Research
A SCIENCE PARTNER JOURNAL

RESEARCH ARTICLE

Evolution of Cancer Metastases via Lineage Trans-Differentiation

Yu Xiao[1,2†], Wan Jin[1,2†], Fangjin Chen[3], Kaiyu Qian[4,5], Lingao Ju[1,6*], and Yi Zhang[2,7*]

[1]Department of Biological Repositories, Human Genetic Resources Preservation Center of Hubei Province, Laboratory of Precision Medicine, Zhongnan Hospital of Wuhan University, Wuhan, China. [2]Institute for Genetics and Molecular Medicine, Chinese Institutes for Medical Research, Beijing, China. [3]High Performance Computing Center, Peking-Tsinghua College of Life Sciences, Peking University, Beijing, China. [4]Department of Urology, Hubei Key Laboratory of Urological Diseases, Zhongnan Hospital of Wuhan University, Wuhan, China. [5]Medical Research Institute, Frontier Science Center for Immunology and Metabolism, Wuhan University, Wuhan, China. [6]Wuhan Research Center for Infectious Diseases and Cancer, Chinese Academy of Medical Sciences, Wuhan, China. [7]School of Basic Medical Sciences, Capital Medical University, Beijing, China.

*Address correspondence to: julingao1990@whu.edu.cn (L.J.); zhangyi@cimrbj.ac.cn (Y.Z.)
†These authors contributed equally to this work.

Citation: Xiao Y, Jin W, Chen F, Qian K, Ju L, Zhang Y. Evolution of Cancer Metastases via Lineage Trans-Differentiation. *Research* 2026;9:Article 1144. https://doi.org/10.34133/research.1144

Submitted 26 September 2025
Revised 5 January 2026
Accepted 22 January 2026
Published 19 February 2026

Copyright © 2026 Yu Xiao et al. Exclusive licensee Science and Technology Review Publishing House. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License (CC BY 4.0).

Most cancer metastases exhibit mutational profiles similar to those of the primary tumor. However, the nongenetic mechanisms driving metastasis remain poorly understood. Here, we show that lineage trans-differentiation is a hallmark of cancer metastasis. Bioinformatic tools capable of reconstructing cancer phenotypic evolutionary trajectories at single-cell resolution were developed, revealing a progressive loss of transcriptional and epigenomic lineage fidelity in cancer cells as they evolve toward metastasis. During premetastatic evolution, cells undergo de-differentiation into a fetal-like state. The mis-expression of alternative-lineage gene programs during re-differentiation from this fetal-like state leads to the formation of trans-differentiated metastatic cells. This trans-differentiation, rather than fetal-like transcription, constitutes a key feature of metastasis in both humans and mice. In clinical samples, trans-differentiation correlates with histopathological grade, metastatic potential, and patient survival. Additionally, trans-differentiation is associated with gain of oncogenic mitogen-activated protein kinase (MAPK) signaling and can be reversed through MAPK inhibition. These findings offer a detailed account of metastatic cancer evolution driven by epigenetic reprogramming while also uncovering the molecular mechanisms underlying this process.

Introduction

Metastatic disease is the leading cause of cancer-related mortality [1]. Extensive molecular genetic profiling has failed to identify consistent alterations in metastatic tumors compared to their primary counterparts [2–9]. Generally, metastases share the same spectrum of driver mutations as the primary tumor, suggesting that new mutations are passive "passengers" that are not required for metastatic dissemination. Although certain metastasis-specific chromosomal aberrations (such as karyotypic or copy number changes) have been observed in some cancers, these are not universally present, and their functional relevance remains unclear [10,11]. In most cases, the genetic mutations in primary tumors appear to facilitate metastatic outgrowth, but do not account for why only a minority of cancer cells in the primary tumor successfully establish distant metastases.

Transcriptional and epigenetic profiling has revealed that metastatic cancers often exhibit substantial intratumoral heterogeneity (ITH), which cannot be fully explained by additional mutations [12–15]. This observation has led to the hypothesis that metastatic potential is associated with a highly plastic subpopulation of cells—often termed "cancer stem cells" or "high plasticity progenitor cells"—which adopt progenitor-like, developmental programs [16–18] that resemble a "fetal-like" state. In fact, many aggressive tumors display gene expression signatures indicative of early embryonic or "fetal-like" cell states, including precursor-, germline-, and placental-like transcriptional programs [19–24]. However, reverting tumor cells to an embryonic state alone is insufficient for malignant progression, as demonstrated by the benign nature of teratomas.

Metastatic tumors are commonly associated with features of epithelial–mesenchymal transition (EMT), an embryonic program that transforms epithelial cells into a mesenchymal

phenotype, leading to the loss of apical–basal polarity and enhanced cellular motility and invasiveness [25–27]. Although EMT is often regarded as a phase-shift-like event during tumor progression, transcriptional profiling has revealed that EMT represents a combination of distinct molecular states. Single-cell analyses have identified a hybrid or partial EMT state as permissive for metastasis [28–32], while cells undergoing full EMT tend to enter a quiescent, dormant state, limiting their metastatic potential [33]. In clinical settings, tumors with high Ki67 expression are typically considered more aggressive. Additionally, single-cell DNA sequencing has shown that highly proliferative cancer subclones are more efficient at forming metastases. Consequently, the EMT phenotype does not always correlate with metastatic potential. Moreover, a causal link between EMT and metastasis has yet to be experimentally established in vivo. Therefore, the role of EMT and fetal-like transcriptional programs in generating ITH and producing metastasis-competent clones remains unclear.

This study provides evidence that metastasis evolves through lineage trans-differentiation—an alteration in cell phenotype across developmental lineage boundaries. By tracking the transcriptional and epigenetic profiles of cancer cells at single-cell resolution, this study demonstrates that tumor cells progressively lose their original lineage identity. They first undergo de-differentiation into a fetal-like, EMT-active state and subsequently re-differentiate into alternative-lineage fates, producing clones that are primed for metastasis. This trans-differentiation process, rather than the fetal-like progenitor state itself, is a defining feature of cancer metastases in both mouse models and human cancers. This process is driven by heightened mitogen-activated protein kinase (MAPK) signaling activity and can be reversed with MAPK inhibition. These findings, consistent with multiple human clinical trial results, suggest that targeted inhibition of lineage trans-differentiation could be a promising strategy to prevent cancer metastasis.

Results

Trans-differentiation of mouse lung cancer during metastasis evolution

To investigate the evolution of metastatic tumors in detail (Fig. 1A), a well-characterized, genetically defined lung adenocarcinoma (LUAD) mouse model was analyzed. In this model, lung alveolar type 2 (AT2) cells were transformed into adenocarcinoma through cell-type-specific Cre-induced overexpression of a constitutively active form of oncogene *Kras* ($Kras^{CA}$) [34], which serves as the central signaling hub in the receptor tyrosine kinase (RTK)–Ras–Raf–MAPK pathway (referred to here as the MAPK pathway) [35] (Fig. S1). The MAPK pathway is one of the most frequently altered molecular pathways in human cancers [36]. In humans, the majority of LUADs are driven by oncogenic mutations that hyperactivate MAPK signaling [37–40].

Multiple published single-cell RNA sequencing (scRNA) [41–43] and single-cell assay for transposase-accessible chromatin using sequencing (ATAC-seq, scATAC) datasets [44–48] were integrated to analyze the in vivo evolution of mouse LUAD samples, spanning early to late metastatic stages (Fig. 1B and C). Single cells in the scRNA dataset were classified based on canonical marker gene expression [49] into normal AT2 (AT2: $Lamp3^+/Sftpb^+/Nkx2\text{-}1^+$), stressed AT2 (trAT2: $Ifit3^+/Irf7^+$), normal AT1 (AT1: $Ager2^+/Hopx^+/Nkx2\text{-}1^+$), bronchioalveolar stem cell-like cells (BASC: $Sox2^+/Cyp2f2^+/Scgb1a1^+/Krt8^+$), primary cancer (Primary: $Gdf15^+/Cd24^+/Apoe^+$), Cldn4$^+$ hyperplastic cancer cells (Cldn4$^+$: $Cldn4^+/Tm4sf1^+/Slc4a11^+$) [41], EMT cells (EMT: Vim^+) [26], and 2 clusters of cells that could be found at metastatic sites [43] (Met1: $Id3^+/Gpx3^+/Sox9^+/Lgals1^+$; Met2: $Id3^+/Gpx3^+/Bmp2^+/Zeb2^+/Runx2^+$) (Fig. 1B and C and Fig. S2A). The scATAC dataset similarly identified AT2, AT1, BASC, primary cancer, Cldn4$^+$, and EMT cells (Fig. 1C), with metastatic cells forming a distinct cluster. Nontumorous ciliated and club cells were also present in the scATAC dataset. Furthermore, a unique cluster of primary cancer cells, found only in KPL (KP, *Lkb1*-deficient) tumors, was identified in the scATAC dataset, distinguishing it from KP ($Kras^{CA}$, $Trp53^{KO}$) or KPS (KP, *Smarca4*-deficient) models.

Genetic lineage tracing of cells from the scRNA dataset [43] revealed a stepwise evolutionary trajectory originating from AT2 cells. AT2 cells first transition to BASC and primary cancer cells, which then give rise to Cldn4$^+$ cells that coexpress airway epithelial progenitor (AEP) markers such as *Tm4sf1* and *Cd24a*. Cldn4$^+$ cells then evolve into EMT cells, which serve as precursors for metastatic (Met1/Met2) cells (Fig. S2B). This evolutionary trajectory was further validated by single-cell replication age estimation [50] in a cell-composition balanced scATAC dataset [46] (Fig. S3A and B). The age of cancer cells progressively increases from AT2-like cells to primary cancer cells, Cldn4$^+$ cells, followed by EMT and metastatic cells (Fig. S3C). Additionally, a phylogenetic analysis of cell clusters based on chromatin accessibility at clock-like loci revealed a trajectory from AT2 through Cldn4$^+$ and EMT cells to metastatic cells (Fig. S3D). Given these results and the observation that some single-cell clusters are not consistently detected across all analyzed datasets, the primary evolutionary trajectory AT2–BASC–Cldn4$^+$–EMT–Met was characterized, with other side branches reserved for future studies.

