Integrative Network Analysis of Early-Stage Lung Adenocarcinoma Identifies Aurora Kinase Inhibition as Interceptor of Invasion and Progression

Seungyeul Yoo
Sema4 https://orcid.org/0000-0001-5616-2293
Abhilasha Sinha
Icahn School of Medicine at Mount Sinai
Dawei Yang
Icahn School of Medicine at Mount Sinai
Nasser Altorki
Weill-medical College of Cornell University
Radhika Tandon
St. George's University School of Medicine
Wenhui Wang
Icahn School of Medicine at Mount Sinai
Deebly Chavez
Icahn School of Medicine at Mount Sinai
Eunjee Lee
Icahn School of Medicine at Mount Sinai
Ayushi Patel
Icahn School of Medicine at Mount Sinai
Takashi Sato
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0002-2338-1665
Ranran Kong
Icahn School of Medicine at Mount Sinai
Bisen Ding
Icahn School of Medicine at Mount Sinai
Eric Schadt
Sema4
Hideo Watanabe
Icahn School of Medicine at Mount Sinai https://orcid.org/0000-0002-6017-6361
Pierre Massion
Vanderbilt Ingram Cancer Center
Alain Borczuk

Weill Cornell Medicine
Jun Zhu
Sema4
Charles Powell (charles.powell@mssm.edu)
Icahn School of Medicine at Mount Sinai

Article

Keywords: early-stage lung adenocarcinoma, network analysis, aurora kinase inhibition

Posted Date: April 13th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-334210/v1

Version of Record: A version of this preprint was published at Nature Communications on March 24th, 2022. See the published version at https://doi.org/10.1038/s41467-022-29230-7.

Integrative Network Analysis of Early-Stage Lung Adenocarcinoma Identifies Aurora Kinase Inhibition as Interceptor of Invasion and Progression

Seungyeul Yoo[1,2,3,#], Abhilasha Sinha[4,5,#], Dawei Yang[4,5,6], Nasser K. Altorki[7], Radhika Tandon[8], Wenhui Wang[1,2], Deebly Chavez[4], Eunjee Lee[1,2,3], Ayushi S. Patel[4,5], Takashi Sato[4,5,9], Ranran Kong[4,5], Bisen Ding[4], Eric E. Schadt[1,2,3,5], Hideo Watanabe[1,2,4,5], Pierre P. Massion[10], Alain Borczuk[11], Jun Zhu[1,2,3,5,*], and Charles A. Powell[4,5,*]

 # co-first authors

***corresponding authors**

Contacts: jun.zhu@mssm.edu, charles.powell@mssm.edu

1. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA
2. Icahn Institute for Data Science and Genomic Technology, New York, NY, USA
3. Sema4, a Mount Sinai venture, Stamford, CT, USA
4. Division of Pulmonary, Critical Care and Sleep Medicine, Icahn School of Medicine at Mount Sinai, New York, NY, USA
5. Tisch Cancer Institute, Icahn School of Medicine at Mount Sinai, New York, NY, USA
6. Department of Pulmonary and Critical Care Medicine, Zhongshan Hospital, Fudan University, Shanghai, China.
7. Department of Cardiothoracic Surgery, Weill Cornell Medicine-New York Presbyterian Hospital, New York, NY, USA
8. School of Medicine, St. George's University, West Indies, Grenada
9. Division of Pulmonary Medicine, Department of Medicine, Keio University School of Medicine, Tokyo, Japan
10. Department of Medicine, Vanderbilt University Medical Center, Nashville, TN, USA
11. Anatomic Pathology, Weill Cornell Medicine, New York, NY, USA

Abstract

We present a gene signature distinguishing invasive and indolent tumors among early-stage lung adenocarcinoma (esLUAD). An Invasiveness Score estimated using the gene signature was strongly associated with survival of esLUAD patients in multiple independent cohorts and with the invasiveness phenotype in lung cancer cell lines. Regulatory network analysis identified aurora kinase as one of master regulators of the gene signature and perturbation of aurora kinases in *vitro* and in *vivo* reduced tumor invasion. Our study suggests aurora kinases as a novel target for treating early-stage invasive lung adenocarcinoma.

Introduction

Lung cancer has the highest cancer incidence and is the leading cause of cancer-related mortality worldwide with annually 2.1 million new lung cancer cases and 1.8 million deaths[1]. Over the past 30 years, the overall five-year survival for lung cancer in the United States has increased from 12% to over 19% and continues to rise[2]. The increase is attributable to implementation of early detection screening programs and development of targeted therapies and immunotherapies that are effective in specific subtypes of metastatic lung cancer. These advances have been essential for the improved outcomes for lung cancer. In this study, we present an approach that focused on therapeutic approaches that target molecular subtypes of early-stage lung adenocarcinoma and promise to improve outcomes for this important patient group.

Here we focus on the molecular characterization of clinically significant histological subtypes of early-stage lung adenocarcinoma (esLUAD), which is the most common histological subtype of lung cancer. Although esLUAD patients have much better prognosis than patients with advanced disease, among early-stage patients treated primarily with surgery, about 30-50% of early-stage patients will develop metastasis with 70% and 35% overall 5-year survival rate for stage 1 and 2 non-small cell lung cancer patients, respectively[3-5]. Within lung adenocarcinoma, histology is heterogeneous and associated with diverse tumor invasion and clinical outcomes[6]. Invasiveness is one of cancer hallmarks and is directly related with metastatic potential and clinical outcomes of the tumor[7].

Yu and colleagues examined integrated gene/protein expression data of resected lung adenocarcinomas from The Cancer Genome Atlas LUAD cohort (TCGA) and showed that the integrated genomic models improved survival prediction of stage 1 tumors[8]. However, the analysis cannot be extrapolated to examine the impact of the histological invasiveness because the TCGA LUAD cohort does not include non-invasive histological subtypes such as minimally invasive adenocarcinoma (MIA) or adenocarcinoma in situ (AIS). Recently, Xing and colleagues

profiled indolent LUAD subsolid nodules and tumor micro-environment and showed a unique transcriptomic signature involving cell-cell interaction pathways that was distinct from primary LUAD with lymph node metastasis[9]. Our previous studies reported the type 2 TGF Beta Receptor (*TGFβRII)* as a significant determinant of invasiveness and metastasis of localized lung adenocarcinoma[4,10,11]. While these previous studies were useful to confirm the molecular and clinical heterogeneity in the esLUAD, the identification of treatment opportunities guided by molecular subtypes of early lung cancer remains underdeveloped[12].

In this study, we characterize invasiveness mechanisms in esLUAD by analyzing gene expression of a novel cohort of 53 histologically heterogeneous esLUAD samples. Based on transcriptomic analysis of this cohort, we identified a gene signature that distinguishes invasive and indolent tumors and validated the prognostic significance in multiple independent cohorts. We then performed a systematic analysis to understand molecular functions of the signature genes. Integrative network analysis highlighted *TPX2*, (an activator/binding partner of *AURKA*[13,14]), and *AURKB* as key regulators of the pro-invasive signature. Aurora kinases are a family of highly conserved serine-threonine kinases and are master regulators of mitotic cell division, chromosomal segregation and spindle formation[15,16]. Among the three members of aurora kinase family (A, -B and -C), aurora kinase A and B are the best characterized. Their role in tumor growth and survival has been reported across several cancer types[17,18], but the association of aurora kinases with lung tumor invasion or its role in early lung cancer therapy has not been studied before. In this article, we define the role of aurora kinases in early lung cancer progression and demonstrate using in *vitro* and in *vivo* models that targeting this pathway reduces lung cancer progression and improves survival.

Results

Lung Adenocarcinoma Invasiveness Signature

We performed molecular profiling of 53 histologically heterogeneous esLUAD samples (Supplementary Table 1) by RNA sequencing and identified signature genes associated with invasiveness of tumors at early-stage (Methods). First, we performed an unsupervised hierarchical clustering for the 53 samples based on the most varying genes (Extended Data Fig. 1a). The unsupervised hierarchical clustering resulted in two distinct groups containing 20 and 33 samples, respectively (Extended Data Fig. 1a). Differentially expressed genes (DEGs) between the two groups were determined based on t-test using cutoffs (fold-change (FC)>1.5 and FDR<0.01) and the samples were re-clustered into two groups of 21 and 32 tumors, respectively, based on the DEGs (Fig. 1a). Samples in each group were consistent with their histological subtypes; Group 1 on the left side (red color in Inv. Class) was enriched for pathological aggressive subtypes acinar (AC), micro papillary (MP), papillary (PAP) and solid (SOL) while Group 2 on the right side (blue color in non-Inv. Class) consisted of less invasive or indolent tumors such as MIA, AIS, and lepidic predominant (LPA) (Fig. 1a). Therefore, we annotated Group 1 as "Invasive" and Group 2 as "Indolent" (Supplementary Table 1). Together, 1,322 DEGs that distinguished the Invasive and Indolent tumors were identified as invasiveness signature genes (Supplementary Table 2). These invasiveness signature genes were separated into 526 genes up-regulated in the Invasive group (pro-invasive signature genes) and 796 genes up-regulated in the Indolent group (indolence signature genes). The signature was not associated with sex and smoking status (detailed in Supplementary Information).

The clustering of 53 tumors aligned with histological subtypes and staging of the samples (Chi-square test $p= 1.4\times10^{-7}$ and 0.001 for histology and nodal stage, respectively, Fig. 1b). The Invasive group was highly enriched for tumors with aggressive histology (AC, MP, PAP, and SOL, 17 out of 20, Fisher's Exact Test (FET) OddRatio (OR)=36.42 and $p=1.5\times10^{-7}$) while the Non-Invasive group included less invasive histological tumors (MIA, AIS, and LPA, 29 out of 32, FET OR= 50.17 and $p= 2.1\times10^{-8}$). Most node positive tumors were included in the Invasive group (9 Out of 10, FET OR=21.8 p=0.0005).

To compare our signature genes derived from clustering with those derived from tumor classification, we identified 793 histology-based DEGs between aggressive tumors (AC, MP, PAP, and SOL) and others (MIA, AIS, and LPA) (Extended Data Fig. 1b) and 80 nodal-based (N.stage) DEGs between node-positive and -negative tumors (Extended Data Fig. 1c). Our signature genes captured larger molecular differences between invasive and indolent tumors than histology- and node stage-based signatures containing 92% and 89% of histology-based and N.stage-based signatures, respectively (Fig. 1c). These results suggest that the signature defined by gene expression driven clustering is more informative for comparing invasiveness phenotypes among esLUAD tumors than signatures determined by histology or node metastasis status alone.

The signature genes are enriched for tumor invasion related functions.

We functionally annotated the pro-invasive and indolence signatures using public databases (Methods). First, gene ontology (GO) terms curated in the MSigDB C5 collection[19] were compared with our signature genes. The pro-invasive signature genes were significantly enriched for the cell cycle (FET OR=5.1, $p=2.9\times10^{-42}$), cell division (FET OR=6.9, $p=5.2\times10^{-31}$), extracellular matrix (FET OR=5.9, 4.6×10^{-22}), kinetochore (FET OR=11.9, $p=1.3\times10^{-19}$), and other cell cycle related functions. The indolence signature genes were enriched for terminally differentiated cell activities such as extracellular matrix (FET OR=3.0, $p=2.4\times10^{-9}$), tissue development (FET OR=1.8, $p=1.8\times10^{-7}$), and plasma membrane region (FET OR= 2.1, $p=3.7\times10^{-7}$). The complete list of GO terms enriched in the pro-invasive and indolence signatures is shown in Supplementary Table 3. Compared with the Hallmark gene sets[20], the pro-invasive signature genes significant overlapped not only with cell cycle associated gene sets such as mitotic spindle, G2M checkpoint, or E2F target genes, but also with EMT and angiogenesis pathways which are major hallmarks of tumor invasion and metastasis functions[7] (Fig. 1d).

Next, we compared our signatures with previously reported gene sets related to lung cancer and to tumor invasive properties (MSigDB C2 collection) (Fig. 1d). The pro-invasive and

indolence signatures significantly overlapped with poor and good survival lung cancer genes reported by Shedden et al.[21] (FET OR=20.7 and 7.6, $p=8.7\times10^{-110}$ and 8.4×10^{-23}, respectively), and they also significantly overlapped with up-regulated and down-regulated in invasive breast cancer[22] (FET OR=14.4 and 3.9, $p=3.2\times10^{-51}$ and 3.1×10^{-6}, respectively). The pro-invasive signature genes were also consistently enriched for other signatures associated with EMT[23-25], metastasis[26-28], and angiogenesis[25,29,30] (Supplementary Table 4). Interestingly, the indolence signature significantly overlapped with known tumor suppressor genes curated in TSGene2.0[31] (FET OR=2.1, $p=3.5\times10^{-8}$). These results suggest that our signature genes derived from lung adenocarcinoma are strongly associated with biological pathways important for tumor progression and are generalizable to invasive and/or metastatic features of epithelial neoplasms (additional annotations of the signature detailed in Supplementary Information).

The signature genes are associated with invasive phenotype of $Kras^{+/-}Tgfbr2^{-/-}$ mouse model.

Previously, we showed the loss of *Tgfbr2* in non-invasive murine inducible lung adenocarcinoma model ($Kras^{+/-}Tgfbr2^{-/-}$) induced a highly invasive phenotype associated with lymph node metastasis and poor survival[32]. *TGFBR2* was included in the human lung adenocarcinoma indolence signature (FDR=0.0016, Supplementary Table 2). We examined the overlay of human lung adenocarcinoma gene signatures with murine gene expression from invasive ($Tgfbr2^{-/-}$) and non-invasive ($Tgfbr2^{wt}$) and show that the pro-invasive and indolence signatures classified tumors from 11 mice into two groups (Methods) that exactly matched their *Tgfbr2* status (Fig. 1e). The $Tgfbr2^{-/-}$ mice (KO) had higher expression of pro-invasive signatures while $Tgfbr2^{wt}$ (WT) showed opposite patterns (Fig. 1e), confirming the significant role of *Tgfbr2* in transforming non-invasive tumors into invasive tumors. Comparing our human esLUAD signatures with murine KO vs WT DEGs (Methods), the pro-invasive signatures strongly overlapped with genes up-regulated in KO (FET OR=5.0, $p=1.6\times10^{-12}$) while the indolence genes were enriched for up-regulated in WT (FET OR=2.1, p=0.002) (Supplementary Table 5).

