyzed immune cell types and (b) is independent of the preliminary CD14-based reordering of the samples, which is instead used in the analysis of the survival of the immune cell types. Additionally, this ranking procedure is unambiguous, with the only exception of two samples, which have the same normalized RSEM expression levels of CD14. However, the presence of these two samples does not introduce any relevant bias in the analyses here described, since both patients belong to the group that expresses CD14 and even share the same subtype (Classical). Altogether, this semiparametric procedure (steps 1–4) is similar to what was previously described for calculating immune gene signatures, but is preceded by the non-parametric steps 1 and 2. This approach is different from deconvolution methods such as CIBERSORT29 and TIMER29 because, in theory, with our algorithm, a sample can be relatively enriched with respect to the broadest range of immune cell types (from none to all). It also differs from scRNA-seq30; since it assesses the under- and over-representation of the genes that belong to a gene set working across the samples (i.e., relatively to them).

**Analysis of the immune cell types: computational visualization of the CD14 + infiltrates.** Individual genes of the 9 CD14 + populations (i.e., ADC, DC, IDC, IMs, M1, M2, Macrophages, MDSC and Neutrophils) were also hierarchically clustered, after the gene expression values were log2 transformed and median centered (see section “Analysis of RNA-Seq data: visualization of the clustered expression matrix”); genes belonging to two distinct immune cell types were used for the HMs of both.

**Analysis of the immune cell types: correlations among immune infiltrates and survival analysis based on the density of the CD14 + cell types.** Immune cell type density scores are used for two main purposes: (i) evaluating the level of correlation between the 29 immune cell types, in order to understand their coordinated biological action in these patients; (ii) calculating, for the 9 CD14 + populations, the different level of enrichment in the ranked list, in the ranked list. For our analysis, the ranking gave the highest priority to the genes of the PS aggregate. We considered statistically significant gene sets having a FDR < 0.05; this significance threshold is considerably lower than the value (i.e., 0.25) originally suggested by the Authors of the GSEA method and provides a high degree of selectivity. Among the selected gene sets, we performed a further refinement according to their relationships with the biological findings described in this article. In consideration of the gene expression gradient of the four subtypes from Classical to Secretory and for the sake of brevity, the word Classical is (extensively) used, in the figures of this article, as representative of CB, while the word Secretory represents the PS ensemble.