As previously reported [51–54], $Kras^{CA}$ induces the transformation of AT2 cells toward BASC (Fig. S4A). However, progression to cancer from BASC requires the loss of *Trp53* and functional *Foxa1*/*Foxa2* activity (Fig. S4A). Following *Trp53* loss, transformation from primary adenocarcinoma to Cldn4$^+$, EMT, and metastatic cells occurs gradually over time (Fig. S4B). Consistent with prior studies [1,12–14], it was observed that non-genetic transcriptional reprogramming continues autonomously without the need for additional genetic alterations, suggesting that the evolution may be driven by epigenomic mechanisms.

Next, transcriptional programs associated with metastasis evolution were systematically explored. From $Kras^{CA}$, $Trp53^{KO}$ cancer cells from primary and metastatic tumors (Fig. S5A), 15 unique transcriptional programs ("modules") [55] were algorithmically identified and manually annotated based on their biological characteristics (Fig. S5B). As expected, lung-like modules (AT1, AT2, Goblet, and AEP) gradually decreased during cancer evolution [56–58], while EMT-associated modules (Early EMT, pEMT, Late EMT) increased over time (Fig. 1D and Fig. S5C) [30,59–64]. Expression of the fetal-like transcriptional program [19–21,24,65–68] was elevated in early metastatic cells (Met1), while alternative-lineage modules (neuro-glial, stromal, and endothelial lineages) [69–72] were particularly increased in late metastatic cells (Met2) (Fig. 1D and Fig. S5C). To validate this analysis, a similarity search was performed comparing cancer cells with normal cells from the Mouse Cell Atlas [73]. Early-stage cancer cells exhibited the highest similarity to normal epithelial cells, whereas metastatic

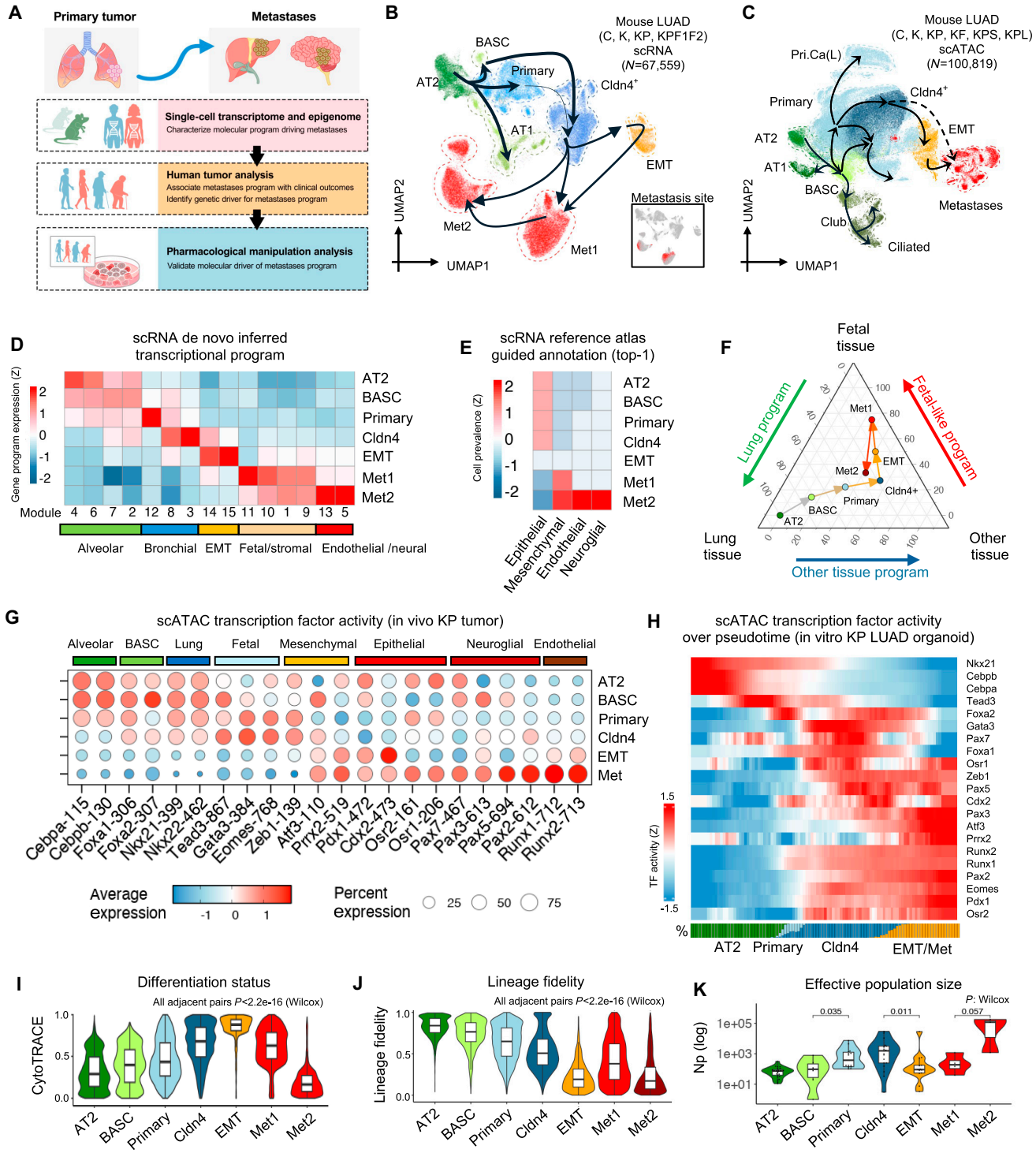


Fig. 1. Trans-differentiation of mouse lung cancer during evolution toward metastases. (A) Study scheme. Metastasis-associated transcriptional programs are identified through comparative analysis of single-cell sequencing data from primary cancer and metastases. Human tumor samples were analyzed to validate the clinical phenotypes linked to the identified programs and to identify the genetic variations driving their expression. Finally, pharmacological perturbation datasets were analyzed to validate the driver events for metastasis programs. (B) UMAP of integrated mouse genetically induced LUAD scRNA cells ($N = 67{,}559$). Inset: Distribution of metastatic cells on the UMAP. (C) UMAP of integrated mouse genetically induced LUAD scATAC cells ($N = 100{,}819$). (D) Relative expression of de novo inferred transcription modules in mouse LUAD scRNA cells. (E) Distribution of the most similar normal cell type from the Mouse Cell Atlas to mouse LUAD scRNA cells. (F) Median expression of lung-like, fetal-like, and alternative-lineage (other tissue) gene programs in each class of mouse LUAD scRNA cells. (G) TF activities of lung-like, fetal-like, and alternative-lineage (other tissue) master TFs in each class of mouse LUAD scATAC cells. (H) TF activities of lung-like, fetal-like, and alternative-lineage (other tissue) master TFs in mouse LUAD organoid scATAC cells along evolution pseudotime. Bottom: Prevalence of each class of cancer cells at given pseudotime. (I) Stemness of mouse LUAD scRNA cells. (J) Lineage fidelity of mouse LUAD scRNA cells. (K) Total size of progenitor cell estimation for each single-cell clusters. *P* value: Wilcox test. Cell clusters: AT2, type 2 alveolar cells; AT1, type 1 alveolar cells; trAT2, stressed AT2 cells; BASC, bronchioalveolar stem cells; Primary, primary cancer cells; Cldn4^{+}, Cldn4^{+} cancer cells; EMT, epithelial–mesenchymal transformation cells; Met1/2, metastatic cluster; Pri.Ca(L), LKB1 mutated primary cancer cells; Club, club cells; Ciliated, ciliated cells of the bronchus. Genotypes: C, wild-type; K, Kras^{CA}; KP, Kras^{CA}/Trp53^{KO}; KF, Kras^{CA}/Trp53^{hypermorph}; KPF1F2, Kras^{CA}/Trp53^{KO}/Foxa1^{KO}/Foxa2^{KO}; KPS, Kras^{CA}/Trp53^{KO}/Smarca1^{KO}; KPL, Kras^{CA}/Trp53^{KO}/Lkb1^{KO}.

cancer cells showed reduced similarity to epithelial cells and increased similarity to mesenchymal, endothelial, or ectoderm-derived neuron or glia (neuroglial) cells (Fig. 1E). Thus, metastatic cancer cells are transcriptionally reprogrammed to express genes associated with alternative lineages.

In metastatic cells, the increased expression of genes with alternative-lineage specificity could result from either the loss of defined lineage specificity (de-differentiation) or the gain of alternative-lineage specificity (trans-differentiation). Furthermore, trans-differentiation could occur either directly from the original lineage or through re-differentiation from a de-differentiated intermediate state. To distinguish between these possibilities, single-cell gene expression profiles were decomposed into lung-like, fetal-like, and alternative-lineage-like components (Fig. S6A and B). During the evolution toward metastatic cancer, cells first transition from a lung-like state to a fetal-like state and then re-differentiate to express alternative-lineage genes (Fig. 1F and Fig. S6A and B). This result was further validated through de novo inferred transcription module expression in single cells (Fig. S6C). Similarly, analysis of transcription factor (TF) activity in in vivo $Kras^{CA}$; $Trp53^{KO}$ (KP) mouse cancer cells and in vitro KP cancer organoid scATAC datasets [74] showed a decrease in the activity of lung-lineage TFs, such as *Nkx2-1*, *Cebpa*, and *Foxa2*, from AT2 cells, followed by the induction of fetal-lineage TFs like *Gata3*, *Tead3*, and *Eomes* in primary cancer and $Cldn4^+$ cells. Alternative-lineage TFs were primarily active in post-EMT metastatic cells (Fig. 1G and H and Fig. S7). In line with this analysis, stemness estimation (cell differentiation state) using cancer cell RNA expression [75] revealed gradual de-differentiation from AT2 cells to EMT cells, followed by re-differentiation from EMT cells to metastatic Met1/Met2 cells (Fig. 1I).

A quantitative metric of single-cell lineage fidelity was introduced (Materials and Methods), defined as the disparity between the highest and remaining expression similarity scores obtained by correlating a single-cell expression profile against a reference normal cell atlas. A large disparity indicates that a cell uniquely aligns with a single lineage, while a small disparity, reflecting comparable similarity across multiple reference types, signifies low lineage fidelity. It was found that cancer cells progressively lose lineage fidelity during metastasis evolution (Fig. 1J).