The signature genes are associated with early-stage patients' survival.

Because signature genes of invasiveness correspond to a critically important biological function of tumor progression that is associated with clinical outcomes (Fig. 1d-e), we hypothesized that our gene signatures could predict survival of esLUAD patients examined by independent studies. Using the pro-invasive and indolence signature genes, we developed an Invasiveness Score (IVS) based on elastic net to estimate tumor's invasiveness (Methods). Gene expression profiles of seven independent lung adenocarcinoma cohorts were downloaded, and stage I and II tumors were selected for further analysis (Methods). Based on the IVS, samples were classified as "Indolent" (peaked at 0) or "Invasive" (peaked at 1) and samples with IVS value between the two peaks were classified as "Intermediate" (Extended Data Fig. 2a, Methods). For each independent lung cancer cohort, we calculated IVS of each sample and associated the scores with patient survival. IVS captured clear transition of expression of gene signatures of tumor samples in each of 7 independent datasets (Extended Data Fig. 2b).

In all 7 independent datasets, "Invasive" groups had significantly inferior survival compared to "Indolent" groups (log-rank test (LRT) $p=8.3\times10^{-5}$, 0.01, 0.0002, 0.0007, 0.0003, 0.02, and 0.05 for Shedden, TCGA, Okayama, Tang, Der, Rousseaux, and Wilkerson datasets, respectively, Fig. 2a-g). With adjustment for age, gender, and tumor stages, the IVS remained significantly associated with survival in these cohorts (Extended Data Fig. 2c).

Considering the heterogeneity of populations from the international datasets, the strong and consistent detection of survival differences highlights the biological significance of the invasive signature genes as important prognostic biomarkers for early lung adenocarcinoma patient survival. Notably, most of the international datasets such as TCGA do not contain MIA or AIS tumors, yet our invasiveness signature genes distinguished patients with different survival, suggesting that our signature genes are generalizable for understanding biological and clinical pathways that determine survival outcomes in early-stage lung adenocarcinoma.

The signature genes rank invasiveness phenotype of cancer cell lines.

To evaluate invasiveness signature mechanisms *in vitro*, we calculated IVS as relative invasiveness in 70 lung adenocarcinoma cell lines from the CCLE dataset[33]. The 70 cell lines were ranked based on the IVS (Fig. 2h). To validate the biological implications of the IVS-based ranking, we experimentally assessed cell invasiveness using *in vitro* migration and invasion assays. We selected 5 cell lines from each of highly invasive (rank above median, NCI-H1373, SK-LU-1, NCI-H1792, A549, and NCI-H2009) and less invasive (rank below median, HCC78, NCI-H3255, Calu3, HCC1833, HCC2279) groups. The 5 highly invasive cell lines showed high transwell migration and invasion within 48hr time period compared to the less invasive cells (t-test p-values=6.5×10^{-21} and 8.7×10^{-24}, respectively, Fig. 2i-j for migration and 2k-l for invasion). Furthermore, the less invasive lines HCC1833, HCC78 and NCI-H3255 did not show any significant migration or invasion until 144hr of assay (Fig. 2j, 2l and Extended Data Fig. 3). Calu3 and HCC2279 had low migration and invasion at 48hr as compared to that of more invasive cells (t-test p-values= 9.7×10^{-18} and 3.7×10^{-22} in migration and invasion, respectively). Although all LUAD cell lines were derived from invasive LUAD tumors, the result above shows that the invasiveness phenotype varies across the LUAD cell lines and that our gene signatures enable us to infer the relative invasive phenotypes.

Integrative Network and Connectivity Map Analysis Identifies Key Drivers and Therapeutic Targets in Early Lung Adenocarcinoma.

The lung adenocarcinoma invasiveness signature distinguished tumors of different histology and more importantly, of distinct clinical outcomes following surgical resection. The biological mechanisms that drive the transformation of tumors from indolence to invasiveness remain incompletely characterized. Hence, we constructed a molecular regulatory network for esLUAD

by integrating gene expression, CNV, and methylation profiles of the stage I LUAD samples in TCGA using the software package RIMBANET (Methods)[34,35]. The final network consisted of 8,533 genes including 389 pro-invasive signature genes and 562 indolence signature genes (Extended Data Fig. 4). Applying key driver analysis (Methods), we identified 13 and 9 key drivers regulating the pro-invasive and indolence signature genes, respectively (Supplementary Table 6). The pro-invasive signature genes formed two large closed subnetworks consisting of 123 and 97 nodes, respectively. The first subnetwork was regulated by *TPX2, AURKB,* and 5 other key regulators (referred as TPX2/AURKB subnetwork in Fig. 3a) and the second subnetwork was driven by collagen associated genes *COL1A2* and *COL11A1,* and 3 other key regulators (referred as COL1A2 subnetwork in Fig. 3b). *TPX2* is a well-known activator of *AURKA* and vice versa and thus together regulate spindle assembly and dynamics[36,37]. *AURKB* is a master regulator of mitosis that forms the enzymatic core of the Chromosomal Passenger Complex (CPC) with Survivin (*BIRC5*)[38,39]. Both *AURKA* and *BIRC5* were also in the TPX2/AURKB subnetwork (Fig. 3a).

We interrogated molecular functions enriched within these subnetworks. Although there were distinct molecular pathways associated with each subnetwork (Supplementary Table 7), i.e. the TPX2/AURKB network was enriched for cell cycle related genes including G2M_CHECKPOINT and E2F_TARGETS (FET OR=66.93 and 56.4 with $p=1.45\times10^{-59}$ and 8.71×10^{-57}, respectively) and the COL1A2 network was enriched for HYPOXIA and ANGIOGENESIS (FET OR=8.07 and 25.16 with $p=1.83\times10^{-7}$ and 5.68×10^{-7}, respectively) in HALLMARK datasets, both subnetworks were enriched in the genesets such as HALLMARK MTORC1 signaling pathways (FET OR=4.9 and 4.93 with $p=9.37\times10^{-5}$ and 0.0004 for the TPX2/AURKB and COL1A2 subnetworks, respectively) and invasive signatures in breast cancer[22] (FET OR=18.68 and 7.42 with $p=5.7\times10^{-29}$ and 1.93×10^{-8}, respectively). In addition, genes in the TPX2/AURKB subnetwork were enriched for up-regulated genes in EMT in breast cancer[25] (FET OR= 53.12 and $p=1.02\times10^{-51}$) and genes in the COL1A2 subnetwork were

associated with HALLMARK EMT pathway genes (FET OR=33.16 and $p=1.85\times10^{-34}$) (Supplementary Table 7). These results suggest that even though the two subnetworks of the pro-invasive signatures were not directly connected with each other, both subnetworks share common downstream pathways associated with tumor invasiveness functions.

To identify drugs that could impact tumor invasiveness and potentially prevent tumor metastasis, we searched the Connectivity map (CMAP) database to identify small molecules that could perturb the two pro-invasive subnetworks[40] (Methods). A list of potential perturbagens for each subnetwork was identified based on CMAP enrichment scores from 2429 small molecule compounds in A549 cells. Aurora kinase inhibitor was among the top 10 perturbagens for both subnetworks (enrichment score < -95, Fig. 3c). When considering the two TPX2/AURKB and COL1A2 subnetworks together, the strongest perturbagen class to reverse the expression of the genes in the two subnetworks was Aurora kinase inhibitor (Fig. 3c). Taken together, the genomic analyses suggest that aurora kinases represent potential regulators of invasiveness that are vulnerable to small molecular inhibitors.

Aurora kinases A and B protein expression in human lung adenocarcinoma tumors is associated with pathological invasiveness and survival

To examine the clinical significance of AURKA and AURKB in esLUAD, we examined protein expression in a large series of resected human tumors represented on a tissue microarray (TMA) (Methods, Supplementary Table 8). Categorizing tumors into indolent and invasive groups based on their histology (MIA, AIS, LPA vs. AC, PAP, MP, SOL), patients with invasive tumors showed significantly worse survival than those with indolent tumors (LRT p=0.001, Extended Data Fig. 5a). AURKA and AURKB expression were significantly correlated with each other at the protein ($p=1.1\times10^{-5}$) level, similar to the correlation at the mRNA level ($p=6.7\times10^{-14}$ and 1.3×10^{-14} noted in the original and TCGA datasets, respectively) (Extended Data Fig. 5b). The AURKA and AURKB immunostaining scores were significantly higher in the invasive group (t-

test $p=2.0\times10^{-12}$ and 2.9×10^{-12}, respectively, Fig. 3d, IHC in Extended Data Fig. 5c-d), consistent with the observations based on mRNA expression data (Extended Data Fig. 5e-f). Tumor clusters based on the AURKA score showed different survival outcomes indicating prognostic significance of AURKA (LRT p-value=0.0004, Fig. 3e). Multivariate Cox-regression analysis showed the survival association of AURKA was independent from age, gender, and histology of patients (Hazard Ratio=1.9 with 95% confidence interval = [1.16, 3.1], p=0.01, Fig. 3f). Additionally, we observed higher levels of AURKA and AURKB protein expression in the more invasive CCLE cells lines as compared to the less invasive cell lines (Fig. 3g).

Aurora kinase A and B regulate migration and invasion phenotypes in lung adenocarcinoma cells

Aurora kinase A and B are aberrantly expressed in several tumor types[41-46], and promote cancer proliferation and survival owing to their functions in cell cycle and mitosis[47]. However, their role in regulating the invasiveness and migratory phenotypes in human tumors is not well characterized. We used genetic and small molecule inhibition approaches to examine the effect of *AURKA* and/or *AURKB* on tumor cell migration and invasion. We performed CRISPR/Cas9 deletion of *AURKA* and *AURKB* in highly invasive NCI-H1792 and A549 cells (Methods, Fig. 4a-b, Extended Data Fig. 6a and 7a). Repressing *AURKA* or *AURKB* alone did not alter migration and invasion in either cell line (Fig. 4c-d and Extended Data Fig. 6b-c) but, deleting *both AURKA* and *AURKB* together demonstrated significant reduction in migration by 60% and 50% and invasion by 70% and 60% in H1792 and A549 cells, respectively, adjusting for cell number (one-way ANOVA for all comparisons $p<0.0001$ Fig. 4c-d and Extended Data Fig. 7b-c). These results suggest that *AURKA* and *AURKB* have redundant functions in regulating lung tumor invasiveness and that pan-inhibition is required to impact tumor cell invasion.

Small molecule pan-Aurora kinase inhibitors suppress activity of Aurora kinases and invasive phenotype of lung adenocarcinoma cells

Because our genetic experiments indicate co-operative regulation of lung cancer cell invasion by *AURKA* and *AURKB*, we focused small molecule inhibitor studies on pan-Aurora kinase inhibitors, AMG900 and PF-03814735. AMG900 is in phase 1 clinical trials for acute myeloid leukemia[48] (NCT01380756) and advanced solid tumors[49] (NCT00858377) and in phase 2 clinical trials for triple negative breast cancer in combination with angiogenic kinase inhibitor ENMD-2076[50] (NCT01639248). PF-03814735 is in phase 1 clinical trial for advanced solid tumors[51] (NCT00424632). We treated highly invasive LUAD cell lines (NCI-H1792, A549, and NCI-H2009) with AMG900 or PF-03814735 for 48hr.

To determine the functional consequences of inhibiting AURKA and AURKB activity, we examined cell migration and invasiveness in five highly invasive cell lines (NCI-H1373, SK-LU-1, NCI-H1792, A549 and NCI-H2009, Fig. 2h-l) by treatment with AMG900 and PF-03814735. Both pan-aurora kinase inhibitors significantly suppressed transwell migration and invasion assays in all 5 cell lines treated with either drug (Fig. 4e, Extended Data Fig. 8a, and statistical test results in Supplementary Table 9). We also showed a significant decrease in cell motility in wound healing assay except for H1373 with PF-03814735 treatment (Fig. 4e, Extended Data Fig. 8a, and Supplementary Table 9). We did not observe any effect on cell viability at 48hr in any of the LUAD cells treated in a concentration range of 0.001μM to 10μM (Extended Data Fig. 8b, details in Supplementary Information).

Aurora kinases inhibition decreases invasive signature gene expression and impairs AKT/mTOR and EMT signaling pathway

Genetic and pharmacological inhibition of AURKA and AURKB impaired cell migration and invasiveness in lung adenocarcinoma cell lines (Fig. 4). To comprehensively understand the molecular mechanisms driven by inhibition of Aurora kinases, we examined gene expression

profiles of A549 and H1792 cells treated with an aurora kinase inhibitor (AMG900) and identified DEGs (FDR<0.01) (Fig. 5a, Extended Data Fig. 9a, and Supplementary Table 10). DEGs were highly consistent in both cell lines (FET OR=10.5 and 14.2 and p-value=10^{-1054} and 10^{-1572} for up-regulated and down-regulated genes, respectively) suggesting common effects of AMG900 treatment. They were enriched for pathways such as MTORC1_SIGNALING set (FET OR=4.3 and 5.8 p=2.5×10^{-20} and 2.6×10^{-27} for A549 and H1792 cells, respectively, Fig. 5b and Extended Data Fig. 9b, Supplementary Table 11). Genes down-regulated by AMG900 treatment in A549 cells were significantly enriched for the pro-invasive signature genes (FET OR=5.6 and p= 7.7×10^{-62}) and more significantly enriched for genes in the TPX2/AURKB subnetwork (FET OR=123.14 and p=1.9×10^{-104}, Fig. 5c). Similar observations were made for H1792 cells (FET OR= 2.0 and 15.0 and p= 7.2×10^{-10} and 5.1×10^{-16} for the prov-invasive signature and TPX2/AURKB genes, respectively, Extended Data Fig. 9c). These indicates that the AMG900 treatment transcriptionally perturbed the TPX2/AURKB subnetwork. Even though the COL1A2 subnetwork was not significantly down-regulated in A549 and H1792 treated cells (FET p=0.61 and 0.02), genes associated with the MTORC1 signaling pathway such as *PLOD2* and *LHDA* as well as an EMT gene, *SNAI2*, were suppressed by aurora kinase inhibition (Fig. 5c and Extended Data Fig. 9c).