**Microarrays-based analysis of the gene expression of a mouse model of lung squamous carcinoma.** The Affymetrix Mouse Gene 2.1 ST Array (http://www.thermofisher.com/us/en/home.html) was used for measuring the gene expression of a normal murine bronchial epithelial cells (MBC), b) the KLN205 murine lung squamous (parental) cell line, and c) its sub-clone LN4K1. The CEL file processing was performed using the Affymetrix Expression Console; background adjustment, quantile normalization and summarization were accomplished using the Robust Multichip Analysis (RMA) algorithm22. Later, the set of RNA probes that are included in this array (41,345) was split into two distinct sets. The list of probes used for this analysis, together with the available Affymetrix annotation, is reported in Supplementary Data 12. After these bioinformatic steps, the differential analysis is further refined by a context-based computational procedure. Details of these steps are described in the Methods section. All gene expressions were selected when (i) are differentially expressed between MBC (4 replicates) and KLN205 (three replicates), or (ii) are differentially expressed between MBC and the metastatic-derived cell line LN4K1 (three replicates) or (iii) fulfill both (i) and (ii). At this stage, a gene is considered differentially expressed between two groups of replicates when a) its expression levels are strictly higher or lower than in the other group for each replicate used (combinatorially-based implied p-value: 0.0286); (b) the difference between the means of the two groups is ≥ 50%; c) its range is in the top 75% in the entire set of genes. The second procedure allows defining which genes follow a gradient of expression (growth or reduction) going from MBC to KLN205 parental to the sub-clone LN4K1 (hence moving from “normal” to “primary tumor” to “metastasis”), so that this expression gradient is sustained across all the replicates of the three experimental conditions. For each gene, the growing pattern (pattern A) is sub-divided into two sub-patterns, namely A-1 and A-2, which are not mutually exclusive. The sub-pattern A-1 is based on the following requirements (which are extensively described for more easily allowing assessing the level of overlap with the sub-pattern A-2): (i) all the LN4K1 samples have a strictly greater expression than all the KLN205 samples; (ii) all the KLN205 samples have a strictly greater expression than all the MBC samples; (iii) the average expression for the LN4K1 samples is strictly greater than the average expression for the KLN205 samples. The sub-pattern A-2 is based on the following requirements: (i) the expression of each LN4K1 sample is strictly greater than the average expression across all the samples considered (10 = 4 + 3 + 3); (ii) the expression of each KLN205 sample is strictly lower than the average expression across all the samples considered; (iii) the expression of each MBC sample is strictly lower than the average expression across all the samples considered. Similarly, for each gene, the dropping pattern (pattern B) is sub-divided into two sub-patterns, namely B-1 and B-2, which are also not mutually exclusive. The sub-
pattern B-1 is based on the following requirements (extensively described, as for A-1): (i) all the LN4K1 samples have a slightly lower expression than all the KLN205 samples; (ii) all the KLN205 samples have a strictly lower expression than all the MBECS samples; (iii) the average expression for the LN4K1 samples is strictly lower than the average expression for the KLN205 samples. The sub-pattern B-2 is based on the following requirements: (i) the expression of each LN4K1 sample is strictly smaller than the average expression across all the samples considered; (ii) the expression of each KLN205 sample is strictly lower than the average expression across all the samples considered; (iii) the expression of each MBECS sample is strictly higher than the average expression across all the samples considered.

Overall, a gene is selected when it differentially expressed and follows either gene pattern from the aforementioned group of genes, and when the standard deviation of expression across all the samples considered is not less than 50% of the highest standard deviation across these two sets of murine samples. Then, orthologous genes between mouse and human and that were defined as differentially expressed between MBECS and LN4K1 samples (see the previous section) were included in the matched list of TCGA LUSC differentially expressed genes. We then considered the percentage of the samples in the matched list of murine Secretory subtypes greater than 60%; (b) the Benjamini-Hochberg FDR 58 calculated on the Wilcoxon rank sum test p-values between classical and secretory (after passing the previous gene filter) was < 0.00005. The choice of these two subtypes was based on their highest level of genomic dissimilarity (see section “Analysis of RNA-Seq data: visualization of the clustered expression matrix”). Then, the corresponding murine genes of the MBEC and LN4K1 samples of the Affymetrix arrays were aligned to the human genes and displayed through two additional heat maps (for ratios of the medians above and below 1 as well), using the visualization style adopted in the main analysis of our mouse model of lung squamous carcinoma (see the previous section). Genes of these two groups were used for two separate GO analyses (see section “Gene ontology analysis of RNA-Seq data”), with “Homo sapiens” as background species and selecting GO-BP terms whose p-values were < 0.001 and, at the same time, containing at least 10 genes of either of these two groups.

ELISA assays. Murine CCL2 protein levels were quantified by ELISA using the Duoset Immunoassay kit (R&D Systems DY749-05 and DY7008) according to the manufacturer’s protocol. To assess secretion of CCL2 in vitro, 344SQ, KLN205, and LN4K1 were seeded at a density of 400,000 cells per well in 3 mL of media in 6-well plates. Supernatant was collected 48 h later and stored at −80 °C. For analysis of plasma CCL2 levels, blood (approximately 200 μL per mouse) from 3 to 5 mice per group was obtained 1-week prior to sacrifice via submandibular bleed using a Goldener lancet (2 mm). Blood was collected into Vacutainer Blood Collection Tubes with anti-coagulant. Tubes were centrifuged at 25,000g for 4 °C for 5 min, then plasma was collected and stored at −80 °C until assay. Samples were assayed in triplicate and data represents the mean concentration.