To validate the decrease in single-cell lineage fidelity during cancer cell evolution, a method was developed using maximum likelihood estimation to compute the probability of a single cell adopting the fate of any reference normal cell in a panel. This was used to derive lineage confusion for each single cell, defined as the total transcriptional distance between reference normal cells most likely to match it (Materials and Methods). The lineage confusion of genetically traced cancer clones [43] along their phylogenetic tree was then tracked (Fig. S8A and B). Compared to normal-like AT2, AT1, and BASC cells, cancer cells were significantly more likely to show increased lineage confusion than their progenitors (Fig. S8C). Furthermore, metastatic cells exhibited increased lineage confusion compared to EMT cells. In metastatic Met2 cells, more than half of the cells showed increased lineage confusion. Collectively, these data suggest that trans-differentiated cancer cells in metastases emerge through de-differentiation followed by re-differentiation.

Early-stage cancers are typically dormant, persisting for many years before clinical detection [76], with metastatic cancer developing many years after initial onset. Several studies have proposed that the evolution of metastatic tumors occurs in a punctuated manner [8,43,77–84]. However, our observations suggest a smooth, continuous transition from the ancestral normal cell to the metastatic cell (Fig. 1J). One possible explanation for this discrepancy is that only a fraction of cancer cells undergo metastatic transformation [84,85]. To test this hypothesis, genetically traced cancer cells [43] were used to compute the total size of progenitor cells (Np, "number of progenitors") for each cell type cluster [86], revealing a population bottleneck between pre-metastatic and metastatic cancer stages: Np gradually increases from AT2 to the $Cldn4^+$ stage, decreases significantly at the EMT stage, and then increases again after metastatic evolution (Fig. 1K). This suggests that only a subset of premetastatic $Cldn4^+$ cells are transformed into EMT cells and beyond.

Analysis of subclusters within the $Cldn4^+$ population revealed that these cells exhibit either lung-like chromatin accessibility profiles or profiles resembling those of alternative-lineage cells (Fig. S9A and C). Interestingly, single-cell replicational age analysis shows that a specific $Cldn4^+$ subcluster (C5), which does not resemble either lung-like or alternative-like profiles, has the lowest replicational age (Fig. S9B and D). Clusters with high replicational age tend to show lung-like or alternative-lineage-like features, suggesting that $Cldn4^+$ cells from the early C5 subcluster could stochastically evolve toward either a lung-like or alternative-like fate (Fig. S9D). Notably, the $Cldn4^+$ subcluster with maximal similarity to EMT and metastatic cells comprises only approximately 8% of all $Cldn4^+$ cells (Fig. S9D). This proportion is close to the ratio of EMT to $Cldn4^+$ progenitor cells inferred from genetic lineage tracing [median Np: 91 (EMT, $N = 13$), 1,553 ($Cldn4^+$, $N = 19$), ratio = 5.86%; Fig. 1K and Table S1], supporting the idea that the population bottleneck from $Cldn4^+$ to EMT cells is driven by the stochastic nature of epigenomic evolution. Collectively, these findings suggest that only the most de-differentiated $Cldn4^+$ cells can undergo EMT, with re-differentiation occurring only after EMT, when global chromatin accessibility is significantly increased [87] (Fig. S10), allowing alternative-lineage TFs to function at full capacity (Fig. 1G and H and Fig. S7).

Trans-differentiation in human lung cancer during metastasis evolution

To assess whether the findings in the mouse model are applicable to human cancer, a re-analysis of an scRNA-seq dataset from patients with LUAD [88] was conducted, covering samples from early to advanced-stage primary tumors and metastases in pleural effusion (PE), lymph nodes (LNs), and brain (Fig. 2A). Similar to the mouse LUAD model, human LUAD cells exhibited a gradual loss of lineage fidelity during their progression (Fig. 2B). Within the primary tumor, a cluster of malignant cells associated with poor clinical outcomes (tS2) [88] demonstrated significantly lower lineage fidelity compared to other malignant cell groups (tS1 and tS3) (Fig. S11A). An alternative validation method, using softmax probability derived from a supervised principal components analysis (PCA)-based cell type annotation approach [89], showed a similar trend of decreasing lineage fidelity with advancing malignancy (Fig. S11B). Concurrently, single-cell stemness progressively increased from normal cells to early and advanced-stage cancer cells, and decreased in metastatic cancer cells, reflecting the trends observed in the mouse model (Fig. 2C and Fig. S11C).

The loss of lineage fidelity could arise from either a reduction in the original cell's transcriptional program, an increase in the transcriptional program of alternative-lineage cells,

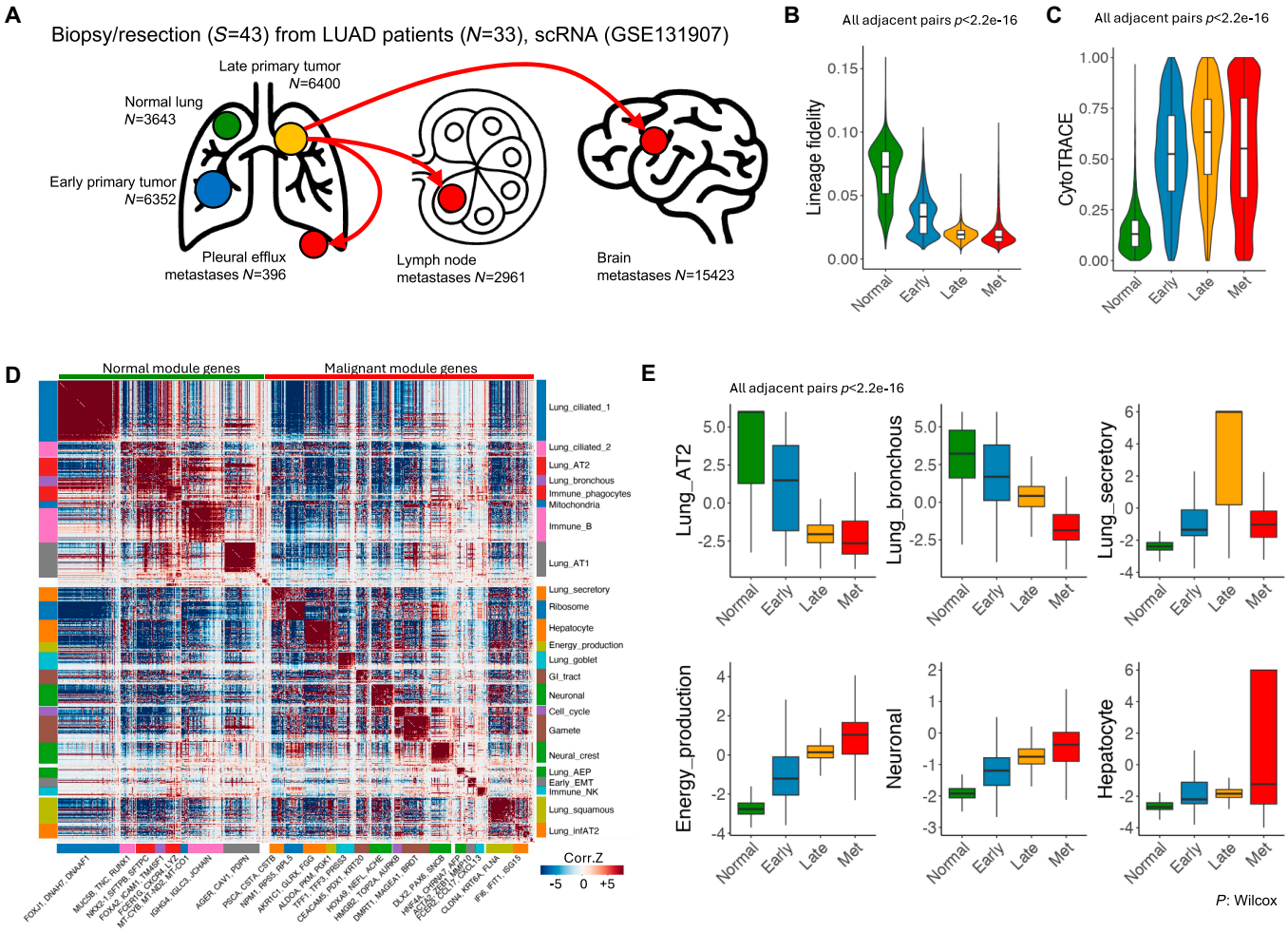


Fig. 2. Trans-differentiation of human lung cancer during evolution toward metastases. (A) Study scheme. Normal lung tissue (Normal), early-stage LUAD resection sample (Early), advanced-stage LUAD biopsy (Late), and metastases (Met) samples underwent scRNA. (B) Lineage fidelity of epithelial cells grouped by their stage. (C) Stemness (CytoTRACE score) of epithelial cells. (D) Cross-correlation revealing de novo identified gene coexpression modules in human LUAD single cells. (E) Expression of identified modules across single cells from different stages. Modules specific for terminally differentiated lung cells (AT1, bronchus), intermediate potential lung cells (secretory), energy production, neuronal-like, and hepatocyte-like cells are shown. *P* value: Wilcoxon test.

or a combination of both. Hotspot [55] gene coexpression modules from these human LUAD cells were segregated into "normal" and "malignant" clusters (Fig. 2D). During LUAD progression, transcription programs associated with the ancestral normal cell (AT2, bronchus) were progressively reduced, while expression of other cell-type-specific transcription programs increased (Fig. 2E and Fig. S12). Thus, at the single-cell level, the reduction in lineage fidelity results from both the loss of original lineage genes and the gain of alternative-lineage gene expression.

A cell with reduced lineage fidelity can adopt either the fate of a promiscuous cell or that of an alternative lineage. Using the maximum likelihood estimation method, it was found that most (97.8% to 98.9%) cancer cells could be unequivocally classified, with >95% probability, as a single terminally differentiated cell type, closely resembling normal cells (98.6% to 99.8%) (Fig. S13A). However, as cancer progresses, the transcriptional profile of cancer cells increasingly deviates from its ancestral cell type (Fig. S13B). In rare instances, cancer cells could be mapped to multiple terminal cell types, with their fate map distribution broadening as cancer evolves, leading to lineage confusion (Fig. S13C and D). These rare multi-mapping cells likely represent an intermediate stage of lineage trans-differentiation.