It has been shown that activation of p-S6-kinase and p-4E-BP1 pathways by mTORC1 promotes cell motility and invasion[52], but there are no data showing a relationship between aurora kinases with AKT/mTOR signaling and tumor cell invasiveness. Consistent with the RNAseq transcriptomic analysis, we observed decreased p-AKT (Thr308) and p-mTOR (Ser2448) protein expression in A549 and NCI-H1792 cells treated with AMG900 (Fig. 5d, Extended Data Fig. 9d). Downstream targets of mTOR pathway activation, p-S6-kinase (Ser371) and p-4E-BP1 (Thr37/46) were similarly suppressed. Aurora kinase inhibitor also suppressed ERK1/2 activity in both cell lines (Fig. 5d, Extended Data Fig. 9d). The lung adenocarcinoma pro-invasive signature was enriched for EMT signature genes (Fig. 1d) that were suppressed in AMG900 treated H1792

cells (Extended Data Fig. 9b, FET OR=3.0, $p=3.5\times10^{-10}$). In cells treated with aurora kinase inhibitor AMG900, aurora kinase inhibition down-regulated mesenchymal markers N-Cadherin and Vimentin in A549 and NCI-H1792 (Fig. 5d, Extended Data Fig. 9d) and up-regulated E-Cadherin in A549. EMT markers such as Snail, Slug and Claudin1 were also suppressed by aurora kinase inhibition, suggesting that aurora kinases mediate invasiveness and cell motility by regulating EMT pathways. Taken together, these results show that Aurora Kinase inhibition impacts lung adenocarcinoma invasiveness through suppression of AKT/mTOR and EMT pathway signaling.

The genes up-regulated by AMG900 treatment did not overlap with the indolence signature genes in either cell line (FET p=0.32 and 0.98 for A549 and H1792 cells, respectively) suggesting that molecular mechanisms underlying the indolence signature genes regulation are independent of aurora kinase activity.

AMG900 intercepts lung adenocarcinoma invasiveness in Kras(G12D)/Tgfbr2$^{-/-}$ genetically engineered mouse model

To investigate the pre-clinical utility of aurora kinase inhibitors and to evaluate its efficacy in suppressing invasive lung adenocarcinoma, we treated our transgenic invasive LUAD mouse model of KrasG12D mutation and inducible Tgfbr2 deletion with AMG900. We previously showed that loss of Tgfbr2 in Kras G12D mutant condition induces highly invasive phenotype[32] in the transgenic mouse model. We imaged tumor formation starting from week 5 after Ad5-CMV-Cre administration and initiated AMG900 treatment 8 weeks after Ad5-CMV-Cre instillation (Fig. 6a). AMG900 was well tolerated in treatment group and there was no effect on animal weight (Extended Data Fig. 10g) and there were no other signs of toxicity such as loss of fur, skin ulceration or decreased motility.

Micro-CT images demonstrated clear differences in tumor density and burden between vehicle and AMG900 treated animals at several time points of tumor progression (Fig. 6b). The

tumor burden in vehicle group animals increased significantly faster compared to the tumor burden in AMG900 treated animals (t-value= -11.0 and p= 1.4×10^{-8} (Methods), Fig. 6c, Extended Data Fig. 10a).

Histopathological analysis (weeks 9, 13 and 17) identified several invasive and in situ tumors with bulky nodules. While solid and acinar patterns increased in control animal lungs, lungs from AMG900 treated animals showed suppression of invasiveness in mouse lungs (t-value= -5.1 and p= 4.0×10^{-5} (Methods), Fig. 6d, Extended Data Fig. 10b-c). The progression difference of lepidic tumor percentages between vehicle and treated animals was also significant (t-value= -3.0 and p=0.007, Fig. 6e, Extended Data Fig. 10b-c). The results together suggest the effect of aurora kinase on inhibiting tumor invasion. AMG900 treatment led to significant reduction in tumor infiltrating $CD31^{+}$ neovessels as compared to vehicle treated group (t-test p=0.03) (Fig. 6f and Extended Data Fig. 10d-e). Vehicle treated tumors showed heavy collagen deposition while AMG900 treated animals showed significantly lower level of collagen (t-test p=0.043 Fig. 6f, Extended Data Fig. 10f). AMG900 therapy led to a significant survival benefit over control group (LRT p= 0.015, Fig. 6g). While, vehicle group animals died within 18 weeks of treatment, AMG900 group survived till 35 weeks of treatment. Taken together, our *in vivo* transgenic data indicated that AMG900 treatment intercepts progression of invasive lung adenocarcinoma and leads to better survival in this animal model of early-stage invasive lung adenocarcinoma.

Discussion

In advanced lung carcinoma, genomic dissection of unique histological and clinical subtypes has identified actionable alterations that have led to the development and dissemination of targeted therapeutics that have remarkably altered the clinical course of this disease and have contributed to increased overall survival in lung cancer. Our premise is that similar approaches applied to

early-stage lung cancer might be similarly impactful. Our initial step was to introduce and confirm the clinical and biological significance of lung adenocarcinoma invasiveness as a key driver of tumor progression, metastasis, and clinical outcome in early-stage disease. In this study, we generated novel signatures of lung adenocarcinoma invasion and indolence and validated the biological and clinical impact of these signatures in independent cohorts, human tissues, human cell lines, and in genetic mouse models of invasive lung adenocarcinoma. Using genomic networks and in-silico approaches to identify actionable therapeutic targets, we identified and validated inhibitors of Aurora kinases A and B as effective interceptors of lung adenocarcinoma progression, invasion in cells and in the mouse model of invasive lung cancer.

Aurora kinases play important role in mitosis and are crucial in cellular processes such as chromosome segregation, chromosome alignment and spindle assembly[47,53-55]. Recent studies have shown AURKA and AURKB regulate cancer cell proliferation and tumor growth[41,43,56,57]. The role of aurora kinases in tumor migration and metastasis has been previously demonstrated in models of breast cancer[58], oral squamous call carcinoma[59] and osteosarcoma[60,61]. Our results in early lung cancer models show an important role for AURKA and AURKB in lung tumor progression through regulation of cell migration and invasiveness in lung adenocarcinoma. AURKA and AURKB function redundantly to influence lung cancer progression as shown by genetic deletion experiments showing that loss of AURKA and AURKB together, but not alone, leads to inhibition of migration and invasion of LUAD cells. We showed that the AKT/MTOR pathway mediates downstream signaling to influence tumor progression in AURKA/B inhibited cells and suggests opportunities for combination strategies to address cells that may bypass Aurora Kinase inhibition.

This is the first study to report potency of AMG900 in genetically engineered mouse model of invasive LUAD with potential application to intercepting invasion in early-stage human lung adenocarcinoma. We mined genomic data acquired from human lung adenocarcinoma specimens to identify signatures of tumor invasiveness and we used network modeling

approaches to identify key drivers susceptible to small molecule inhibition *in vitro* and *in vivo*. Our preclinical studies exhibit abrogation of progression and spread of invasive LUAD tumor with AMG900 treatment thus illustrating the importance of Aurora kinase inhibition as powerful therapeutic opportunity in clinic.

Clinical application of small molecule inhibitors (SMI) is expanding for targeted therapy for several cancers. In lung adenocarcinoma, inhibitors to oncogenic drivers such as EGFR, BRAF and ALK have been in use to treat advanced lung adenocarcinoma for several years in clinic[62]. Recently, proof of principle for application to early-stage lung adenocarcinoma was established in the ADAURA trial that demonstrated administration of adjuvant EGFR TKI Osimertinib after surgical resection prevented recurrence and death[63] (PMID 32955177). We anticipate the initial clinical application for testing of the invasiveness signature and AMG900 will be in the setting of unresected subsolid lung nodules that are commonly detected as solitary or multiple lesions in patients screened for lung cancer by chest CT examination. The objective will be to determine lesions most at risk for progression and to determine if AMG900 interception reduces tumor progression and metastasis.

In summary, we report a gene signature capable of classifying esLUAD patients into indolent and invasive groups, and inferred Aurora kinases as key regulators of the signature. We show that Aurora kinase A and B together regulate migration and invasiveness in LUAD through activating AKT/mTOR pathway and EMT pathway. Our findings support stratification of esLUAD patients based on our gene signature and suggest a potential role of pan-aurora kinase small molecule inhibitors in esLUAD patients with tumors of an invasive gene signature.

Methods

Patient samples and Histopathological analysis

Tumor specimens were acquired from 53 patients of resected lung adenocarcinoma with histology classification that included adenocarcinoma in situ and minimally invasive adenocarcinoma. Tumors tissues were collected from 2000-2010 and retrieved via an IRB approved protocol. All cases of AIS, MIA and LPA for which frozen tissue was collected were obtained. In addition, small invasive adenocarcinomas (prioritized as <2.0 cm, but all <3.0 cm) with node metastasis were identified (10 cases) and matched to 10 cases of the same T stage size grouping and with the same predominant invasive pattern but lacking node metastasis. Demographic details of these 53 patients are listed in Supplementary Table 1.

Tumor frozen sections were stained using 1% cresyl violet and were microscope guided needle dissected in an RNase free collection tip under vacuum suction to enrich for tumor >75% for further analysis. These samples were immediately processed for RNA extraction using the Qiagen RNeasy kit. RNA quality was assessed by Agilent Bioanalyzer, and tumors with RIN values <7.0 were excluded.

RNA-sequencing for LUAD patients and cell lines

Total RNA (approximately 100-300 ng per sample) for each resected tumor sample was poly(A)-selected, fragmented, converted to cDNA and barcoded for multiplex sequencing on the Illumina HiSeq 2500. The RNA sequencing reads were processed through a standard protocol for TopHat and Cufflinks[64]. For reference genome and transcriptome, hg19 reference genome and UCSC refseq gtf files were downloaded (https://genome.ucsc.edu/). The average number of reads per sample is 45 million and the mapped reads were further used to estimate abundance of genes in Fragments per kilobase of transcript per million fragments (FPKM) (details in Supplementary Information). For RNA sequencing of A549 and H1792 cells treated with DMSO or 0.1μM

AMG900 for 48hr, total RNA was purified from cells using Rneasy kit (Qiagen) for sequencing performed on NextSeq 500 (Illumina). Deseq2[65] was used to identify differentially expressed genes (DEGs) (FDR<0.01) between control (DMSO) and AMG900 treated cells. Gene expression omnibus accession numbers: GSE166722 (GSE166720 for 53 primary tumors and GSE166721 for AMG900 treated cells).

Identification of invasive signatures

Unsupervised hierarchical clustering used the top 1000 most varying genes (variance>1.0). Differentially expressed genes (DEGs) between the two clusters were determined by t-test (Fold-Change (FC) >1.5 & FDR<0.01). The samples were clustered again using the DEGs, and after two rounds of iteration, we reached the static state with no more change in DEGs nor cluster memberships. The final clustering result contained two groups of 21 and 32 samples, labeled as "invasive" and "non-invasive", respectively, based on the predominant pathology histology in each group. Then the final signature genes were determined by t-test with the same cutoffs above and separated into "pro-invasive" and "indolence" signatures according to their direction across tumors.

For functional annotation, Hallmark gene sets, curated gene sets, and GO terms in Molecular Signatures Database (MSigDB) were used[20]. The significance of enrichment was determined based on Fisher's Exact Test (FET) p-value considering multiple testing (FET p<0.05/(# of dataset tested)) in each category. For curated gene sets, "C2" collection in the MsigDB, we filtered gene sets using keywords such as "Lung cancer", "Epithelial mesenchymal transition (EMT)", "Invasion", "Migration", "Metastasis", and "Angiogenesis". Known tumor suppressor genes (TSGs) were collected from TSGene2.0[31].

Gene expression data from lung tumors of KrasTgfbr2 $^{-/-}$ mouse model

Gene expression microarray data of 11 mouse lung tumors was downloaded from GSE27717. The dataset consists of 5 invasive lung tumors from KrasTgfbr2$^{-/-}$ mice and 6 non-invasive lung tumors from KrasTgfbr2^{wt} mice[32]. The pro-invasive and indolence genes were used to cluster the murine tumors according to unsupervised hierarchical clustering. The DEGs between two groups of lung tumors were determined by t-test (FC>1.5 and FDR<0.01).

Validation of the Invasiveness Signature in Independent Human Lung Tumor Datasets

We collected additional LUAD cohorts to test whether our invasiveness signature genes could classify samples into distinct subgroups associated with dissimilar patient's survival. For this purpose, we downloaded gene expression data from 7 independent publicly available LUAD cohorts with survival information of the samples: a RNAseq data set from TCGA LUAD[66] and 6 microarray data sets including Shedden et al. (GSE68465)[21], Okayama et al. (GSE31210)[67], Der et al. (GSE50081)[68], Tang et al. (GSE42127)[69], Rousseaux et al. (GSE30219)[70], and Wilkerson et al. (GSE26939)[71]. We restricted our analysis to early-stage (I and II) tumor specimens. The number of samples used in this study is 371, 397, 204, 127, 111, 85, and 62 for Shedden et al., TCGA LUAD, Okayama et al., Der et al., Tang et al., Rousseaux et al., and Wilkerson et al., respectively. In addition to seven primary tumor datasets, we downloaded gene expression profiles of 70 LUAD cell lines from Cancer Cell Line Encyclopedia (CCLE) microarray data matrix[33].

Estimation of invasiveness score (IVS)

First, using the pro-invasive and indolence signatures, we trained a classifier based on elastic net to classify our original dataset whose invasiveness was determined (0=indolent and 1=invasive)[72]. To account for any potential differences between gene expression platforms, both the original and the testing gene expression matrix were z-transformed. The elastic net[73] is a regularized

regression model that uses a linear combination of the L_1 and L_2 penalties of the lasso and ridge methods as below;

$$\min_{\beta_0,\beta}\left[\sum_{i=1}^{N}(y_i-\beta_0-{x_i}^T\beta)^2+\lambda\sum_{j=1}^{|x_i|}\left(\frac{1-\alpha}{2}\beta_j^2+\alpha|\beta_j|\right)\right]$$

where, y_i corresponds to binary class (invasive or non-invasive) and x_i is a vector of features (z-transformed gene expression) for the i^{th} samples. The β's are regression coefficients that we estimate. The tuning parameter λ is the weight of the regularization terms and is chosen to minimize mean square errors. α is the elastic net penalty. Using a R package, "glmnet", we performed cross validation to select the optimal regularized parameters of the elastic net as the elastic net penalty (α=0.02) and the regularization parameter (λ=0.1). Then the elastic network classifier was applied to samples in each cohort to estimate relative IVS as predicted probability with distribution [0,1]. The association between IVS and patients' survival was tested using Cox regression with age, gender, and tumor stage as confounding factors $Survival \sim age + gender + stage + IVS$. For survival analysis in the TCGA LUAD dataset[66], we used overall survival (OS) data recently updated by Liu et al[74]. For other datasets, vital status and the last contact or death days were used as deposited into Gene Expression Omnibus (GEO) database.