Proliferation assays. KLN205 and LN4K1 cells were seeded at a density of 25,000 cells per well in 6-well plates in triplicate and counted on a hemocytometer using a Trypan Blue counterstain.

Monocyte fibrin cross-linking protocol. Low (25k) and high (100k) density of monocytes were incubated with unfractonated or Peak 1 fibrinogen (Enzyme Research Laboratories, South Bend, IN), in the absence and presence of T101 (Zedira, Darmstadt, Germany) for 15 min at 37 °C. Clotting was triggered with thrombin (Enzyme Research Laboratories, South Bend, IN) and CAI (1 U/mL and 10 mM, final, respectively) and clot formation proceeded for 2 h. Samples were then treated in 50 mM Na bicarbonate-0.1% Triton X-100, 12°C for 1 h, diluted 120-fold in 6 × reducing SDS sample buffer (Baston Bioproducts, Ashland, MA), boiled, separated on 10% Tris-Glycine gels (Bio-Rad, Hercules, CA), and transferred to polyvinylidene difluoride membranes (Invitrogen, Carlsbad, CA). Membranes were blocked for 1 h at room temperature with Odyssey Blocker (LI-COR Biosciences, Lincoln, NE), incubated overnight at 4 °C with primary anti-human fibrinogen polyclonal antibodies (Clone A0080, Dako, Glostrup, Denmark), and then incubated with Alexa Fluor® 488 fluorescent-labeled anti-rabbit secondary antibodies (Immunoresearch, West Grove, PA) for 1 h at room temperature. Membranes were scanned on a GE Typhoon FLA-9000 (GE Healthcare, Piscataway, NJ) by a total number of cells in each of six fields.

Invasion assays. LN4K1 cell lines (50,000 cells) were mixed with 200 μg/mL Fibronectin, 2 mM CaCl2, and either 2 mg/mL unfractionated fibrinogen, Peak 1 fibrinogen, or BD Matrigel (total final volume of 50 μL) on top of a Boyden 8 μm migration chamber. Coagulation was induced with 2.5 U/mL thrombin by incubating at 37 °C for 15 min. Cultured cells were incubated with serum free MEM medium (100 μL), and MEM +10% FBS was used as a chemoattractant in the bottom chamber. Invasion was assessed 24 h later. For experiments using T101, a concentration of 50 μM was used.

Analysis of Nfkb-dependent CCL2 induction. LN4K1 cells were subjected to NF-κB subunit p65 silencing or IKKβ inhibition by transfecting with 25 nM p65-specific or non-specific control siRNA using Lipofectamine RNAiMAX (Invitrogen, Carlsbad, CA) or by treating with 5 μM Compound A, an IKKβ inhibitor, for 5 h. Twenty-four hours after siRNA transfection, the medium was replaced, after which the cells were incubated for an additional 24 h and then treated with 100 ng/ml recombinant TNF for 2 h. After all treatments, the cells were harvested in RNA lysis buffer and subjected to RT-qPCR analysis.

Western blotting. After treatments, cells were lysed by scraping in RIPA buffer (ThermoFisher, cat no. P8390) containing 1 mM PMSF, 1 mM NaVO3, 1 mM dithiothreitol, and 1 x protease inhibitor cocktail. Equal amounts of lysates (20–30 μg of total protein) were run on 10% SDS-PAGE gels, after which protein was transferred to nitrocellulose membranes (BioRad, Hercules, CA). Membranes were blocked in 5% BSA/Tris-buffered saline-Tween 20 (TBS-T) for one hour at room temperature prior to probing with primary antibodies overnight at 4 °C. Primary antibodies included anti- phospho-p65 (Ser 536, clone 9E11, #3303) and anti-p65 (clone D14E12, #8242) from Cell Signaling Technology (Danvers, MA), and anti-vinculin (clone hVIN-1, #9391) from Sigma. After probing with primary antibodies, membranes were washed three times in TBS-T and then probed with the appropriate horseradish peroxidase-conjugated secondary antibodies (anti-mouse (HRP-conjugated, #7130-03-003) or anti-rabbit (#111-035-003) from Jackson ImmunoResearch). Then, the membranes were washed four times in TBS-T and developed using Clarity Western ECL substrate (BioRad, #1705060). Membranes were visualized using a BioRad ChemiDoc MP system (BioRad, Hercules, CA).