As trans-differentiation advances, the cancer cell may either adopt the transcriptional profile of an alternative cell within its developmental branch or cross boundaries to resemble cells from other developmental branches, such as cells from different organs. The organ-origin of the best-correlated reference normal cells for these LUAD cells was estimated, and the first hypothesis was rejected: as cancer cells evolve, they gradually transition out of the lung lineage, adopting developmentally distant lineages from other organs (Fig. S14).

Trans-differentiation is a hallmark of human metastatic cancers

Lineage trans-differentiation was identified during the evolution of both mouse and human LUAD toward metastasis, with multiple measures used to distinguish it from other potential confounding events (Fig. S15). However, it remains unclear whether this phenomenon is specific to LUAD or if it can be generalized to other cancer types. To explore this, scRNA-seq

data from primary and metastatic human tumor tissues were analyzed. Primary and metastatic data were obtained from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) for 8 tumor types, including 4 adenocarcinomas [breast cancer (BRCA) [90], colorectal cancer (CRC) [91], prostate adenocarcinoma (PRAD) [92,93], and stomach adenocarcinoma (STAD) [94]], 3 non-adenocarcinomas [clear cell renal carcinoma (ccRCC) [95–97], hepatocarcinoma (HCC) [98], and head-and-neck squamous cancer (HNSC) [99]], and 1 melanoma type [uveal melanoma (UVM) [100]]. In all adenocarcinomas and non-adenocarcinomas, except for HNSCs, metastatic cancer cells exhibited a decrease in lineage fidelity compared to primary cells (Fig. 3A to F). Additionally, the stemness of metastatic cancer cells was reduced relative to primary

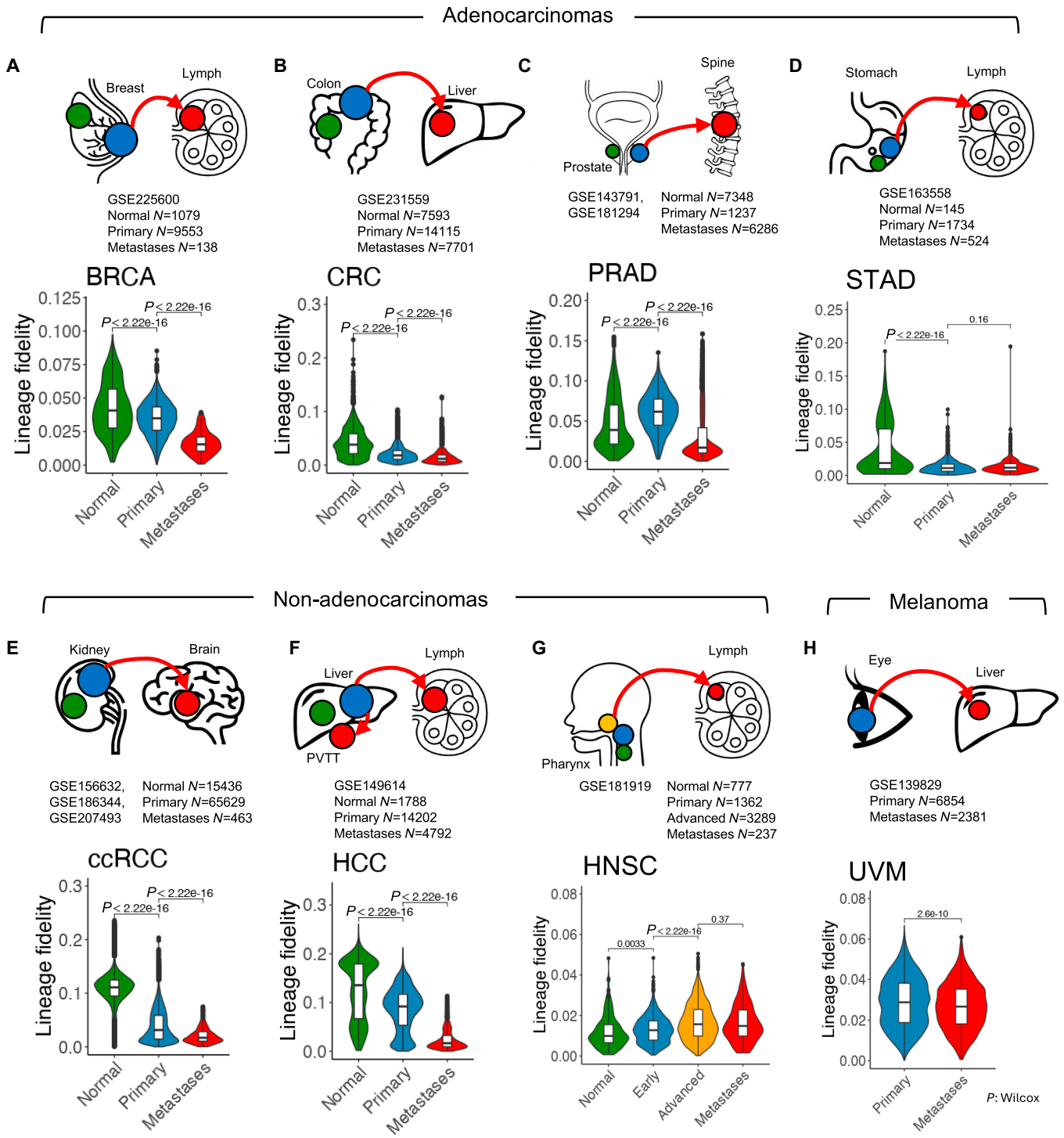


Fig. 3. Trans-differentiation is a hallmark of human metastatic cancers. Lineage fidelity scores of normal, primary cancer, and metastatic cells in (A) breast cancer (BRCA), (B) colorectal cancer (CRC), (C) prostate adenocarcinoma (PRAD), (D) stomach adenocarcinoma (STAD), (E) clear cell renal carcinoma (ccRCC), (F) hepatocarcinoma (HCC), (G) head-and-neck squamous carcinoma (HNSC), and (H) uveal melanoma (UVM). For HNSC, primary cancer samples are staged into early and late stages (copresent with metastases). For UVM, no normal cells were available for analysis. *P* value: Wilcoxon test.

cancer cells (Fig. S16). In HNSC and UVM, although lineage fidelity in metastatic cells did not decrease as markedly as in other tumor types (Fig. 3G and H), the stemness of metastatic cells was still significantly lower (Fig. S16). Therefore, evidence of trans-differentiation in metastatic cells was found across all the tumor types analyzed.

Epigenomic trans-differentiation in human metastatic cancer

In the mouse model, lineage trans-differentiation at the transcriptional level is accompanied by epigenomic reprogramming (Fig. 1G and H). A method was then established to use DNA methylation (DNAm) to detect lineage trans-differentiation in human clinical tumor samples. Inspired by the observation that cancer cells switch to transcriptionally distinct fates resembling cells from alternative organs, non-negative least squares (NNLS) deconvolution was performed on the DNAm profiles of samples against known normal tissue DNAm profiles [101], yielding a vector of "tissue similarity scores" for each sample (Fig. 4A).

This method was tested on 35 primary LUAD resection specimens from an in-house cohort (Challenge-Lung). Based on their DNAm similarity to a panel of normal tissue methylation references, the tumors were segregated into 4 groups: low-grade (LG), mid-grade (MG), high-grade (HG), and trans-differentiated (Trans) (Fig. S17A). Notably, this classification aligned with pathological grade (Fig. S17A and B): The LG group mostly consisted of low-grade G1/2 lepidic and acinar cancers [102], the MG and HG groups contained high-grade G3 cancers, and the Trans group included histologically non-adenocarcinoma cancers, such as small cell lung carcinoma and lung adenosquamous carcinoma, likely transformed from adenocarcinoma precursors. This segregation was largely attributed to the loss of lung-specific DNAm profiles and the acquisition of non-lung epithelial DNAm profiles (Fig. S17C). Importantly, since the DNAm analysis was performed on primary tumor samples, the presence of non-lung DNAm profiles is unlikely to result from contamination by non-lung normal cells, but rather from the transformation of cancer cells into alternative lineages.

From the cancer methylation deconvolution profiles, a DNAm deviation index (DNAmDev index) was derived, defined as the difference between the DNAm weight of non-origin epithelial tissue and the original tissue (Fig. 4A). As expected, the DNAmDev index increased from LG to MG and HG, peaking in the Trans group (Fig. 4B). This finding was validated by analyzing the external TCGA (The Cancer Genome Atlas) LUAD cohort [39]. In the TCGA LUAD cohort, the DNAmDev index correlated with histopathological grade, with more aggressive histopathological types exhibiting higher DNAmDev index values (Fig. S17D). Correlation of the DNAmDev index with RNA expression profiles in the same samples revealed that the DNAmDev index was negatively correlated with lung-specific TFs such as *NKX2-1* and *CEBPA*, and positively correlated with non-lung, alternative-lineage TFs like *SOX9*, *HNF4A*, *CDX1*, and *PDX1* (Fig. 4C). Thus, the DNAmDev index provides a means to detect lineage trans-differentiation in human cancer.

To validate whether trans-differentiation is associated with metastasis in human cancers, the DNAm profile of primary LUAD and brain metastases [103] was decomposed against a normal reference panel, revealing a significant increase in non-lung DNAm signals in the metastases (Fig. 4D). This result was further validated by profiling DNAm in an in-house tumor sample cohort (Challenge-PANCAN, $N = 44$ samples from 24 patients, including 12 with paired primary-metastatic tumor samples). An increase in non-origin alternative tissue DNAm signals was observed in all tested primary-metastasis pairs across 5 tumor types, including LUAD, STAD, PRAD, meningioma (MEN), and pancreatic neuroendocrine tumor (PNET) (Fig. 4E and F). This suggests that trans-differentiated cells are more prevalent in metastases than in primary sites, and this phenomenon is not exclusive to LUAD.

In a patient with prostate cancer, primary adenocarcinoma, neuroendocrine-transformed foci in the prostate, and liver metastatic lesions were sequenced (Fig. 4F). The liver metastatic clone was derived from the neuroendocrine-transformed clone in the prostate, as both lesions shared a similar RB1 mutation. Nonprostate DNAm signals increased from the primary to the premetastatic site, and further increased in the liver metastases. Similarly, using cell replicational age estimation from scATAC data, it was found that late-derived trans-differentiated cancer cell clones dominated the metastases (Fig. S18) in a human CRC PDX model [104]. These findings suggest that epigenomically trans-differentiated cancer clones are responsible for seeding metastasis in various cancer types.