Classification of samples based on IVS

Samples in each independent primary tumor cohort were classified into three groups (invasive, intermediate, and indolent) based on IVS. We divided IVS into 40 bins (bin size=0.025 between 0 and 1) and identified the smallest and largest local minima If IVS<smallest minima, the tumor was classified as "indolent", else if IVS>largest, the tumor was classified as "invasive". Otherwise, the tumor was classified as "intermediate" (Extended Data Fig 2a). The survival difference across these group was compared. We censored the maximum time of survival at 5 years and measured the 5-year survival rates across tumor groups in each dataset with likelihood-

ratio test (LRT) p-value from a cox proportional hazards model. The "coxph" function from an R package "survival" was used to perform the survival analysis.

Construction of a molecular causal network for esLUAD

A molecular causal network for esLUAD was constructed by integrating methylation, CNV, and gene expression profiles of stage I patients in the TCGA LUAD dataset. RSEM data for gene expression, Illumina HumanMethylation450 matrix for DNA methylation and CBS segment mean values for CNV were downloaded for stage I patients from the TCGA data portal (https://portal.gdc.cancer.gov). Sample alignment[75] was performed to confirm that different types of data were from the same individuals (details in Supplementary Information) and 216 stage 1 tumor samples with all three types of data were included in the further analyses. A total of 8,533 informative genes with detectable expression levels and large variances across samples were selected to be included in the network reconstruction process. Among them, the expression of 3,476 and 761 genes were cis-regulated by CNVs or promoter methylation (FDR < 0.01), respectively, and cis-CNVs and cis-methylations were included as root nodes in the network construction as described in Supplementary Information.

The gene expression, CNV, or methylation profile was discretized into three states: low, normal, and high level, guided by k-means clustering (k = 3) and biological meaningful cutoff values. The gene expression, cis-CNV, and cis-methylation nodes were then imported into the software suite, Reconstructing Integrative Molecular Bayesian Network (RIMBANet), to construct a biological causal network given the data and priors, as previously described[34,35]. Briefly, the network reconstruction process searches for a directed acyclic graph (DAG) structure G and associated parameters Θ that can best explain the given data D, $P(G, \Theta|D)$. If the structure G is a DAG, then $P(G, \Theta|D)$ can be decomposed into a series of sub-structures $P(G, \Theta|D) = \prod_i P(G^i, \Theta^i|D)$. With cis-CNV and cis-methylation nodes included, the structures X→Y, given

by $p(X \to Y|D) = p\left(Y\middle|X, CNV_y, Methyl_y D\right)p(X|D)$, and Y→X, given by $p(Y \to X|D) = p(X|Y, CNV_x, Methyl_x D)p(Y|D)$, are no longer equivalent, so that potential causal relationships between X and Y can be inferred unambiguously. To speed up the searching process, for each gene, the bottom 20% genes based on their mutual information were excluded as potential candidate regulators (sparse candidate search[76]). The network reconstruction process is a Monte Carlo Markov chain (MCMC) process. Given different random seeds, we might end up with different structures. Thus, we ran 1000 independent MCMC processes based on 1000 random seed numbers that resulted in 1000 candidate structures. Then, we selected consensus structure features with posterior probabilities >0.3 among candidate structures[77]. Finally, loops in the consensus network were removed by deleting the weakest link in the loops. The resulting network was visualized using Cytoscape3.7[78]. Given a set of seed nodes N_s, $SN_s = \bigcup_i d(node, N_s^i) \leq l)$ is the union of nodes that are within l steps from the seed node N_s^i, and the subnetwork for the seed nodes N_s is the set of connections among N_s.

Subnetworks enriched for invasive signatures

Key driver analysis (KDA)[79,80] was performed to identify master regulators associated with pro-invasive or indolence signatures in two steps. First, for each node in the network, we extracted a list of nodes within two layers from the seeding node and tested the significance of enrichment against the pro-invasive or indolence signatures (Fisher's exact test (FET) p-value $<10^{-8}$, adjusting multiple testing). Sorted on FET p-values, a gene with the strongest p-value was determined as the top key regulator. Any candidate regulators in its two-layered neighbors were excluded from the candidate list. Then we identified the next key regulator with the strongest p-value among remaining candidates. The process was iterated throughout the sorted candidate list.

To extract the pro-invasive and indolence subnetwork, we overlaid the pro-invasive and indolence signatures on the network and collected two large closed subnetworks; one with 123

genes, and the other with 97 genes. Drug treatment perturbation profiles curated in Connectivity Map (CMAP)[81,82] were used to investigate potential drug candidates that impact expression of the genes in the two subnetworks. The CMAP database (LINCS database) encompasses 1.3 million L1000 profiles covering 8 different cell lines responding to 7977 drugs[40]. Using the 123 and 97 genes from the two invasive subnetworks defined above, we focused on results from the lung adenocarcinoma cell line A549 (https://clue.io/). The output small molecule classes were compared based on the negative enrichment score (enrichment score < -95).

Human tissue microarray and Immunohistochemistry

Tissue microarrays of 768 consecutive lung adenocarcinoma cases from 1997-2000, 2002, 2010-2013 were constructed with 1.0 mm tumor plugs in triplicate with the exception of 2011, 2012 and 2013 which were in duplicate. The 2014 and 2015 arrays included Stage I adenocarcinoma only. We focused on 396 stage I tumors in our analysis. TMA patients have similar demographic of age and gender with our esLUAD cohort (Supplementary Table 8). Immunohistochemistry for AURKA and AURKB in human TMA was performed using the Leica Bond system. Sections were pre-treated with heat mediated antigen retrieval with Tris-EDTA buffer (pH=9, epitope retrieval solution 2) for 20 mins. They were incubated with primary antibodies for Aurora A or Aurora B for 30 mins at room temperature and detected using an HRP conjugated compact polymer system with DAB as the chromogen and hematoxylin as counterstain. Aurora A and B staining was measured by an H-score that integrates a semi-quantitative score of 0, 1 or 2+ for intensity and the percentage of positive cells. Nuclear staining for Aurora kinase A was required in all cases, but when present, both nuclear and cytoplasmic staining was included for H-score. For Aurora kinase B, only nuclear staining was scored.

We performed t-test to assess significance of differences of AURKA and AURKB protein expression between non-invasive (AIS, MIA, and LPA) and invasive (AC, MP, SOL, and PAP) tumors. The association between the AURKA expression and survival was tested on

patients with survival information (107 non-invasive and 158 invasive) using Cox regression with age, gender, and histology (invasive vs non-invasive) as confounding factors $Survival \sim age + gender + histology + AURKA$.

Cell culture

Lung adenocarcinoma cell lines NCI-H1373, NCI-H1792, NCI-H2009, HCC-78, NCI-H3255, Calu-3, HCC-1833 and HCC2279 were maintained in RPMI 1640 (Gibco) supplemented with 10% FBS, 1% penicillin and 1% streptomycin. SK-LU-1, A549 and HEK293T cells were maintained in DMEM (Gibco) supplemented with 10% FBS, 1% penicillin and 1% streptomycin. All cells were maintained in 37°C and 5% CO_2. All cell lines were regularly tested for mycoplasma using the mycoAlert Detection Kit (Lonza).

Lentivirus production

HEK293T cells were seeded in 10cm culture dish. After reaching a confluency of 70-80%, cells were co-transfected with 10μg target plasmid construct, 7.5 μg of psPAX2 (Addgene #12260) and 2.5 μg pMD2.G (Addgene #12259) vectors using TransIT®-Lenti Transfection Reagent (Mirus) according to the protocol. Lentivirus particles were harvested after 48hr of transfection and filtered with 0.45μM filter for transduction into target cells.

CRISPR-Cas9 genome editing

Lenti-Cas9-Blast plasmid (Addgene # 52962) or hUBCp_Cas9_3xNLS_p2a_puroR plasmid (Addgene #81251) were used to generate cells with stable human S. pyogenes Cas9 expression. Cells were infected with lentivirus and supplemented with polybrene (Sigma) at a final concentration of 8μg/ml. Transfected cells were selected with 8-10μg/ml of blastidicin or 3-5μg/ml puromycin for 6-10 days. For generating CRISPR knock out cell lines, short guide RNA (gRNA) target sequences for *AURKA* and *AURKB* and non-target sgRNAs were acquired from

human CRISPR knockout pooled Brunello library[83]. sgRNAs were cloned into pLKO.1 GFP or pLKO.1 mCherry vector at Bbs1 site downstream of human U6 promoter. pLKO.1-GFP plasmid was a gift from Dr. Brian D Brown laboratory (ISMMS, New York, NY) and pLKO.1 mCherry was purchased from Addgene (#128073) (Supplementary Table 12 for plasmid details). Sequences of sgRNAs are listed in Supplementary Table 13. Cas9 expressing cells were infected with pLKO.1-GFP-sgRNA or pLKO.1-mCherry-sgRNA.

Migration and Invasion

For migration assay, 50 X 10^3 cells were seeded on an 8μM cell culture insert (Fisher Scientific) in triplicate wells in serum free media in a 24-well plate and 800μL of 10% FBS supplemented media was added to the lower compartment of the well. Cells were incubated for 48hr in 37°C incubator and then the inserts were washed with phosphate-buffered saline. Cells on the top of transwell were scraped and washed away using cotton tip applicator. Cells on the bottom side of transwell were fixed with 70% ethanol and stained with 0.2% crystal violet. Images were analyzed using ImageJ software for each replicate.

For invasion assay, 8μM cell culture insert was coated with 300 μg/ml of Corning matrigel basement membrane matrix (Corning) and incubated for 1hr in 37°C incubator. 50 X 10^3 cells were then seeded over the layer of matrigel in the 8μM cell culture insert and was stained in a similar fashion as migration assay. To test whether phenotype changes were significant or not, we measured the significance of a fit line from a generalized linear regression mode (n=3); $percentage(invasion\ or\ migration) \sim dose$. All tests were significant ($p<0.05$) and results are reported in Supplementary Table 9.

Wound healing assay

Cells were cultured to 100% confluency in a 24 well plate and wound was carefully created using P20 pipette tip. Images of wound were captured at 0hr, 24hr, 48hr and 72hr to evaluate closure of

wound. Images were analyzed using ImageJ software. A linear model $percentage_{wound\_healing} \sim dose$ was used to test significance of phenotype changes in a dose dependent manner. Results are reported in Supplementary Table 9.

In vivo transgenic mouse model and drug treatment

All mouse studies were conducted in compliance with regulations and guidelines from the Mount Sinai Institutional Animal Care and Use Committee. Health of mouse was monitored every day. To develop *in vivo* invasive lung adenocarcinoma animal model, we generated transgenic mouse LSL-KrasTgfbr2$^{-/-}$ by crossing LSL-KrasG12D positive mice with Tgfbr2$^{flox/flox}$ mice[32]. Mouse genotypes were confirmed at 21 day from birth by PCR amplification of genomic DNA isolated from tail snips (primer sequence available on request). To induce tumor formation, 2.5 X 10^7 pfu particles of Ad5-CMV-Cre (University of Iowa) were administered intranasally to LSL-KrasTgfbr2$^{-/-}$ mice at the age of 6-8 weeks, as described previously[32]. To begin AMG900 treatment, KrasTgfbr2$^{-/-}$ mice were imaged by micro-CT every week starting from 5 weeks to identify earliest signs of tumor formation. At week 8 post Ade Cre administration animals were randomly divided into 2 groups and treated with either vehicle (0.5% hydroxy propyl methyl cellulose, 0.1% Tween 80, pH=2.2) or 3mg/kg AMG900[84] by oral gavage twice every day (n=10 per group). Mice were sacrificed at several time points during treatment for histopathological analysis for tumor lesion comparison. The tumor burdens of mice from two groups (vehicle and AMG900 treated) were compared using a generalized linear model $tumor\ percentage \sim time + time{:}\ group$, and significance of the interacting term (time:group) was measured with t-value and p-value.

Micro CT imaging and quantitation

KrasTgfbr2$^{-/-}$ mice were imaged regularly starting from week 5, to monitored tumor spread by micro-CT imaging. For micro-CT imaging, animals were anesthetized using a cocktail of

100mg/kg ketamine and 10mg/kg xylazine injected via intra-peritoneal injection at the dose of 0.1g per 10g of body weight. Mice were intubated with 20G X 1 inch catheter with the help of Fiber optic lightening kit (Kent Scientific Corporation, CT). Mice were ventilated on small animal ventilator MiniVent (Hugo Sachs Electronik, Germany) to control the breathing and imaged on nanoScan PET/CT (Mediso USA). Images were acquired with a 300ms scan at 150kVp and 610μA. 3 to 5 mice were imaged per group at several time points before and after starting AMG900 treatment. Horos v3.3.6 was used to mark the tumor region of interest and total lung and % tumor burden was calculated for each lung.

Histopathological analysis

Lung sections from LSL-KrasTgfbr2$^{-/-}$ animals treated with vehicle or AMG900 collected at several time points were formalin fixed paraffin embedded and stained with hematoxylin and eosin for analysis by expert pathologist. Each lung section was carefully studied by pathologist and marked for invasive and lepidic tumor areas. Lepidic and invasive tumor area per lung was calculated using Aperio ImageScope 12.1 software.

Competing interests

SY, EL, EES, and JZ are employees of Sema4, a for-profit organization that promotes personalized patient care through information-driven insights. Other authors declare that they have no competing interests.

Figure legends

Fig. 1. Pro-invasive and indolence signature genes associated with tumor invasion at early-stage lung adenocarcinoma.