Flow cytometry. Blood, bone marrow and tumors were collected for flow cytometry analysis. Lung tissues were washed and mechanically minced using a sterile scalpel in low glucose DMEM and digestion media (1 mL collagenase at 2 mg/ml, 100,000 cells) were mixed with 200 μl Fibronectin, 2 mM CaCl2, and either 2 mg/mL unfractionated fibrinogen, Peak 1 fibrinogen, or BD Matrigel (total final volume of 50 μL) on top of a Boyden 8 μm migration chamber. Coagulation was induced with 2.5 U/mL thrombin by incubating at 37 °C for 15 min. Clot-embedded cells were incubated with serum free MEM medium (100 μL), and MEM +10% FBS was used as a chemoattractant in the bottom chamber. Invasion was assessed 24 h later. For experiments using T101, a concentration of 50 μM was used.
cells to remove alveolar macrophages and/or eosinophils. IMs were identified as CD45 +/CD11b +/Ly6G−/Ly6C− cells, RMs as CD45 +/CD11b +/Ly6CLo−Ly6G−/Ly6C+ cells, and ANs as CD45 +/CD11b +/Ly6C−Ly6G−/Ly6C+ cells, for use as a biologic feed forward control. Immunohistochemistry was performed on tissue sections dewaxed in Bond reagent with antigen retrieval, 3% H2O2 was used to block the endogenous peroxidase activity for 10 min. Tumor sections (8 µm thickness). After deparaffinization, the epitope retrieval was followed with 5 min enzymatic de-waxing and cross-linking (Zedira GmbH, Darmstadt, Germany) and anti-D-Mimics (cross-linked fibrin), # A079 (Zedira GmbH, Darmstadt, Germany). Between the stains the appropriate antigen retrieval (10 min) and peroxidase blocking steps were inserted. Stained slides were counterstained with Hoechst 33258 (H3569, Life Technologies) and mounted with ProLong® Diamond Antifade Mountant (P36961, Life Technologies). Single stain controls were done for each of the primary and secondary antibodies to control for cross-reactivity between the antibodies did not occur. For single IHC stain D-Mimics antibody (1:1500) was applied for 30 min and detection was done using Bond Polymer Refine kit with 3,3′-diaminobenzidine (DAB) visualization and Hematoxylin counterstain (Dako). Stained slides were dehydrated and coverslipped. Positive and negative controls (no primary antibody) were included for IHC and IF stains. IHC were digitally imaged in the Aperio ScanScope XT (Leica Biosystems Inc., Norwell, MA) using 20 × objective. High resolution acquisition of CD14-CRCR2 CK IF slides in the DAPI, AF 488, Cy3 and Cy5 channels was performed in the Aperio ScanScope FL (Leica) using 20 × objective. Nuclei were visualized in DAPI channel (blue), CD14 in AF 488 (green), CK in Cy3 (cyan) and CCR2 in Cy5 (red). Using Aperio software, following color deconvolution, a previously described HSC scoring criteria using Aperio software was utilized to obtain H-scores for CD14 and cross-linked fibrin expression. For automated scoring of multiplexed images, slides containing fluorescently labeled TMA sections were scanned in the Aperio ScanScope FL (Leica Biosystems) using 20 × objective and images were archived in TPL’s eSlide Manager database (Leica Biosystems). Cytokeratin staining was used to digitally separate tissue cores into cytoplasmic positive and negative regions (Tissue Studio Composer; Tissue Section version 2.5 with Tissue Studio Library version 4.2; Definiens Inc., Carlsbad CA). Automated digital analysis of individual tissue cores was performed in these two regions, similarly to Tissue Studio Composer. The Nuclei and Simulated Cells algorithm in the IF Portal, was then used to detect and enumerate cells that co-expressed biomarkers of interest in the annotated regions. Briefly, nuclei were digitally detected by the presence of Hoechst stain (nuclear counterstain). From these nuclei, a cell simulation was performed – cells margins were grown out from nuclear boundaries. For this dataset, positivity thresholds for CD14 + and CCR2 + were determined by measuring the average staining intensities both inside and outside simulated cells. Only thresholds were set, the algorithm evaluated each cell individually for the presence of CD14 and CCR2. Cells that were negative for both markers or positive for CD14, CCR2 and both CD14 + and CCR2 were enumerated by the algorithm.