Tumor DNAm profiles in the TCGA cohort ($N = 9,837$ samples) [105] were systematically analyzed using this method (Fig. S19A). The DNAmDev index varied across cancer types (Fig. S19B), with tumor types exhibiting higher DNAmDev indices showing significantly worse overall survival (OS) compared to those with lower DNAmDev values (Fig. S19C).

A previous study used machine learning to develop a DNAm-based stemness scoring algorithm (mDNAsi) for tumors, which measures the similarity between the observed DNAm profile and that of embryonic stem cells [68]. In the TCGA cohort, the DNAmDev index showed inconsistent correlation with mDNAsi (Fig. S20), indicating that the 2 indices represent distinct biological processes: The DNAmDev index likely reflects trans-differentiation, while mDNAsi is more likely associated with a fetal-like state. In the TCGA cohort, regional remission and distal metastases were strongly associated with an increased DNAmDev index (Fig. S21A), but only weakly with mDNAsi (Fig. S21B). In multivariate Cox regression analysis ($N = 8,890$ samples with available survival data), including DNAmDev, mDNAsi, and clinical stage, the DNAmDev index was the only factor contributing to the progression-free interval (PFI) (Fig. S21C), and it independently predicted worsened OS regardless of tumor clinical stage (Fig. S21D). Among various cancer types, including bladder cancer (BLCA), LUAD, kidney clear cell carcinoma (ccRCC), PRAD, and low-grade glioma (LGG), patients with high DNAmDev tumors exhibited significantly worse progression-free survival (Fig. 4G). In summary, these results indicate that lineage trans-differentiation primes metastasis and progression in human cancer.

Genetic gain of oncogenic MAPK signaling associates with lineage trans-differentiation

The association between genetic mutations and the DNAmDev index was systematically evaluated in the Challenge-Lung cohort samples (Fig. S17A and Table S2). Since the sample size is limited, we reasoned that single gene analysis is unlikely to yield significant association. Therefore, we also gathered oncogenic MAPK alterations [germline pathogenic epidermal growth factor receptor (EGFR) mutations, MAPK oncogene amplification on

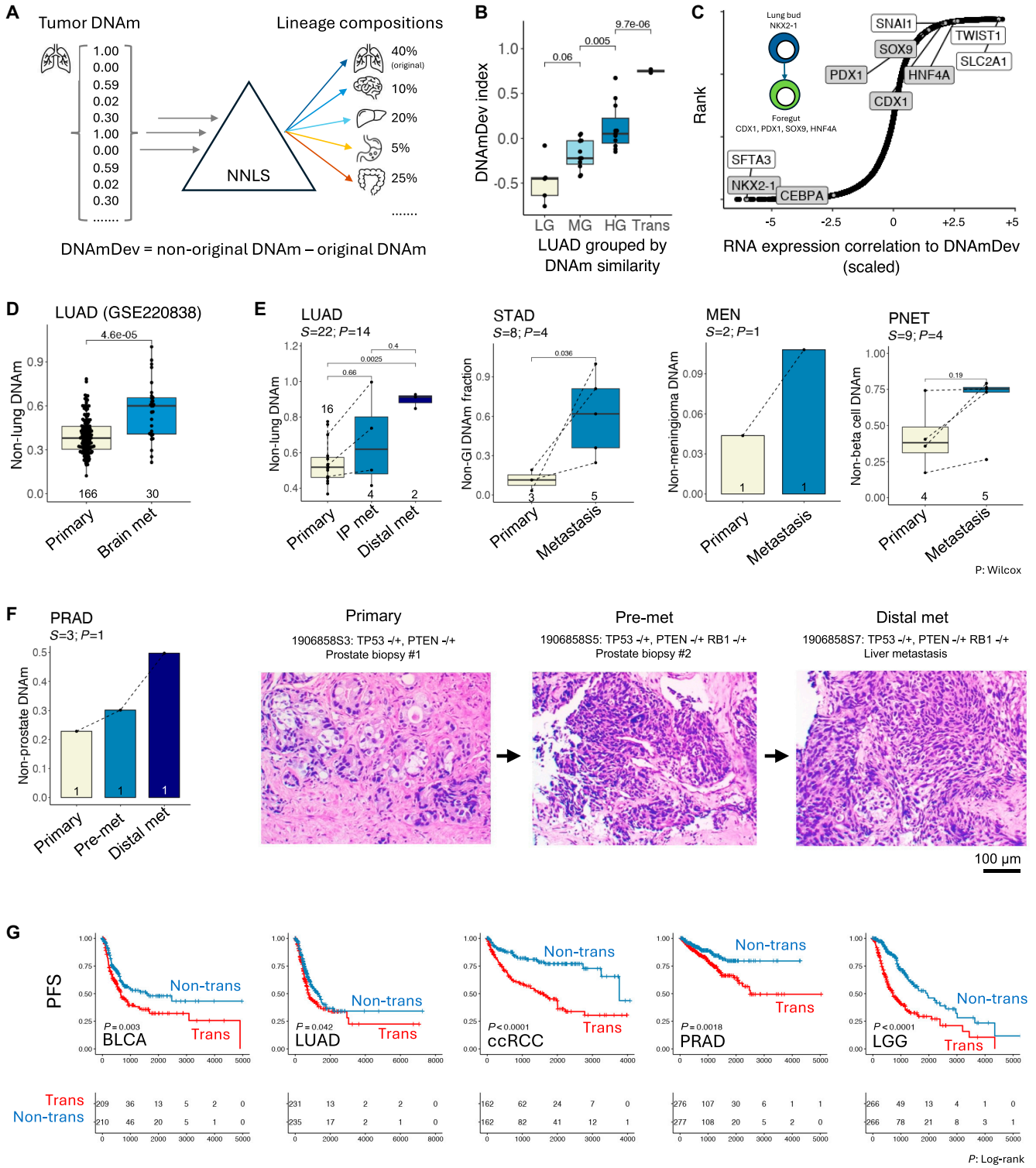


Fig. 4. Lineage trans-differentiation in human cancer. (A) Schematic of the process to compute the DNAm-deviation (DNAmDev) index: DNAm values of lineage-specific methylation sites are decomposed using an NNLS method to infer the lineage compositions (weights) of the tumor. DNAmDev is calculated as the difference between the non-original epithelial tissue weight and the original tissue weight. (B) DNAmDev of the Challenge-Lung cohort samples. Samples were grouped by their DNAm similarity to normal tissue references into LG (low-grade), MG (mid-grade), HG (high-grade), and Trans (transformed) classes. (C) Correlation of DNAmDev index with RNA expression of each gene in the TCGA LUAD cohort. Inset: Schematic showing TFs of lung cells and primitive foregut progenitor cells. *P* value: Wilcoxon test. (D) DNAmDev index of primary LUAD and brain metastases in the external cohort [103]. (E) DNAmDev index of in-house cohort primary cancers and metastases. Lines connect samples from the same patient. S, samples; P, patients; STAD, stomach adenocarcinoma; LUAD, lung adenocarcinoma; PNET, pancreatic neuroendocrine tumor; MEN, meningioma. (F) Left: DNAmDev index of samples from a patient with prostate adenocarcinoma (PRAD). Right: Corresponding histological slides from primary, premetastasis, and liver metastasis sites. Scale bar, 100 μm. (G) Progression-free survival (PFS) of TCGA patients with trans-differentiated tumors (DNAmDev > median value) or non-trans-differentiated tumors (DNAmDev ≤ median value). BLCA, bladder cancer; LUAD, lung adenocarcinoma; KIRC, kidney clear cell carcinoma; PRAD, prostate adenocarcinoma; LGG, low-grade glioma. *P* value: log-rank test (for survival) and Wilcoxon test (for 2-group comparison).

extrachromosomal DNA (ecDNA) [106], and structural variants leading to high-expression MAPK fusions (kinase fusions) (Fig. S1)] or tumor suppressor mutations (including p53, CDKN2A/2B/1A, RB1, KEAP1, and APC) as a group for analysis. Aside from the gain-of-function in MAPK signaling, no other genetic alterations were found to be significantly linked with the DNAmDev index. Particularly, a strong gain in MAPK signaling (MAPK-high) was more prevalent in the HG and Trans groups. MAPK-high tumors exhibited a significantly increased DNAmDev index compared to MAPK-low tumors (Fig. 5A). In contrast, the loss of tumor suppressors was not associated with changes in the DNAmDev index (Fig. 5B).

In the Challenge-Lung cohort, 4 patients carried germline TP53 loss-of-function mutations, and 2 patients had pathogenic