- **a.** Invasiveness signature genes separating 53 early-stage lung adenocarcinoma into invasive and non-invasive tumors. "complete" method was used for unsupervised clustering.
- **b.** Comparison of the tumor clusters based on the signature genes with histological subtypes and nodal stages of the tumors. P-values of association were measured by Chi-Square test.
- **c.** Comparison of the gene numbers of our signatures (pro-invasive and indolence signatures) with differentially expressed genes based on histological subtypes (AIS, MIA, LPA vs AC, MP, PAP, SOL) and nodal stage (N0 vs N1+).
- **d.** Association of pro-invasive and indolence signatures with MSigDB genesets including Hallmark, lung cancer related genes, tumor invasion or metastasis related genesets.
- **e.** Unsupervised clustering of 11 tumors (5 Tgfbr2 -/- and 6 Tgfbr2 wt mice from GSE27717) based on the signature genes. "complete" method was used for unsupervised clustering.

Fig. 2. Invasiveness score associated with patient survival and tumor cells phenotype

a. Stratification of tumor samples of Shedden et al. cohort into three groups; invasive (high IVS), intermediate (middle IVS), and indolent (low IVS) tumors determined based on IVS (Methods, Extended Data Fig 2a). Five-year survival of tumors was shown in a KM curve with corresponding LRT p-values.

b. KM plot for TCGA LUAD cohort.

c. KM plot for Okayama et al. cohort.

d. KM plot for Tang et al. cohort.

e. KM plot for Der et al. cohort.

f. KM plot for Rousseaux et al. cohort.

g. KM plot for Wilkerson et al. cohort.

h. Relative invasiveness scores (values 0-1) of 70 LUAD cell lines estimated via IVS. Tumor cells above median score were labeled as "More Invasive" and below as "Less Invasive". In each group, we selected 5 for further experimental validations as marked in asterisk (red for more invasive and blue for less invasive cells).

i. Representative images of migrated cells in transwell migration assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells.

j. Quantification of migrated cells through transwell at 48hr, 96hr, and 144hr. Data presented as mean (n=3).

k. Representative images of invaded cells in transwell matrigel invasion assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells

l. Quantification of invaded cells through transwell matrigel at 48hr, 96hr, and 144hr. Data presented as mean (n=3).

Fig. 3. AURKA and AURKB are master regulators regulating pro-invasive signatures

a. A subnetwork of 123 pro-invasive genes driven by *TPX2* and *AURKB*. Seven master regulators are marked as squares with bigger size.

b. A subnetwork of 97 pro-invasive genes driven by *COL1A2*. Five master regulators are marked as squares with bigger size.

c. Top 10 small molecular classes significantly reversing expression of genes in the two subnetworks in A549 using CMAP database (enrichment score < -95, Methods): left for TPX2/AURKB subnetwork, middle for COL1A2 subnetwork, and right for both.

d. Comparison of AURKA and B protein scores between non-invasive (AIS, MIA, LPA) and invasive (AC, MP, PAP, SOL) tumors from human tissue microarray data. AURKB score were not available for all patients. T-test p-values are shown.

e. KM curve showing the survival differences between patients classified into 0, 1, 2+ based on AURKA score and the percentage of positive cells (Methods). The LRT p-value was estimated.

f. Hazard ratio and 95% confidence interval was measured by multivariate cox regression from $Survival \sim age + gender + histology + AURKA\_Group$. Significant p-values were marked in asterisk.

g. Western blot of AURKA and AURKB for panel of 5 less invasive and 5 more invasive LUAD cell lines from Fig. 2i.

Fig. 4. Aurora kinase A and B are both regulators of invasive phenotype.

a. Western blot for H1792 cells transduced with indicated sgRNAs.

b. Western blot for A549 cells transduced with indicated sgRNAs

c. Quantitation of percent migration of H1792 and A549 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA n=3, $p<0.0001$)

d. Quantitation of percent invasion of H1792 and A549 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA n=3, $p<0.0001$)

e. Heat map for H1373, SK-LU-1, H1792, A549 and H2009 showing effect of AMG900 and PF-03814735 on invasion and migration phenotype, and wound healing ability, at indicated drug concentrations. (n=3, statistical test results in Supplementary Table 9)

Fig 5. Aurora kinases drive invasiveness in lung adenocarcinoma through activating AKT/mTOR and EMT pathways.

a. Differentially expressed genes between DMSO and 0.1μM AMG900 treated A549 cells (FDR<0.01)

b. Top 10 down-regulated HALLMARK pathways in A549 treated with 0.1μM AMG900.

c. Overlaying the DEGs in A549 cells onto the TPX2/AURKB and COL1A2 subnetworks. Nodes filled in red are up-regulated and ones in green are down-regulated by AMG900 treatment. Genes included in HALLMARK_MTORC1_SIGLANING pathways are indicated with diamonds (◊) within a dashed circle.

d. Western blot for A549 cells treated with DMSO and indicated concentrations of AMG900 for 48hr for AKT/mTOR pathway and EMT pathway.

Fig 6. Aurora kinase inhibition suppresses invasive lung adenocarcinoma progression in Kras(G12D)/Tgfbr2$^{-/-}$ mouse model.

a. Scheme of treatment of Kras(G12D)/Tgfbr2$^{-/-}$ mouse with AMG900.

b. micro-CT images of vehicle and AMG900 treated mice at week 3, 6 and 9 after treatment. Treatment was given at 3mg/kg twice daily by oral gavage. (vehicle n=10, AMG900 n=8)

c. Quantitation of tumor burden from micro-CT imaging for vehicle and AMG900 treated mice at week 3, 6 and 9 after treatment. (n=3). A fitting line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice.

d. Comparison of invasive tumor percentage from vehicle and AMG900 treated mice between baseline and week 9, 13 and 17 from histopathological analysis. (n>=2). A fitting line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice.

e. Comparison of lepidic tumor percentage from vehicle and AMG900 treated mice between baseline and week 9, 13, and 17 from histopathological analysis. (n>=2). A fit

line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice.

f. (Top) IHC images for CD31 staining in Vehicle and AMG900 treated mouse tumor. (Bottom) Masson's trichome staining in Vehicle and AMG900 treated mouse lung.

g. Survival curve for vehicle and AMG900 treated mice. (vehicle n=10, AMG900 n=8). LRT p-value was measured (p=0.015).

Extended Data Figures

Extended Data Fig 1. Identification of invasive signature genes.

- **a.** Unsupervised clustering of 53 early-stage lung adenocarcinoma using top 1000 most varying genes (left) and another unsupervised clustering of the tumors using the DEGs between two groups separated by the most varying genes (right).
- **b.** Histology based DEGs from early-stage lung adenocarcinoma. Using the same cutoffs (FC>1.5 & FDR<0.01), there were 313 genes up-regulated in tumors with aggressive histology (AC, MP, PAP, and SOL) compared to AIS/MIA/LPA tumors and 480 genes down-regulated in aggressive tumors.
- **c.** Node stage based DEGs. Twenty-three genes were up-regulated and 57 genes were down-regulated in node positive tumors compared to no nodal metastasis tumors.

Extended Data Fig 2. Invasiveness Score (IVS) associated with patient's survival

a. Distribution of IVS in 7 independent lung cancer cohorts (stage I and II only). Tumors were further clustered into high, middle, and low IVS groups based on local minima of IVS based on the histogram (bin size=0.025).

b. Heatmap of expression of the signature genes of tumors in 7 cohorts. Samples were sorted from lowest IVS (left) to highest IVS (right).

c. Forest-plots showing significant association of IVS with patients' survival in multiple lung cancer cohorts. Hazard ratios with 95% confidence interval is shown with corresponding p-values. Significant p-values ($p<0.05$) are shown in red.

Extended Data Fig 3. Migration and invasion assay in less invasive LUAD cells.

- **a.** Representative images of migrated cells in transwell migration assay after 96hr and 144hr in 5 less invasive LUAD cells.

b. Representative images of invaded cells in transwell matrigel invasion assay after 96hr and 144hr in 5 less invasive LUAD cells.

Extended Data Fig 4. Integrative regulatory network for early-stage lung adenocarcinoma using TCGA LUAD stage I tumors.

The final network consisted of 8,533 genes including 389 pro-invasive signature genes (red) and 562 indolence signature genes (green).

Extended Data Fig 5. AURKA and AURKB expression between invasive and non-invasive tumors

a. Survival difference between invasive and non-invasive tumors in stage I TMA data. LRT p-value was measured.

b. AURKA and AURKB protein expression from stage I tumor in the TMA dataset (left), mRNA expression across 53 samples from our original dataset (Sinai) (middle), and mRNA expression across 216 TCGA Stage 1 RNAseq samples (right). Pearson correlation coefficient and p-values were estimated.

c. Representative images from immunohistochemical staining of AURKA in AIS, MIA, LPA vs AC, MP, PAP and SOL in human TMA.

d. Representative images from immunohistochemical staining of AURKB in AIS, MIA, LPA vs AC, MP, PAP and SOL in human TMA.

e. Comparison of mRNA levels of *AURKA* and *AURKB* between invasive and non-invasive tumors from the Sinai dataset. T-test p-value was measured.

f. Comparison of mRNA levels of *AURKA* and *AURKB* between invasive and non-invasive tumors from the TCGA Stage I samples. T-test p-value was measured.

Extended Data Fig 6. CRISPR deletion of Aurora kinase A or -B alone has no effect on migration and invasion in LUAD cells.

- **a.** Western blot for H1792 cells transduced with indicated sgRNAs.
- **b.** Quantitation of percent migration and invasion of H1792 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA $p>0.9$, n=3)
- **c.** Representative images of migrated and invaded cells in transwell assay for H1792 transduced with indicated sgRNAs.

Extended Data Fig 7. CRISPR deletion of both Aurora kinase A and B suppress migration and invasion in LUAD cells.

- **a.** Western blot for H1792 cells transduced with indicated sgRNAs.
- **b.** Quantitation of percent migration and invasion of H1792 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA $p<0.0001$, n=3)
- **c.** Representative images of migrated and invaded cells in transwell assay for H1792 transduced with indicated sgRNAs.

Extended Data Fig 8. Pan aurora kinase inhibitors suppress migration and invasion in panel of invasive LUAD cells via suppressing aurora kinases activity.

- **a.** Quantification of %Migration, %Invasion from transwell migration invasion assays respectively, and migration index from wound healing assay, for panel of 5 invasive LUAD cells treated with DMSO, AMG900 (0.1μM, 1μM) and PF-03814735 (0.1μM, 1μM). Data presented as mean ± s.e.m. Significant comparison from two-way ANOVA test (n=3) is marked with asterisk.
- **b.** Relative viability of panel of 5 invasive LUAD cells treated with DMSO vs serial dilution of AMG900 or PF-03814735 for 48hr. n=3, Data presented as mean ±s.e.m.

Extended Data Fig 9. Aurora kinases drive invasiveness in lung adenocarcinoma through activating AKT/mTOR and EMT pathways

- **a.** Differentially expressed genes between DMSO and 0.1μM AMG900 treated H1792 cells.
- **b.** Top 10 down-regulated hallmark pathways by AMG900 treatment.
- **c.** Overlaying the DEGs onto the TPX2/AURKB and COL1A2 subnetworks. Nodes filled in red are up-regulated and ones in green are down-regulated by AMG900 treatment. Genes included in HALLMARK_MTORC1_SIGLANING pathways are indicated with diamond shapes (◊).
- **d.** Western blot for indicated proteins in H1792 cells treated with DMSO and indicated concentrations of AMG900 for 48hr.

Extended Data Fig 10. Aurora kinase inhibition suppresses progression of invasive LUAD in Kras(G12D)/TGFBR2$^{-/-}$ mouse model

a. Waterfall plot showing percent change in tumor burden from micro-CT data for vehicle and AMG900 treated animals. (n=3)

b. (left) Representative images showing annotated invasive and lepidic tumor areas in vehicle and AMG900 treated animal's lung. (right) photographs of dissected lungs from vehicle and AMG900 treated animals. Several tumor nodules (arrows) are seen by naked eye in vehicle treated animal lung.

c. Waterfall plot showing percent change in invasive and lepidic tumor areas from histopathological analysis for vehicle and AMG900 treated animals. (n=2)

d. Average vessel density from CD31 IHC staining for tumor infiltrating neovessels in in vehicle and AMG900 treated animal's lung. (t-test p=0.03, n=2)

e. Average vessel lumen area quantified from CD31 IHC staining in vehicle and AMG900 treated animal's lung. Most vessels in the vehicle group tumor appeared collapsed while

in AMG900 treated mouse tumor vessels were wider, however the difference wasn't significant between both groups. (t-test p=0.42, n=2)

f. Quantification of collagen deposition from Masson's trichome staining in vehicle and AMG900 treated animal's lung. (t-test p=0.043 n=2)

g. Average animal weight for vehicle and AMG900 treated animals after treatment starting till week 18. Drop in animal weight in vehicle group reflects loss in body weight of sick animal, which died at that time point.