Immunocytochemistry. Cells were centrifuged at 1,000 rpm for 5 min in a Cytospin 3 (Shandon). Cells were then fixed with 4% PFA for 15 min and permeabilized with 0.25% Tween 20 in PBS for 15 min at RT. Protein blocking was done with 2% BSA and 0.25% Tween 1 for 1 h at RT. Slides were incubated with primary antibody for CD1b (rabbit, 1:100, Abcam ab133573) and/or FLX (sheep, 1:200, AAbbott Aase Labs (PA)). After washing, slides were incubated with appropriate secondary antibodies, goat anti-rabbit (Alexa Fluor 488) and/or goat anti-sheep (Alexa Fluor 549), diluted 1:500 in blocking buffer for 1 h at RT. Hoechst (1:10,000) was used for nuclear staining. Coverslips were mounted with Prolong Gold (Invitrogen). A Leica DM8 inverted microscope was used for fluorescence microscopy. A Zeiss 710 confocal microscope was used for confocal imaging. All image processing was done with Fiji software.

Statistical analysis for experiments and tissue microarrays. Between 5 and 10 mice were assigned per treatment group; this sample size gave approximately 80% power to detect a 50% reduction in tumor weight with 95% confidence. For this reason,
each group were compared using Student t test (for comparisons of two groups) and analysis of variance (for multiple group comparisons). For values that were not normally distributed (as determined by the Kolmogorov-Smirnov test), the Mann-Whitney rank sum test was used. A P-value less than 0.05 was deemed statistically significant. Aggregated data of Supplementary Figure 4 were analyzed through a two-tailed binomial test (CD206+ vs. CD206− cells in the population of CD14+ /CCR2+ cells), with an expected frequency of 0.5 for each ‘trial’, using R. All other statistical tests for in vitro and in vivo experiments were performed using GraphPad Prism 7 (GraphPad Software, Inc., San Diego, CA). The multiple hypothesis testing correction of these results was made using the FDR38.

Online content. Supplementary and Source Data are available in the online version; references unique to these sections appear only in the online version.

Data availability. The Affymetrix microarray data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) data bank, accession code GSE12585.

Received: 25 August 2017 Accepted: 25 April 2018

Published online: 18 May 2018

References

1. Jemal, A. et al. Cancer statistics, 2009. Ca. Cancer J. Clin. 59, 225–249 (2009).

2. Kamangar, F., Dores, G. M. A. & Anderson, W. F. Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. J. Clin. Oncol. 24, 2137–2150 (2006).

3. Kris, M. G. et al. Using multiplexed assays of oncogenic drivers in lung cancers to select targeted drugs. JAMA 311, 1998–2006 (2014).

4. Herbst, R. S. et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, non-small-cell lung cancer. N Engl. J. Med. 372, 1547–1555 (2015).

5. Brahmer, J. et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. N Engl. J. Med. 373, 1625–1635 (2015).

6. Reck, M. et al. Pembrolizumab versus chemotherapy for PD-L1-positive non-small-cell lung cancer. N Engl. J. Med. 373, 1823–1833 (2015).

7. Network, T. C. G. A. R. Comprehensive genomic characterization of squamous cell lung cancers. Nature 489, 319–525 (2012).