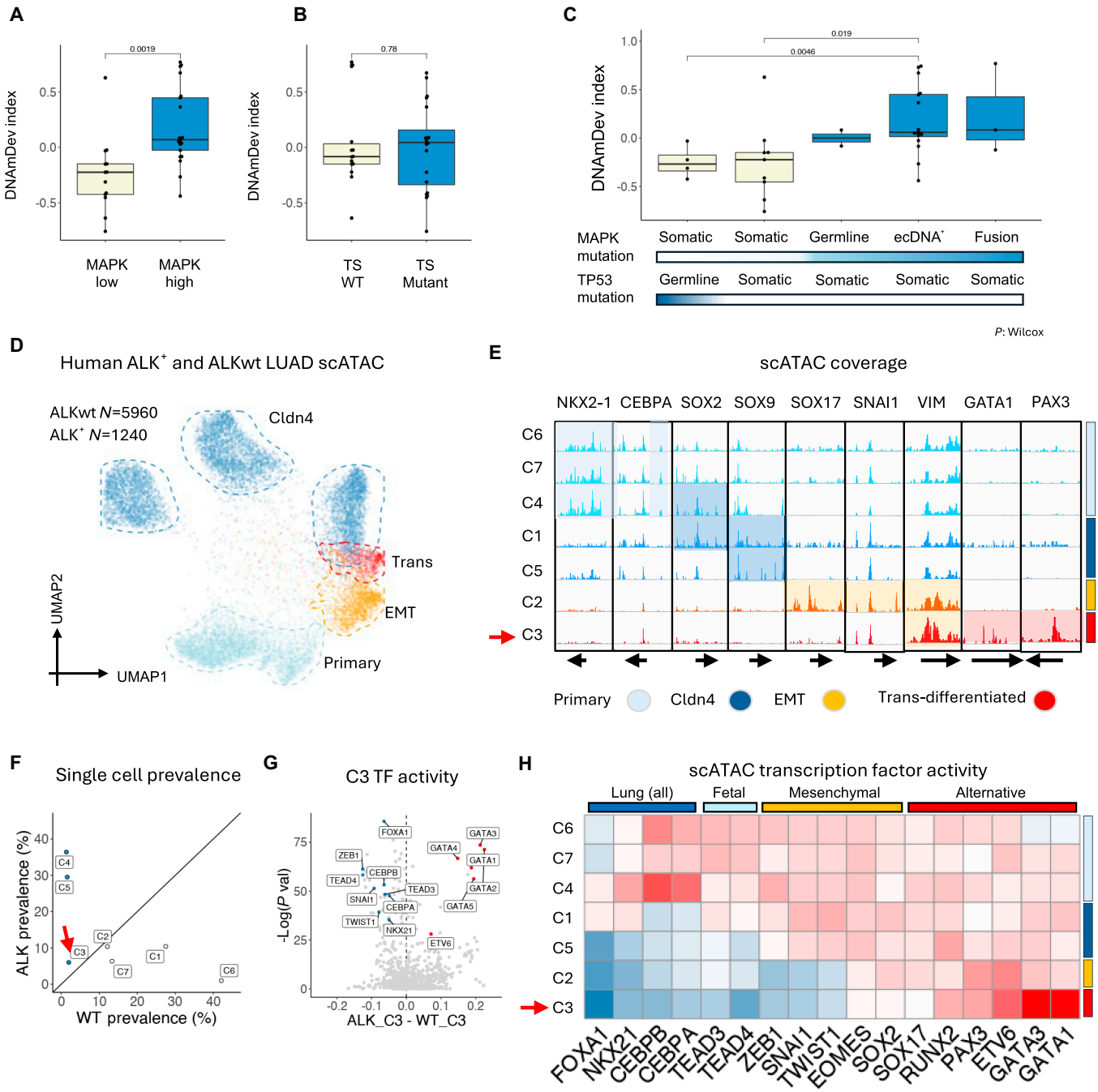


Fig. 5. Genetic gain of MAPK signaling associates with lineage trans-differentiation. (A) DNAmDev index of cancers with high and low MAPK activity. MAPK high: carrying germline pathogenic MAPK mutations, extrachromosomal MAPK amplification, or MAPK fusions. (B) DNAmDev index of tumor suppressor-mutated and wild-type cancers. (C) DNAmDev index of cancer samples segregated by TP53 mutation (somatic or germline) and MAPK mutation [somatic, germline, extrachromosomal DNA (ecDNA$^+$), or fusion via structural variation]. (D) UMAP of human ALK^{wt} and ALK^{mut} scATAC cells. (E) Chromatin accessibility around marker genes for each scATAC cell cluster. Arrow indicates the trans-differentiated cluster. (F) Relative prevalence of each scATAC cell cluster in ALK^{wt} and ALK^{mut} samples. Arrow indicates the trans-differentiated cluster. (G) Differential transcription factor activity between ALK^{wt} and ALK^{mut} scATAC cells of the trans-differentiated cluster. (H) Transcription factor activity profile in each scATAC cell cluster. *P* value: Wilcoxon test.

(oncogenic/activating) germline EGFR mutations. Loss of tumor suppressors prior to somatic MAPK mutations in germline TP53 variant carriers did not increase the DNAmDev index (Fig. 5C). In contrast, germline EGFR mutations led to an increase in the DNAmDev index (Fig. 5C). Additionally, in 3 patients with STAD exhibiting brain metastases from the Challenge-PANCAN cohort, the copy numbers of MAPK oncogene amplification on ecDNA not only increased along the evolutionary trajectory but also correlated with lineage trans-differentiation, as indicated by the DNAmDev index (Fig. S22).

To further validate these findings, scATAC-seq data from EML4-ALK fusion-positive and fusion-negative human LUAD were analyzed (Fig. 5D and Fig. S1) [107]. Trans-differentiated cancer cells, which express alternative-lineage TFs (Fig. 5, E and H), were found to be more enriched in EML4-ALK$^+$ cancers (Fig. 5F). Moreover, within the trans-differentiated cancer subcluster, the activity of alternative-lineage TFs was significantly higher in EML4-ALK$^+$ cells, while lung-specific TF activity was notably reduced (Fig. 5G). Similarly, in the scRNA dataset from the same tumors, EML4-ALK$^+$ cells exhibited decreased expression of fetal-like and EMT-related genes, coupled with an increase in expression of genes from alternative lineages (Fig. S23). These results suggest that in vivo trans-differentiation is linked to the genetic activation of MAPK signaling.

MAPK signaling downstream of RTK promotes de novo DNAm [108–111]. The DNAm surrounding TF-binding sites (TFBS) across the genome was computed for tumors in the Challenge-Lung cohort. MAPK-high tumors exhibited higher DNAm levels around most (88%) TFBS compared to MAPK-low tumors (Fig. S24A and B), indicating that the gain of MAPK signaling induces global de novo DNAm in cancer, as has been previously observed in other tissues. Significantly hypermethylated TFBS in MAPK-high tumors affected NKX2-family, TP53, and EMT-related TFs (Fig. S24C), suggesting that MAPK signaling activation may down-regulate lung-like and EMT gene programs by inhibiting their respective driver TF functions. In contrast, the loss of tumor suppressors did not appear to affect DNAm at TFBS related to trans-differentiation (Fig. S24D).

Our computational analysis identified the alternative-lineage TFs ETV6, GATA1, GATA3, SOX2, and SOX17 as up-regulated in the trans-differentiated cells in MAPK-activated lung cancer (Fig. 5H). To functionally validate whether these alternative-lineage TFs drive metastasis-associated phenotype in these cells, we performed small interfering RNA (siRNA)-mediated knockdown of each TF in the human lung cancer cell lines PC9 and A549. Transwell migration and invasion assays revealed that silencing these trans-differentiation TFs significantly impaired the migratory and invasive capacities of both cell lines compared with control siRNA-treated cells (Fig. S25).

These results provide experimental evidence that lineage trans-differentiation is functionally required for the invasive phenotypes associated with metastasis. This complements our computational findings and strongly supports the conclusion that lineage trans-differentiation is a driver, not just a passenger, of metastatic progression.

MAPK inactivation reverses human cancer lineage trans-differentiation

The functional requirement of MAPK signaling for post-EMT lineage trans-differentiation was tested by analyzing gene expression and DNAm changes following ALK pharmacological inhibition in an EML4-ALK$^+$ human cancer cell line in vitro (Fig. 6A and B). ALK inhibition in these cells led to a switch from an epithelial-like phenotype to an EMT-like phenotype (Fig. 6C), a reduction in alternative-lineage DNAm profiles (Fig. 6D), decreased expression of alternative-lineage genes, and the re-induction of fetal-like and EMT-related genes (Fig. 6E).

To determine whether this phenomenon applies to other MAPK drivers, 7 bulk RNA-seq datasets of EGFR- and RET-driven lung cancer cells treated with tyrosine kinase inhibitors (TKIs) were systematically evaluated (Fig. S26A). Gene set enrichment analysis (GSEA) revealed that acute TKI application induces the activation of myogenic and EMT gene sets, with this activation maintained in TKI-resistant cell lines selected by chronic TKI exposure (Fig. S26B). In fetal lung-tip organoids, treatment with extracellular signal-regulated kinase (ERK) or MAPK kinase (MEK) inhibitors, but not AKT inhibitors, similarly induced myogenic and EMT gene programs (Fig. S26B), consistent with the understanding that oncogenic RTK signaling primarily activates ERK/MEK rather than AKT [36]. Collectively, these findings suggest that oncogenic RTK signaling through ERK or MEK is necessary for the lineage trans-differentiation of EMT cancer cells.

This result was further validated using a human clinical dataset of LUAD scRNA data from pre-TKI (naive, TN) and post-TKI (response, PR; and TKI-resistant progression, PD) biopsies [112] (Fig. 6F). In these cells, TKI treatment was associated with increased lineage fidelity, while disease progression was linked to a reduction in lineage fidelity (Fig. 6G). TKI response resulted in the suppression of alternative-lineage-specific transcription programs and the activation of fetal and EMT-related programs (Fig. 6H). These events were reversed in the TKI-resistant progression samples. Together, these data demonstrate that MAPK signaling drives post-EMT lineage trans-differentiation.

Discussion

Through integrated scRNA and scATAC analyses of the mouse LUAD model, we characterized the evolutionary trajectory from primary cancer to metastatic-ready clone, and found that cancer cells first undergo de-differentiation during early oncogenesis before EMT, by down-regulating cell-of-origin-specific transcription factors and their downstream targets. These pre-metastatic cells then re-differentiate to adopt alternative-lineage transcriptional programs, by mis-expressing and activating alternative-lineage transcription factors. These results are further supported by scRNA, scATAC, and DNAm analyses across multiple human cancer types, including primary samples, organoid systems, and xenograft models. Thus, fetal-like primitive progenitor state primarily functions as a gateway to trans-differentiated cancer. De-differentiated EMT cells, characterized by a fetal signature, act as precursors for metastases. These cells subsequently re-differentiate, adopting alternative-lineage gene expression profiles, thereby forming trans-differentiated metastatic clones (Fig. 7).