References

1. Bray, F.*, et al.* Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394-424 (2018).
2. Howlader, N.*, et al.* The Effect of Advances in Lung-Cancer Treatment on Population Mortality. *N Engl J Med* **383**, 640-649 (2020).
3. Uramoto, H. & Tanaka, F. Recurrence after surgery in patients with NSCLC. *Transl Lung Cancer Res* **3**, 242-249 (2014).
4. Borczuk, A.C., Kim, H.K., Yegen, H.A., Friedman, R.A. & Powell, C.A. Lung adenocarcinoma global profiling identifies type II transforming growth factor-beta receptor as a repressor of invasiveness. *Am J Respir Crit Care Med* **172**, 729-737 (2005).
5. Myers, D.J. & Wallen, J.M. Cancer, Lung Adenocarcinoma. in *StatPearls* (Treasure Island (FL), 2020).
6. Travis, W.D.*, et al.* International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* **6**, 244-285 (2011).
7. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).
8. Yu, K.H.*, et al.* Association of Omics Features with Histopathology Patterns in Lung Adenocarcinoma. *Cell Syst* **5**, 620-627 e623 (2017).
9. Xing, X.*, et al.* Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. *Sci Adv* **7**(2021).
10. Toonkel, R.L., Borczuk, A.C. & Powell, C.A. Tgf-beta signaling pathway in lung adenocarcinoma invasion. *J Thorac Oncol* **5**, 153-157 (2010).
11. Wilgus, M.L.*, et al.* Lysyl oxidase: a lung adenocarcinoma biomarker of invasion and survival. *Cancer* **117**, 2186-2191 (2011).
12. Pikor, L.A., Ramnarine, V.R., Lam, S. & Lam, W.L. Genetic alterations defining NSCLC subtypes and their therapeutic implications. *Lung Cancer* **82**, 179-189 (2013).
13. Bird, A.W. & Hyman, A.A. Building a spindle of the correct length in human cells requires the interaction between TPX2 and Aurora A. *J Cell Biol* **182**, 289-300 (2008).
14. Moss, D.K., Wilde, A. & Lane, J.D. Dynamic release of nuclear RanGTP triggers TPX2-dependent microtubule assembly during the apoptotic execution phase. *J Cell Sci* **122**, 644-655 (2009).
15. Barr, A.R. & Gergely, F. Aurora-A: the maker and breaker of spindle poles. *J Cell Sci* **120**, 2987-2996 (2007).
16. Shrestha, R.L.*, et al.* Aurora-B kinase pathway controls the lateral to end-on conversion of kinetochore-microtubule attachments in human cells. *Nat Commun* **8**, 150 (2017).
17. D'Assoro, A.B.*, et al.* The mitotic kinase Aurora--a promotes distant metastases by inducing epithelial-to-mesenchymal transition in ERalpha(+) breast cancer cells. *Oncogene* **33**, 599-610 (2014).
18. Li, D.*, et al.* Overexpression of oncogenic STK15/BTAK/Aurora A kinase in human pancreatic cancer. *Clin Cancer Res* **9**, 991-997 (2003).
19. Subramanian, A.*, et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).

20. Liberzon, A.*, et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* **1**, 417-425 (2015).
21. Director's Challenge Consortium for the Molecular Classification of Lung, A.*, et al.* Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* **14**, 822-827 (2008).
22. Poola, I.*, et al.* Identification of MMP-1 as a putative breast cancer predictive marker by global gene expression analysis. *Nat Med* **11**, 481-483 (2005).
23. Gotzmann, J.*, et al.* A crucial function of PDGF in TGF-beta-mediated cancer progression of hepatocytes. *Oncogene* **25**, 3170-3185 (2006).
24. Jechlinger, M.*, et al.* Expression profiling of epithelial plasticity in tumor progression. *Oncogene* **22**, 7155-7169 (2003).
25. Sarrio, D.*, et al.* Epithelial-mesenchymal transition in breast cancer relates to the basal-like phenotype. *Cancer Res* **68**, 989-997 (2008).
26. Jaeger, J.*, et al.* Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. *Clin Cancer Res* **13**, 806-815 (2007).
27. Bidus, M.A.*, et al.* Prediction of lymph node metastasis in patients with endometrioid endometrial cancer using expression microarray. *Clin Cancer Res* **12**, 83-88 (2006).
28. Liao, Y.L.*, et al.* Identification of SOX4 target genes using phylogenetic footprinting-based prediction from expression microarrays suggests that overexpression of SOX4 potentiates metastasis in hepatocellular carcinoma. *Oncogene* **27**, 5578-5589 (2008).
29. Lu, Z.*, et al.* Long non-coding RNA HULC promotes tumor angiogenesis in liver cancer by up-regulating sphingosine kinase 1 (SPHK1). *Oncotarget* **7**, 241-254 (2016).
30. Vart, R.J.*, et al.* Kaposi's sarcoma-associated herpesvirus-encoded interleukin-6 and G-protein-coupled receptor regulate angiopoietin-2 expression in lymphatic endothelial cells. *Cancer Res* **67**, 4042-4051 (2007).
31. Zhao, M., Kim, P., Mitra, R., Zhao, J. & Zhao, Z. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. *Nucleic Acids Res* **44**, D1023-1031 (2016).
32. Borczuk, A.C.*, et al.* Progression of human bronchioloalveolar carcinoma to invasive adenocarcinoma is modeled in a transgenic mouse model of K-ras-induced lung cancer by loss of the TGF-beta type II receptor. *Cancer Res* **71**, 6665-6675 (2011).
33. Barretina, J.*, et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).
34. Zhu, J.*, et al.* Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. *Nat Genet* **40**, 854-861 (2008).
35. Zhu, J.*, et al.* Stitching together multiple data dimensions reveals interacting metabolomic and transcriptomic networks that modulate cell regulation. *PLoS Biol* **10**, e1001301 (2012).
36. Tsai, M.Y.*, et al.* A Ran signalling pathway mediated by the mitotic kinase Aurora A in spindle assembly. *Nat Cell Biol* **5**, 242-248 (2003).
37. Takahashi, Y.*, et al.* The AURKA/TPX2 axis drives colon tumorigenesis cooperatively with MYC. *Ann Oncol* **26**, 935-942 (2015).
38. Gautschi, O.*, et al.* Aurora kinases as anticancer drug targets. *Clin Cancer Res* **14**, 1639-1648 (2008).
39. Bavetsias, V. & Linardopoulos, S. Aurora Kinase Inhibitors: Current Status and Outlook. *Front Oncol* **5**, 278 (2015).

40. Subramanian, A.*, et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **171**, 1437-1452 e1417 (2017).
41. Bischoff, J.R.*, et al.* A homologue of Drosophila aurora kinase is oncogenic and amplified in human colorectal cancers. *EMBO J* **17**, 3052-3065 (1998).
42. Sen, S., Zhou, H. & White, R.A. A putative serine/threonine kinase encoding gene BTAK on chromosome 20q13 is amplified and overexpressed in human breast cancer cell lines. *Oncogene* **14**, 2195-2200 (1997).
43. Tanaka, T.*, et al.* Centrosomal kinase AIK1 is overexpressed in invasive ductal carcinoma of the breast. *Cancer Res* **59**, 2041-2044 (1999).
44. Lin, Z.Z.*, et al.* Significance of Aurora B overexpression in hepatocellular carcinoma. Aurora B Overexpression in HCC. *BMC Cancer* **10**, 461 (2010).
45. Vischioni, B., Oudejans, J.J., Vos, W., Rodriguez, J.A. & Giaccone, G. Frequent overexpression of aurora B kinase, a novel drug target, in non-small cell lung carcinoma patients. *Mol Cancer Ther* **5**, 2905-2913 (2006).
46. Qi, G.*, et al.* Aurora-B expression and its correlation with cell proliferation and metastasis in oral cancer. *Virchows Arch* **450**, 297-302 (2007).
47. Willems, E.*, et al.* The functional diversity of Aurora kinases: a comprehensive review. *Cell Div* **13**, 7 (2018).
48. Kantarjian, H.M.*, et al.* A phase 1 study of AMG 900, an orally administered pan-aurora kinase inhibitor, in adult patients with acute myeloid leukemia. *Am J Hematol* **92**, 660-667 (2017).
49. Carducci, M.*, et al.* A phase 1, first-in-human study of AMG 900, an orally administered pan-Aurora kinase inhibitor, in adult patients with advanced solid tumors. *Invest New Drugs* **36**, 1060-1071 (2018).
50. Diamond, J.R.*, et al.* A phase II clinical trial of the Aurora and angiogenic kinase inhibitor ENMD-2076 for previously treated, advanced, or metastatic triple-negative breast cancer. *Breast Cancer Res* **20**, 82 (2018).
51. Schoffski, P.*, et al.* Phase I, open-label, multicentre, dose-escalation, pharmacokinetic and pharmacodynamic trial of the oral aurora kinase inhibitor PF-03814735 in advanced solid tumours. *Eur J Cancer* **47**, 2256-2264 (2011).
52. Zhou, H.Y. & Wong, A.S. Activation of p70S6K induces expression of matrix metalloproteinase 9 associated with hepatocyte growth factor-mediated invasion in human ovarian cancer cells. *Endocrinology* **147**, 2557-2566 (2006).
53. Carmena, M. & Earnshaw, W.C. The cellular geography of aurora kinases. *Nat Rev Mol Cell Biol* **4**, 842-854 (2003).
54. Carmena, M., Ruchaud, S. & Earnshaw, W.C. Making the Auroras glow: regulation of Aurora A and B kinase function by interacting proteins. *Curr Opin Cell Biol* **21**, 796-805 (2009).
55. Wang, G., Jiang, Q. & Zhang, C. The role of mitotic kinases in coupling the centrosome cycle with the assembly of the mitotic spindle. *J Cell Sci* **127**, 4111-4122 (2014).
56. Gritsko, T.M.*, et al.* Activation and overexpression of centrosome kinase BTAK/Aurora-A in human ovarian cancer. *Clin Cancer Res* **9**, 1420-1426 (2003).
57. Araki, K., Nozaki, K., Ueba, T., Tatsuka, M. & Hashimoto, N. High expression of Aurora-B/Aurora and Ipll-like midbody-associated protein (AIM-1) in astrocytomas. *J Neurooncol* **67**, 53-64 (2004).
58. Wang, X.*, et al.* Overexpression of Aurora-A enhances invasion and matrix metalloproteinase-2 expression in esophageal squamous cell carcinoma cells. *Mol Cancer Res* **10**, 588-596 (2012).

59. Chen, C.H.*, et al.* Overexpression of Rap-1A indicates a poor prognosis for oral cavity squamous cell carcinoma and promotes tumor cell invasion via Aurora-A modulation. *Am J Pathol* **182**, 516-528 (2013).
60. He, J.Y.*, et al.* Knockdown of Aurora-B alters osteosarcoma cell malignant phenotype via decreasing phosphorylation of VCP and NF-kappaB signaling. *Tumour Biol* **36**, 3895-3902 (2015).
61. Zhu, X.P.*, et al.* Inhibition of Aurora-B suppresses osteosarcoma cell migration and invasion. *Exp Ther Med* **7**, 560-564 (2014).
62. Matej, R., Rohan, Z., Nemejcova, K. & Dundr, P. [Molecular pathology of lung cancer in routine diagnostic practice: 2017 update]. *Cesk Patol* **53**, 159-166.
63. Wu, Y.L.*, et al.* Osimertinib in Resected EGFR-Mutated Non-Small-Cell Lung Cancer. *N Engl J Med* **383**, 1711-1723 (2020).
64. Trapnell, C.*, et al.* Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7**, 562-578 (2012).
65. Love, M.I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
66. Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-550 (2014).
67. Okayama, H.*, et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* **72**, 100-111 (2012).
68. Der, S.D.*, et al.* Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J Thorac Oncol* **9**, 59-64 (2014).
69. Tang, H.*, et al.* A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin Cancer Res* **19**, 1577-1586 (2013).
70. Rousseaux, S.*, et al.* Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med* **5**, 186ra166 (2013).
71. Wilkerson, M.D.*, et al.* Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS One* **7**, e36530 (2012).
72. Das, J., Gayvert, K.M., Bunea, F., Wegkamp, M.H. & Yu, H. ENCAPP: elastic-net-based prognosis prediction and biomarker discovery for human cancers. *BMC Genomics* **16**, 263 (2015).
73. Zhu, H. & Hastie, T. Regularization and Variable Selection via the Elastic Net. *ournal of the Royal Statistical Society. Series B (Statistical Methodology)* **67**, 301-320 (2005).
74. Liu, J.*, et al.* An Integrated TCGA Pan-Cancer Clinical Data Resource to Drive High-Quality Survival Outcome Analytics. *Cell* **173**, 400-416 e411 (2018).
75. Yoo, S.*, et al.* MODMatcher: multi-omics data matcher for integrative genomic analysis. *PLoS Comput Biol* **10**, e1003790 (2014).
76. Friedman, N., Linial, M., Nachman, I. & Pe'er, D. Using Bayesian networks to analyze expression data. *J Comput Biol* **7**, 601-620 (2000).
77. Zhu, J.*, et al.* Increasing the Power to Detect Causal Associations by Combining Genotypic and Expression Data in Segregating Populations. *PLoS Comput Biol* **3**, e69 (2007).
78. Su, G., Morris, J.H., Demchak, B. & Bader, G.D. Biological network exploration with Cytoscape 3. *Curr Protoc Bioinformatics* **47**, 8 13 11-24 (2014).
79. Wang, I.M.*, et al.* Systems analysis of eleven rodent disease models reveals an inflammatome signature and key drivers. *Mol Syst Biol* **8**, 594 (2012).
80. Liu, Y.*, et al.* A Network Analysis of Multiple Myeloma Related Gene Signatures. *Cancers (Basel)* **11**(2019).

81. Corsello, S.M.*, et al.* The Drug Repurposing Hub: a next-generation drug library and information resource. *Nat Med* **23**, 405-408 (2017).
82. Brum, A.M.*, et al.* Connectivity Map-based discovery of parbendazole reveals targetable human osteogenic pathway. *Proc Natl Acad Sci U S A* **112**, 12711-12716 (2015).
83. Doench, J.G.*, et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).
84. Payton, M.*, et al.* Preclinical evaluation of AMG 900, a novel potent and highly selective pan-aurora kinase inhibitor with activity in taxane-resistant tumor cell lines. *Cancer Res* **70**, 9846-9854 (2010).