8. Wilkerson, M. D. et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin. Cancer Res. 16, 4864–4875 (2010).

9. Joyce, J. A. & Pollard, J. W. Microenvironmental regulation of metastasis. Nat. Rev. Cancer 9, 239–252 (2009).

10. Pollard, J. W. Tumour-educated macrophages promote tumour progression and metastasis. Nat. Rev. Cancer 4, 71–78 (2004).

11. Qian, B. Z. et al. CCL2 recruits inflammatory monocytes to facilitate breast-tumour metastasis. Nature 501, 23–27 (2013).

12. Bronte, V. Coordinated regulation of myeloid cells by tumours. Nat. Rev. Immunol. 10, 1650–1659 (2010).

13. Malik, G. et al. Plasma fibronectin promotes lung metastasis by contributions to fibrin clots and tumor cell invasion. Cancer Res. 70, 4327–4334 (2010).

14. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. Nat. Med. 19, 1432–1437 (2013).

15. Gabrilovich, D. I., Ostrand-Rosenberg, S. & Bronte, V. Coordinated regulation of myeloid cells by tumours. Nat. Rev. Immunol. 12, 253–266 (2012).

16. Kumar, V., Patel, S., Tyczynow, E. & Gabrilovich, D. I. The nature of myeloid-derived suppressor cells in the tumor microenvironment. Trends Immunol. 37, 208–220 (2016).

17. Kim, S. et al. Carcinoma produced factors activate myeloid cells via TLR2 to stimulate metastasis. Nature 457, 102–106 (2009).

18. Dvorak, H. F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. N. Engl. J. Med. 315, 1650–1659 (1986).

19. Dvorak, H. F. Tumors: wounds that do not heal-revis. Cancer Immunol. Res. 3, 1–11 (2015).

20. Bucheci, G., Torchio, P. & Ferrigno, D. Plasma levels of D-dimer in lung carcinoma: clinical and prognostic significance. Cancer 97, 3044–3052 (2003).

21. Komurcuoglu, B., Ulusoy, S., Gayaf, M., Guler, A. & Ozden, E. Prognostic value of plasma D-dimer levels in lung carcinoma. Tumori 97, 743–748 (2011).

22. Embembo, J. S. et al. Factor XIIIa transglutaminase supports hematogenous tumor cell metastasis through a mechanism dependent on natural killer cell function. J. Thromb. Haemost. 6, 812–819 (2008).
53. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*. 12, 323 (2011).

54. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA*. 100, 9440–9445 (2003).

55. de Hoon, M. J., Imoto, S., Nolan, J. & Miyano, S. Open source clustering software. *Bioinformatics*. 20, 1453–1454 (2004).

56. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA*. 95, 14863–14868 (1998).

57. Saldanha, A. J. Java Treeview—extensible visualization of microarray data. *Bioinformatics*. 20, 3246–3248 (2004).

58. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate—a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B-Methodol*. 57, 289–300 (1995).

59. Schmider, E., Ziegler, M., Danay, E., Beyer, L. & Bühner, M. Is it really robust? re-investigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology*. 6, 147–151 (2010).

60. Dean, A., Voss, D. & Dragulj, D. *Design and Analysis of Experiments*. (Springer, Berlin, 2017).

61. Hosack, D. A., Dennis, G. Jr, Sherman, B. T., Lane, H. C. & Lempicki, R. A. Identifying biological themes within lists of genes with EASE. *Genome Biol*. 4, R70 (2003).

62. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc*. 4, 44–57 (2009).

63. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 13, 2498–2504 (2003).

64. Mantel, N. Evaluation of survival data and two new rank order statistics arising in its consideration. *Cancer Chemother. Rep.* 50, 163–170 (1966).