Lineage trans-differentiation has been observed in various cancer types, in both mouse models and human clinical samples. Clinically, lineage trans-differentiation is strongly associated with histopathological grade, metastatic potential, and poor clinical prognosis. Although the presence of a fetal-like

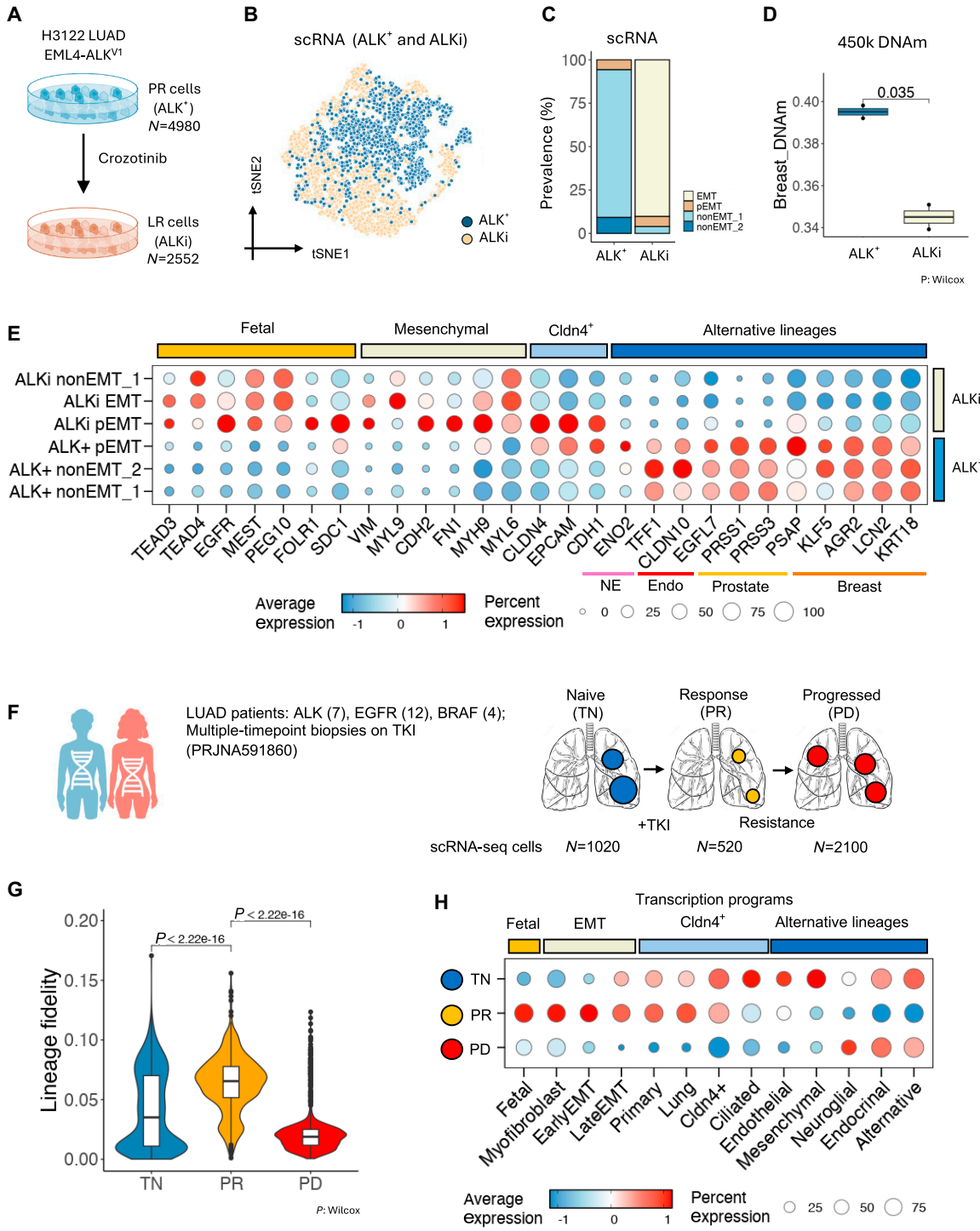


Fig. 6. MAPK inactivation reverses human cancer lineage trans-differentiation. (A) Schematic of the experiment to determine gene expression profile changes after ALK inhibitor application in the EML4-ALK$^+$ human LUAD cell line H3122. (B) UMAP of ALKi H3122 cells and ALK$^+$ treatment-naive H3122 cells. (C) Prevalence of each cell type in ALKi and ALK$^+$ cells. (D) Similarity to breast DNAm profiles of ALKi and ALK$^+$ cells. (E) Key gene expression profile in ALKi and ALK$^+$ cells. (F) Schematic of the experiment to determine gene expression profile changes after TKI application in human cancers in vivo. (G) Lineage fidelity score of treatment-naive (TN), response stage (PR), and progression stage (PD) biopsy single cells. (H) Gene program expression profile in TN, PR, and PD biopsy single cells. *P* value: Wilcoxon test.

state predicts future trans-differentiation, it represents a transient intermediate stage during metastatic cancer evolution and may be absent in advanced-stage endpoint samples, such as tumor biopsies. Therefore, using the fetal-like state alone for clinical prediction may lead to false negatives. In contrast, the presence of a trans-differentiated state in endpoint tumors indicates that a fetal-like state likely existed during the tumor's evolutionary history. This temporal relationship may explain why lineage trans-differentiation outperforms the fetal-like state in predicting clinical outcomes.

While fetal-like states are shared among tumors, best characterized by up-regulation of the EMT-associated and myogenic

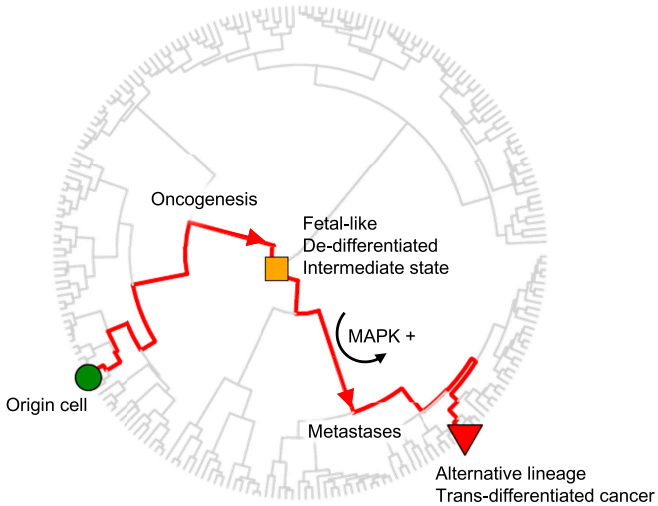


Fig. 7. Summary of findings: Evolution of cancer metastases through trans-differentiation. The minimal transition trajectory (red) between the origin lung AT2 cell and an observed LUAD single cell, which has adopted an alternative-lineage normal cell, is shown on a reference normal cell phylogenetic tree (plotted in a circular structure, with the most ancestral/developmentally immature cell at the center and terminally differentiated cells at the outer edge). The metastatic cancer cell first evolves from the origin normal cell through a series of de-differentiation steps toward a fetal-like state and then trans-differentiates to adopt an alternative-lineage fate, enabling the formation of metastases. MAPK-driven cancers rely on MAPK signaling for trans-differentiation.

transcriptional program and global activation of chromatin, the trans-differentiated states are more diverse. On the single-cell level, each trans-differentiated cancer cell adopts a distinct state (Fig. S13B). Occasionally, we observe intermediate-state cells showing transcriptional similarities to more than one cell type, and the developmental distance between the target cell types increases as tumor stage progresses (Fig. S13C and D). These observations suggest that the "off-target" lineage signatures may reflect a generalized state of high cellular plasticity rather than a shared common signature across metastatic cancers.

While the activation of the MAPK pathway has been extensively studied in early-stage oncogenesis [54,113–115], its role in promoting metastasis remains less clear. Our results indicate that lineage trans-differentiation is associated with the genetic gain of MAPK signaling and can be reversed by MAPK inhibition. Specifically, in human LUAD, alternative-lineage transcription factors are more active in EML4-ALK$^+$ cells than in wild-type cells (Fig. 5D to H), suggesting that elevated oncogenic MAPK signaling is linked to acquisition of the trans-differentiated program. Consistently, pharmacological inhibition of ALK in an EML4-ALK$^+$ human cancer cell line suppresses alternative-lineage genes and restores fetal/EMT-associated programs at both the transcriptomic and DNA methylome levels (Fig. 6A to E). Extending beyond ALK-driven tumors, we re-analyzed additional EGFR- and RET-driven cancer datasets and observed that MAPK-targeting TKI treatment reproducibly shifts tumor cells toward a less trans-differentiated, more lineage-faithful state (Fig. S26). Crucially, this pattern is also evident in paired patient biopsies: TKI exposure in vivo is associated with increased lineage fidelity, whereas subsequent resistance and progression coincide with its loss (Fig. 6F to H). Although direct causal evidence in human tumors is still lacking, these convergent observations strongly suggest that oncogenic MAPK signaling is necessary to maintain, and may drive or stabilize, lineage trans-differentiation. Interventions targeting this molecular pathway could therefore potentially prevent cancer metastasis. In clinical trials, postoperative adjuvant TKI treatment has proven beneficial in preventing metastasis in driver MAPK mutation-positive LUAD, BRCA, and melanoma individuals, as demonstrated in the ADAURA, ALINA, COMBI-AD, NEfERT-T, and NALA trials [116–121]. Our findings not only help explain these results but also suggest future directions for developing and improving clinical interventions to combat metastasis.

The adoption of an alternative-lineage fate through epigenomic reprogramming may provide a selective advantage during metastatic seeding. Consistent with the population bottleneck we observed at the early metastatic stage (EMT/Met1) (Fig. 1K), trans-differentiation could expand ITH and generate phenotypic variants that better survive immune attack and treatment, while alternative-lineage transcriptional profiles may enhance adaptation to distal metastatic niches. In this context, trans-differentiated cells might become locked in a partially reprogrammed, intermediate progenitor-like epigenomic state, enabling both increased motility and aggressive proliferation, which together could underlie their high metastatic potential. We additionally considered whether such "off-target" lineage programs dictate organ tropism. Although DNAm deconvolution revealed distinct non-self lineage methylation patterns across different cancer types (Fig. S19), we did not establish a clear causal link between these patterns and metastatic site preference. For instance, glioblastoma (GBM) and LGG tumors sometimes exhibit pancreatic-, breast-, or prostate-like methylation patterns, which do not correspond to their typical sites of metastasis. However, this does not exclude the possibility that metastatic niche preference is determined by the trans-differentiation state at the transcriptional level. In support of this hypothesis, while this paper is under review, another group reported that pancreatic adenocarcinoma cells adopting lung- or liver-parenchymal-like transcriptional features preferentially form metastases in the corresponding organs [122]. Together with our prior finding that trans-differentiation can facilitate immune evasion by down-regulating immunogenic molecules and up-regulating immunosuppressive programs [123], these observations support a model in which trans-differentiation promotes metastatic fitness through niche adaptation, immune escape, and a hybrid motility-proliferation phenotype. Future in vivo perturbation and lineage-tracing experiments will be required to test these hypotheses directly.