Figures

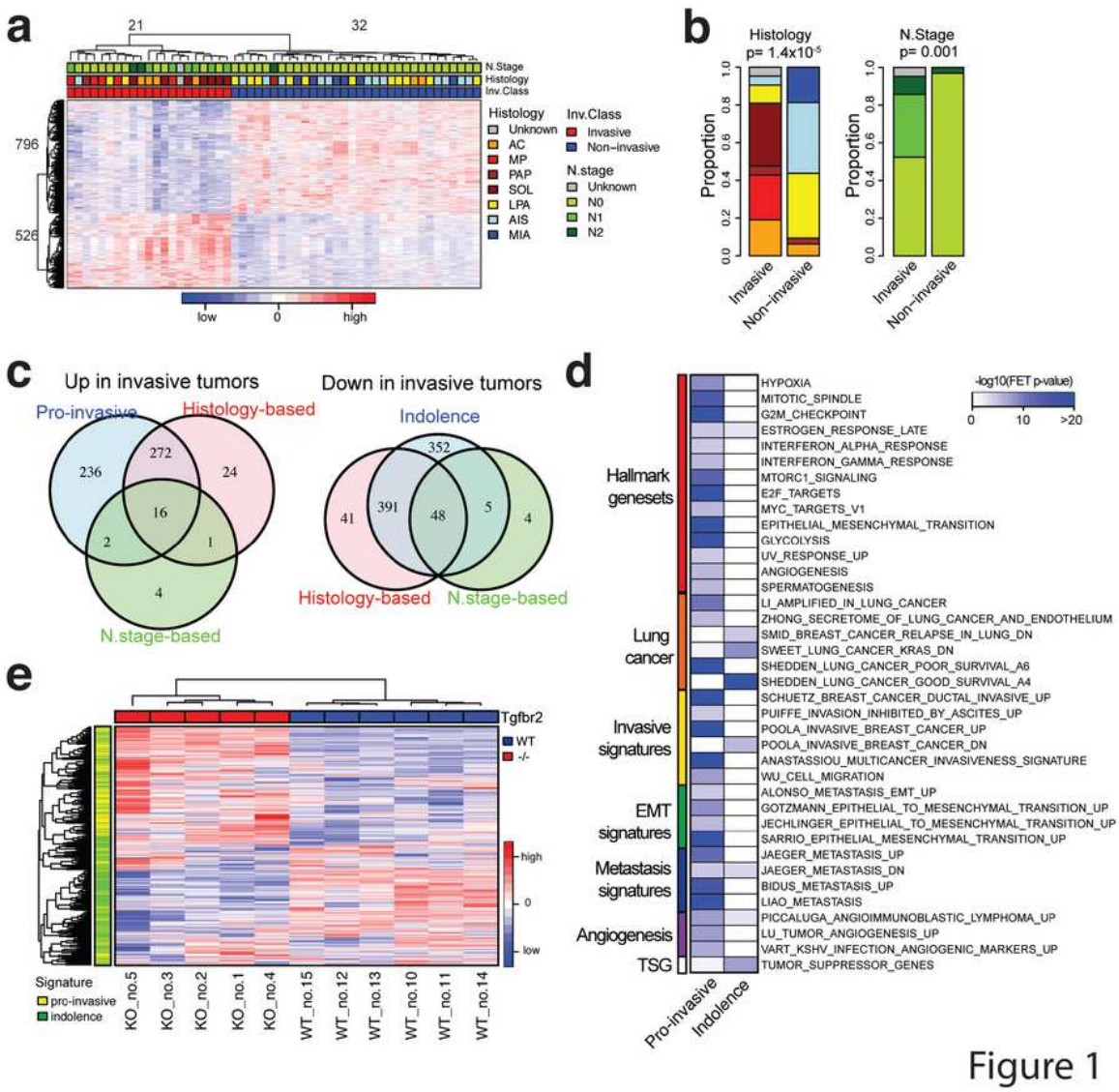

Figure 1

Figure 1

Pro-invasive and indolence signature genes associated with tumor invasion at early-stage lung adenocarcinoma. a. Invasiveness signature genes separating 53 early-stage lung adenocarcinoma into invasive and non-invasive tumors. “complete” method was used for unsupervised clustering. b.

Comparison of the tumor clusters based on the signature genes with histological subtypes and nodal stages of the tumors. P-values of association were measured by Chi-Square test. c. Comparison of the gene numbers of our signatures (pro-invasive and indolence signatures) with differentially expressed genes based on histological subtypes (AIS, MIA, LPA vs AC, MP, PAP, SOL) and nodal stage (N0 vs N1+). d. Association of pro-invasive and indolence signatures with MSigDB genesets including Hallmark, lung cancer related genes, tumor invasion or metastasis related genesets. e. Unsupervised clustering of 11 tumors (5 Tgfbr2 -/- and 6 Tgfbr2 wt mice from GSE27717) based on the signature genes. “complete” method was used for unsupervised clustering.

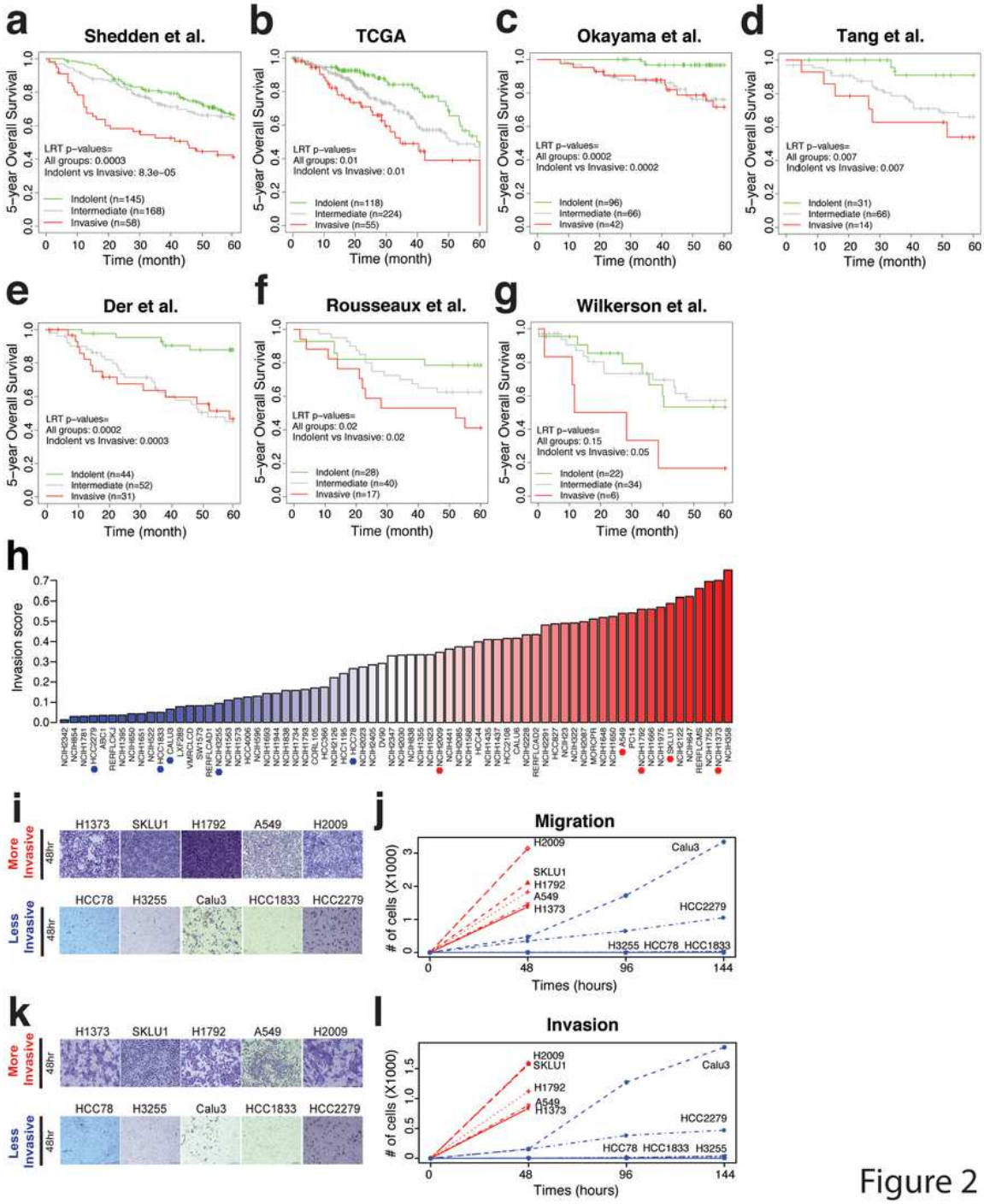

Figure 2

Figure 2

Invasiveness score associated with patient survival and tumor cells phenotype a. Stratification of tumor samples of Shedden et al. cohort into three groups; invasive (high IVS), intermediate (middle IVS), and indolent (low IVS) tumors determined based on IVS (Methods, Extended Data Fig 2a). Five-year survival of tumors was shown in a KM curve with corresponding LRT p-values. b. KM plot for TCGA LUAD cohort. c. KM plot for Okayama et al. cohort. d. KM plot for Tang et al. cohort. e. KM plot for Der et al. cohort. f.

KM plot for Rousseaux et al. cohort. g. KM plot for Wilkerson et al. cohort. h. Relative invasiveness scores (values 0-1) of 70 LUAD cell lines estimated via IVS. Tumor cells above median score were labeled as “More Invasive” and below as “Less Invasive”. In each group, we selected 5 for further experimental validations as marked in asterisk (red for more invasive and blue for less invasive cells). i. Representative images of migrated cells in transwell migration assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells. j. Quantification of migrated cells through transwell at 48hr, 96hr, and 144hr. Data presented as mean (n=3). k. Representative images of invaded cells in transwell matrigel invasion assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells l. Quantification of invaded cells through transwell matrigel at 48hr, 96hr, and 144hr. Data presented as mean (n=3).

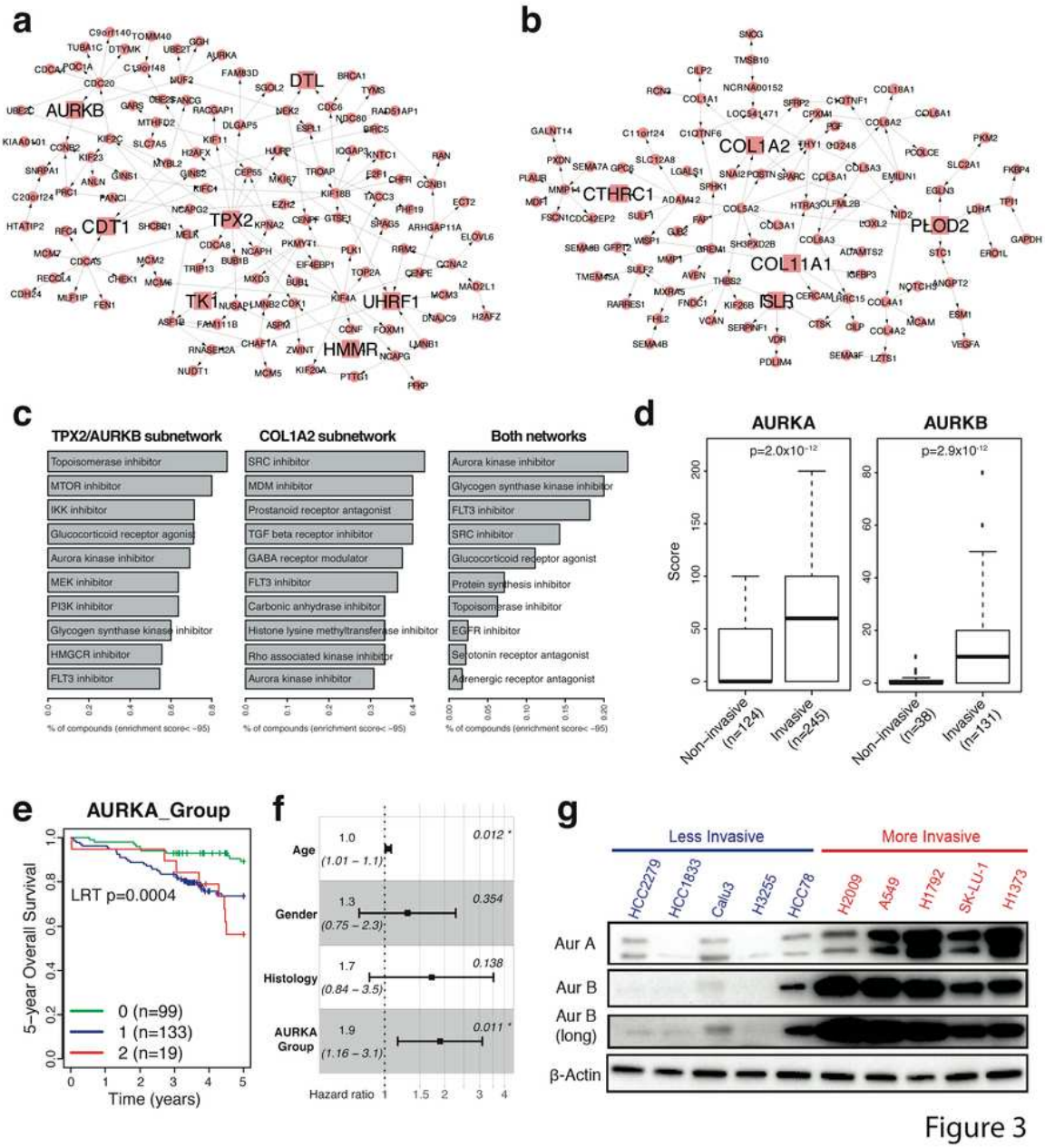

Figure 3

Figure 3

Invasiveness score associated with patient survival and tumor cells phenotype a. Stratification of tumor samples of Shedden et al. cohort into three groups; invasive (high IVS), intermediate (middle IVS), and indolent (low IVS) tumors determined based on IVS (Methods, Extended Data Fig 2a). Five-year survival of tumors was shown in a KM curve with corresponding LRT p-values. b. KM plot for TCGA LUAD cohort. c. KM plot for Okayama et al. cohort. d. KM plot for Tang et al. cohort. e. KM plot for Der et al. cohort. f.

KM plot for Rousseaux et al. cohort. g. KM plot for Wilkerson et al. cohort. h. Relative invasiveness scores (values 0-1) of 70 LUAD cell lines estimated via IVS. Tumor cells above median score were labeled as “More Invasive” and below as “Less Invasive”. In each group, we selected 5 for further experimental validations as marked in asterisk (red for more invasive and blue for less invasive cells). i. Representative images of migrated cells in transwell migration assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells. j. Quantification of migrated cells through transwell at 48hr, 96hr, and 144hr. Data presented as mean (n=3). k. Representative images of invaded cells in transwell matrigel invasion assay after 48 hours in 5 more invasive and 5 less-invasive LUAD cells l. Quantification of invaded cells through transwell matrigel at 48hr, 96hr, and 144hr. Data presented as mean (n=3).