65. Cox, D. R. & Oakes, D. *Analysis of Survival Data*. viii, 201 (Chapman and Hall, London, 1984).

66. Bland, M. *An Introduction to Medical Statistics*, xvi, 405 (Oxford University Press, Oxford, 2000).

67. Isljesa, M. D. et al. Genomic analysis of immune cell infiltrates across 11 tumor types. *J. Natl. Cancer Inst*. 108, 116 (2016).

68. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods*. 12, 453–457 (2015).

69. Li, T. et al. TIMER: a web server for comprehensive analysis of tumor-infiltrating immune cells. *Cancer Res*. 77, e108–e110 (2017).

70. Barbie, D. A. et al. Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112 (2009).

71. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*. 102, 15545–15550 (2005).

72. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 19, 185–193 (2003).

73. Lamprecht, M. R., Sabatini, D. M. & Carpenter, A. E. *CellProfiler*: free, versatile software for automated biological image analysis. *Biotechniques*. 42, 71–75 (2007).

74. Allott, E. H. et al. Performance of three-biomarker immunohistochemistry for intrinsic breast cancer subtyping in the AMBER consortium. *Cancer Epidemiol. Biomark.* 25, 470–478 (2016).

Acknowledgements

The authors would like to especially thank Drs. Stephanie Cohen, Nana Feinberg and Merri Efera from the UNC Translational Pathology Lab for their help with immunohistochemistry and processing of the tissue microarrays, the UNC Animal Histopathology Core, and Janet Dow from the UNC Flow Cytometry Core Facility. The authors thank Dr. Antonio Amelio for providing the fluorescent-Nanoluciferase plasmids. The UNC Translational Pathology Laboratory is supported in part by grants from the NCI (2-P50-CA016086–40), NIEHS (2-P50ES012016-15A1), UCRF, and NC日下午 (2015-IDG-1007). C.V.P. was supported in part by the National Institutes of Health RO1CA1215075, a Mentored Research Scholar Grants in Applied and Clinical Research (MRSG-14-222-01-RMC) from the American Cancer Society, the Jimmy V Foundation Scholar award, the UCRF Innovator Award, the Stuart Scott V Foundation/ Lung Cancer Initiative Award for Clinical Research, the University Cancer Research Fund, the Lung Cancer Research Foundation, the Free to Breathe Metastasis Research Award and the Susan G. Komen Career Catalyst Award. S.H.A. was supported in part by a grant from the National Institutes of Health under award number T32CA19658. S.H.R. was supported in part by NIH P30DK065988. L.M. was supported in part by the University Cancer Research Fund and NCI CA-180134. A.B. was supported by NIH R35CA197684. A.S.W. was supported in part by NIH R01HL126974. S.K. was supported in part by NIH R31HL139100. The UNC Flow Cytometry Core Facility and Lineberger Comprehensive Cancer Center Animal Histopathology and Animal Studies Cores are all supported in part by an NCI Center Core Support Grant (CA016086) to the UNC Lineberger Comprehensive Cancer Center. The UNC Flow Cytometry Core Facility is also supported in part by the North Carolina Biotech Center Institutional Support Grant 2012-IDG-1006. The views expressed in this article are those of the authors and do not reflect the official policy of the Department of Defense or U.S. Government.

Author contributions

Conception and design: C.V.P.; Development of methodology: A. Porrello, E.H.H., P.L., A.H., A.B., D.N.H., C.O., S.K., A.S.W., C.V.P.; Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): C.V.P., S.K., A.S.W., A.P., J.H., S.H.A., E.H.H., P.L., S.B., S.K.G., T.A.W., S.H.R.; Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): A.P., E.H.H., P.L., S.H.A., S.K.G., A.H., E.A., M.D.W., T.A.W., S.K., A.S.W., C.V.P.; Writing, review, and/or revision of the manuscript: All authors, Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): C.V.P., A.P., M.W.; Study supervision: C.V.P.

Additional information

Supplementary Information accompanies this paper at https://doi.org/10.1038/s41467-018-04355-w.

Competing interests: C. Oderup is an employee of Pfizer, Inc. The other authors disclosed no potential conflicts of interest.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2018