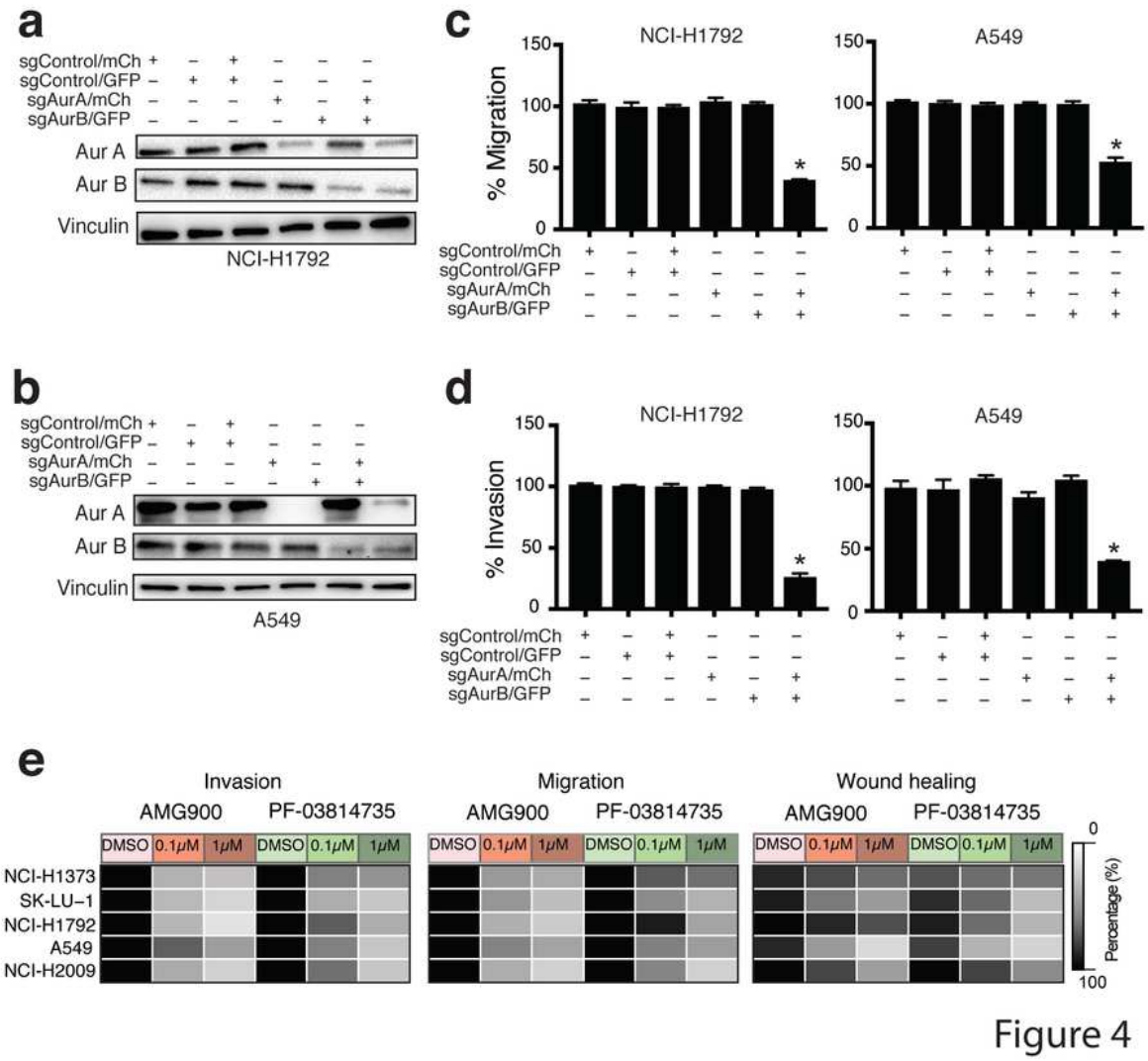

Figure 4

Figure 4

Aurora kinase A and B are both regulators of invasive phenotype. a. Western blot for H1792 cells transduced with indicated sgRNAs. b. Western blot for A549 cells transduced with indicated sgRNAs c. Quantitation of percent migration of H1792 and A549 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA n=3, $p<0.0001$) d. Quantitation of percent invasion of H1792 and A549 cells transduced with indicated sgRNAs at 48hr. (one-way ANOVA n=3, $p<0.0001$) e. Heat map for H1373, SK-LU-1, H1792,

A549 and H2009 showing effect of AMG900 and PF-03814735 on invasion and migration phenotype, and wound healing ability, at indicated drug concentrations. (n=3, statistical test results in Supplementary Table 9)

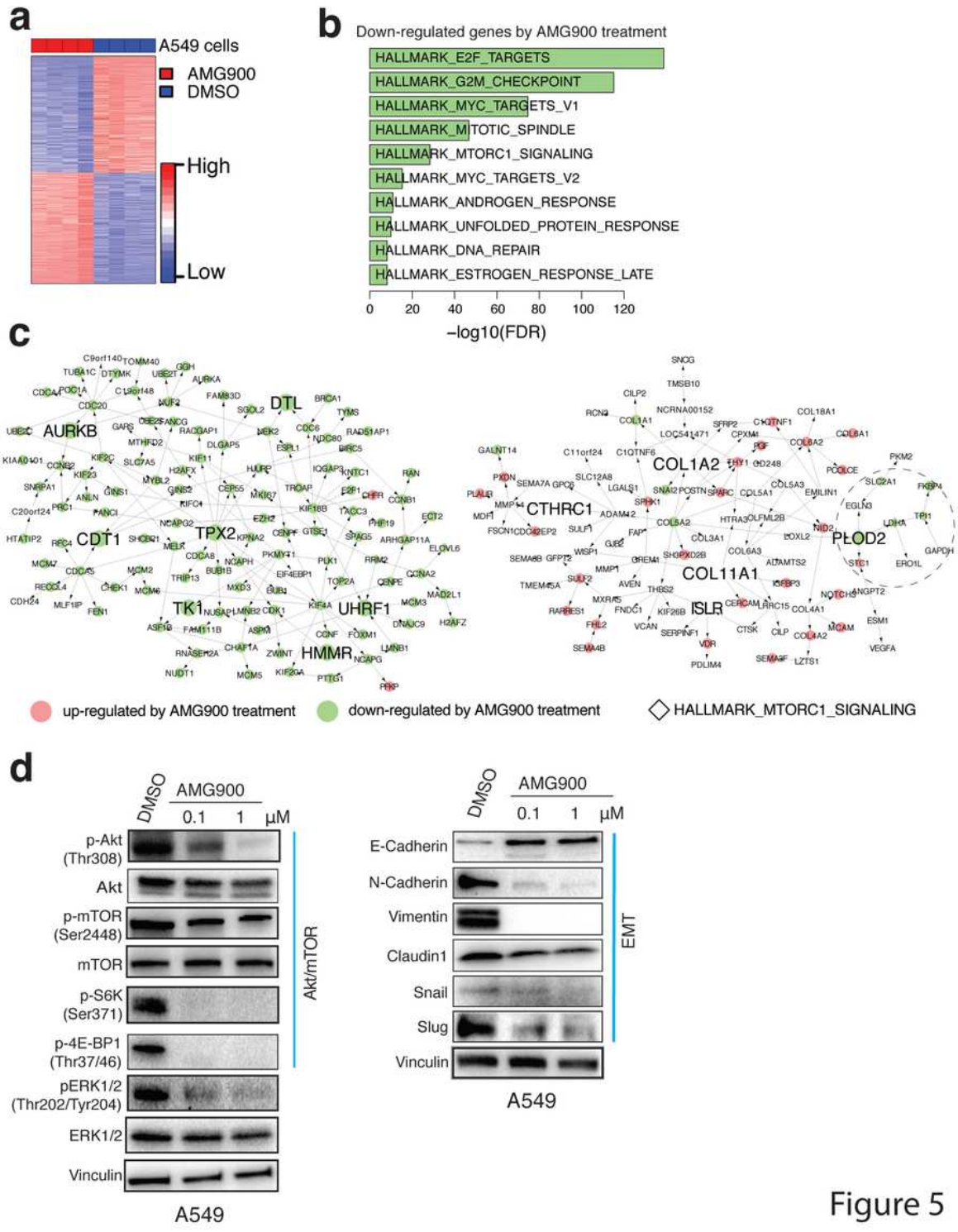

Figure 5

Figure 5

Aurora kinases drive invasiveness in lung adenocarcinoma through activating AKT/mTOR and EMT pathways. a. Differentially expressed genes between DMSO and 0.1μM AMG900 treated A549 cells

(FDR<0.01) b. Top 10 down-regulated HALLMARK pathways in A549 treated with 0.1μM AMG900. c. Overlaying the DEGs in A549 cells onto the TPX2/AURKB and COL1A2 subnetworks. Nodes filled in red are up-regulated and ones in green are down-regulated by AMG900 treatment. Genes included in HALLMARK_MTORC1_SIGLANING pathways are indicated with diamonds (◊) within a dashed circle. d. Western blot for A549 cells treated with DMSO and indicated concentrations of AMG900 for 48hr for AKT/mTOR pathway and EMT pathway.

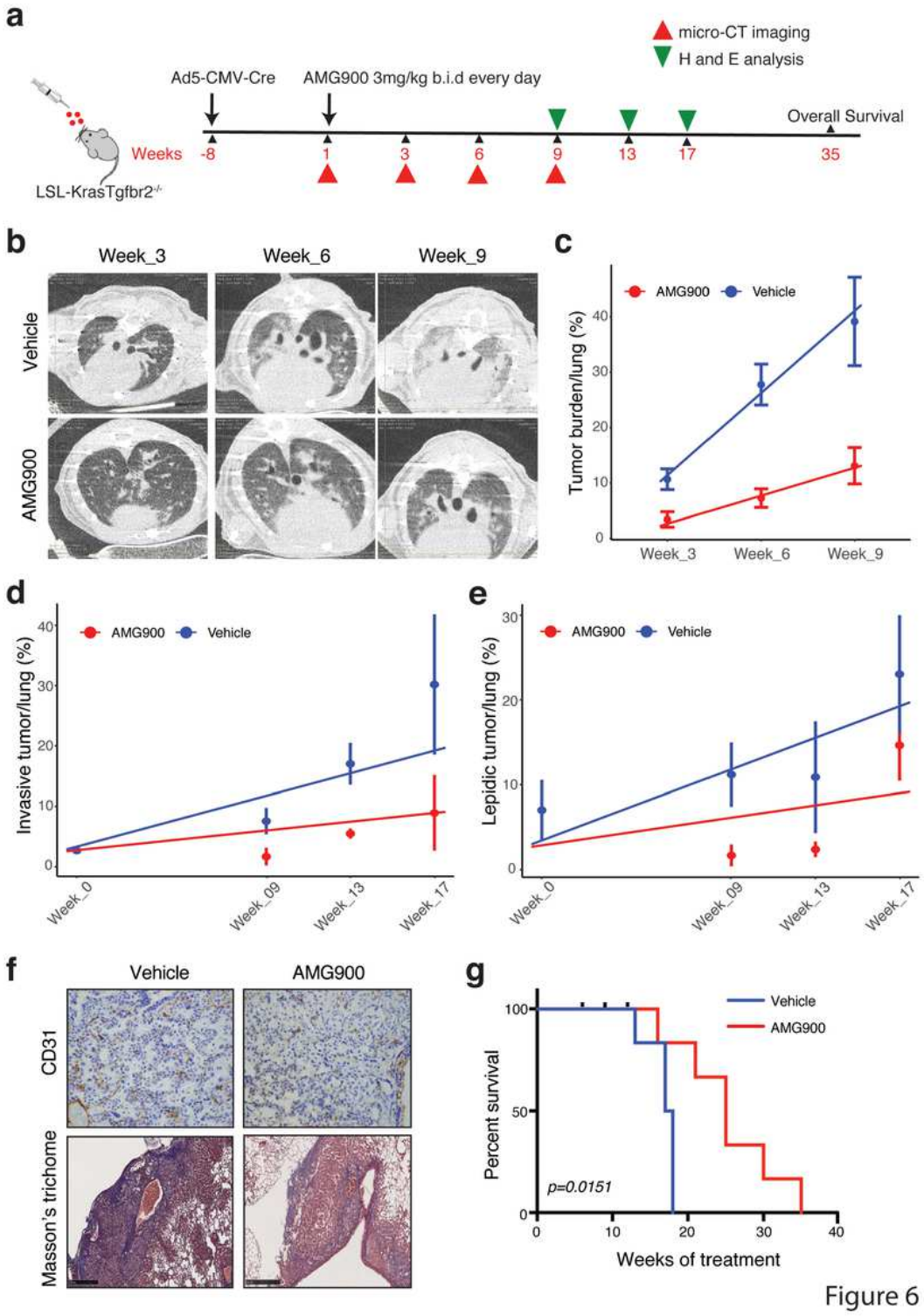

Figure 6

Figure 6

Aurora kinase inhibition suppresses invasive lung adenocarcinoma progression in Kras(G12D)/Tgfbr2-/- mouse model. a. Scheme of treatment of Kras(G12D)/Tgfbr2-/- mouse with AMG900. b. micro-CT images of vehicle and AMG900 treated mice at week 3, 6 and 9 after treatment. Treatment was given at 3mg/kg twice daily by oral gavage. (vehicle n=10, AMG900 n=8) c. Quantitation of tumor burden from micro-CT imaging for vehicle and AMG900 treated mice at week 3, 6 and 9 after treatment. (n=3). A fitting line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice. d. Comparison of invasive tumor percentage from vehicle and AMG900 treated mice between baseline and week 9, 13 and 17 from histopathological analysis. (n>=2). A fitting line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice. e. Comparison of lepidic tumor percentage from vehicle and AMG900 treated mice between baseline and week 9, 13, and 17 from histopathological analysis. (n>=2). A fit line from a linear regression model (percentage ~ time) is shown for vehicle (blue) and AMG900 treated (red) mice. f. (Top) IHC images for CD31 staining in Vehicle and AMG900 treated mouse tumor. (Bottom) Masson's trichome staining in Vehicle and AMG900 treated mouse lung. g. Survival curve for vehicle and AMG900 treated mice. (vehicle n=10, AMG900 n=8). LRT p-value was measured (p=0.015).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- STable1.pdf
- STable2.pdf
- STable3.pdf
- STable4.pdf
- STable5.pdf
- STable6.pdf
- STable7.pdf
- STable8.pdf
- STable9.pdf
- STable10.pdf
- STable11.pdf
- STable12.pdf
- STable13.pdf
- SupplementaryInformation.docx
- ExtendedDataFigures.pdf