The current study has some limitations. Our data are primarily based on re-analysis of published datasets and lack direct manipulation experiments demonstrating that inducing trans-differentiation converts nonmetastatic cancer into metastatic cancer, which should be addressed in future studies. The re-analysis of a large number of publicly available datasets from different laboratories and technologies may introduce inconsistencies and potential false negatives. While our DNAm analysis provides an important layer of epigenomic evidence linking MAPK signaling to cell fate regulation in metastasis, histone modifications and higher-order chromatin structure were not analyzed, which are beyond the scope of the current work. Moreover, although our data demonstrate that MAPK signaling is necessary to maintain the trans-differentiated state, future genetic perturbation and lineage-tracing experiments, such as activating and inactivating MAPK to change metastatic potential of cancer cells, are needed to confirm whether MAPK signaling is sufficient to drive this process. As our study primarily

relies on bioinformatic analysis, the results have not been validated through reverse genetics in animal models. Finally, this study did not investigate whether trans-differentiation enables metastatic cancer cells to adapt to a non-native metastatic microenvironment, evade immune surveillance, or enhance their mobility in vivo. These questions should be addressed in future experimental studies.

Materials and Methods

Human biospecimens

Clinical information and samples (including blood and surgical tissue samples) were collected from the multi-center, prospective, observational cohort for cancer, known as the Cancer HALLmark Epigenetics aNd Genetics (CHALLENGE). The study included 35 samples ($n = 35$, from 35 donors) from the lung cancer subcohort (Challenge-Lung) and 44 samples (from 24 patients) from the pan-cancer cohort (Challenge-PANCAN). These samples were collected from Wuhan and Beijing, China, and were processed and preserved by the Hubei Biobank, an official member of the International Society for Biological and Environmental Repositories. The study received approval from both the Institutional Ethics Committee of Zhongnan Hospital of Wuhan University (approval numbers: 2017038-1 and 2024061K) and the China Human Genetic Resources Management Office, Ministry of Science and Technology of China (approval number: 2022BC0128). Written informed consent was obtained from all participants involved in the study.

Public sequencing datasets

This study incorporated a large number of publicly available sequencing datasets from previous research. A comprehensive list of these datasets can be found in Table S3.

Histopathology of clinical cancer samples

For hematoxylin and eosin (H&E) staining, formalin-fixed, paraffin-embedded (FFPE) tissue sections were deparaffinized in xylene and rehydrated through a graded ethanol series. Nuclei were stained with hematoxylin (5 min), followed by differentiation in 1% acid ethanol (30 s) and bluing in running tap water (5 min). Cytoplasmic counterstaining was carried out with eosin (2 min). Sections were then dehydrated through graded ethanol, cleared in xylene, and mounted with neutral resin. All procedures were performed at room temperature, and stained slides were evaluated using bright-field microscopy.

Cell culture

Human lung cancer cell lines PC9 and A549 were kindly provided by Cell Bank, Chinese Academy of Sciences (Shanghai, China). PC9 and A549 cells were cultured in RPMI 1640 supplemented with 10% fetal bovine serum (FBS) and 1% penicillin–streptomycin (10,000 U/ml). Each cell line was authenticated and tested negative for mycoplasma contamination.

Cell transfection, RNA extraction, and qRT-PCR

siRNAs were purchased from GenePharma (Shanghai, China). The target sequences of the siRNAs used in this study are listed in Table S4. Cells were transfected with siRNAs using Lipofectamine 3000 (Invitrogen, L3000150) according to the manufacturer's instruction. Total RNA was isolated with the HiPure Total RNA Mini Kit (Magen, R4111-03). cDNA was synthesized using the ReverTra Ace qPCR RT Kit (TOYOBO, FSQ-101). Quantitative reverse transcription polymerase chain reaction (qRT-PCR) was performed on StepOnePlus real-time PCR system (Thermo Fisher, USA) with SYBR Green SuperMix (Bio-Rad, 1725125). Primer sequences are provided in Table S4.

Cell migration and invasion assays

Cell migration and invasion were evaluated using 24-well Transwell inserts (Falcon, 353097). For the migration assay, 5×10^4 PC9 cells or 6×10^4 A549 cells were resuspended in 200 µl of serum-free RPMI 1640 media and seeded into the upper chambers. For the invasion assay, 1×10^5 PC9 cells or 2.1×10^5 A549 cells were resuspended in 200 µl of serum-free RPMI 1640 media and added to the upper chambers that had been precoated with Matrigel (Yeasen, 40183ES08). In all experiments, 600 µl of RPMI 1640 supplemented with 20% FBS was added to the lower chambers as a chemoattractant. The Transwell chambers were incubated for 24 h to allow cells to migrate or invade through the porous membrane. After incubation, nonmigrated or non-invaded cells remaining on the upper surface of the membranes were carefully removed with a cotton swab. Cells that had traversed the membranes and adhered to the lower surface were fixed in 4% paraformaldehyde and stained with crystal violet.

scRNA data preprocessing, stemness inference, and cross-dataset integration

The analyzed scRNA data were presented as gene-x-cell count matrices. Datasets were processed individually using Seurat (4.1.3) for reading, normalization, variable gene selection, PCA, neighbor finding via K-nearest neighbors (KNN), clustering, and uniform manifold approximation and projection (UMAP). Default parameters were used for all preprocessing. The top 30 PCA vectors were applied to all datasets. Differentially expressed gene (DEG) analysis based on single-cell clusters was performed using the presto::wilcoxauc (1.0.0) package. For each cluster, the top 20 to 30 DEGs ranked by area under the curve (AUC) were used for manual annotation of cell types with assistance from the DeepSeek API (https://platform.deepseek.com/). Epithelial cells were selected and cleaned up from non-epithelial stromal and hematopoietic lineage cells by cluster annotation. Since estimation of stemness can be influenced by batch effects, CytoTRACE (0.3.3) analysis was performed individually for epithelial cells from each dataset using default parameters.

For cross-dataset integration, epithelial cells from each dataset were combined using Harmony with default parameters. For both mouse normal and cancerous lung cells, standards set by LungMAP [49] and Marjanovic et al. [41] were adhered to. Rather than breaking down the analysis to a very fine granularity for cancerous cells, they were broadly classified into primary cancer, Cldn4$^+$, EMT, and metastasis-associated (Met) cells. The maximum number of datasets contained cells from all of these clusters.

Inferring the likely cell type of last common ancestor and estimating progenitor population size in scRNA lineage-tracing data

The occurrence timing of single-cell clusters (AT2, BASC, Primary cancer, Cldn4$^+$, EMT, Met) was determined by pseudotime analysis using Slingshot (2.4.0) within the integrated

dataset. The phylogenetic tree derived from the scRNA lineage-tracing data [43] was provided by the authors on Zenodo (see Data Availability statement). Single-cell type annotations were used as leaf labels. The tree was recursively backtracked from each leaf, and at each node, the node was annotated by the most ancestral (early-emerged) label among its leaves. This process was completed across the tree to determine the ancestor cell type for each observed single cell (leaf).

To estimate the total progenitor population from a given cell type within these samples, each sample was processed individually using TarCA (TarCA.beta, 0.7.0) following the authors' guidelines. The Np value estimated by TarCA was used to represent the progenitor size (Table S1).

Transcription module extraction from scRNA (Hotspot)

The scRNA-seq count matrix was exported from the integrated Seurat object and saved as h5ad using MuDataSeurat (0.0.0.9000) in R (4.1.3), then read in Python (3.9.7) with Scanpy (1.10.1). The data were normalized and log1p-transformed in Scanpy using default parameters. Genes with mean expression within ±5 SDs and dispersions ≥0.5 were selected. The top 3,000 variable genes from this selected set were used and combined with known markers (e.g., *Nkx2-1*, *Cebpa*, *Cebpe*, *Cebpb*, *Foxa1*, *Foxa2*, *Sox17*, *Sftpb*, *Sftpc*, *Sftpa1*, *Scgb1a1*, *Scgb3a2*, *Scgb3a1*, *Clu*, *Hopx*, *Pdpn*, *Ager*, *Cyp2f2*, *Sox2*, *Cldn4*, *Apoe*, *Cd24a*, *Gdf15*, *Tm4sf1*, *Slc4a11*, *Onecut2*, *Vim*, *Sox9*, *Lgals1*, *Zeb1*, *Zeb2*, *Snai1*, *Snai2*, *Twist1*, *Twist2*, *Id3*, *Gpx3*, *Bmp7*, *Runx1*, *Runx2*, *Gata3*, *Gata1*, *Gata5*, *Hnf4a*, *Pdx1*, *Cdx2*, *Osr1*, *Prrx1*, *Meis1*, *Meis2*, *Myc,* and their human orthologs). Hotspot (1.1.1) was used to run the analysis with the parameters "n_neighbors = 30" for KNN graph creation, "FDR < 0.05" for marker gene selection, and "min_gene_threshold = 30, core_only = False, fdr_threshold = 0.05" for module creation. Module expression scores for each single cell were generated automatically by Hotspot. Manual annotation of each module was done with assistance from the DeepSeek API.

Inferring scRNA cell lineage fidelity using correlation

Both the raw scRNA count matrix ("observation") and the respective reference normal cell matrix ("reference") (mouse: Mouse Cell Atlas; human: Human Protein Atlas single-cell RNA expression) were depth-normalized to a total of 1e+5 and log1p-transformed. Pearson correlation was performed using WGCNA::cor (1.71) between the observation and reference matrices. For each single cell, the top 10 correlation coefficients were extracted, and lineage fidelity was estimated as follows: stdev(cor_coef)/mean(cor_coef).

Single-cell annotation and lineage fidelity inference using Azimuth

The pan-human Azimuth API (0.1.0) with the DISCO database [124] was employed to annotate the human LUAD scRNA dataset, using default parameters. After annotation, the final_level_softmax_prob parameter for each cell was extracted as the confidence level of the prediction.

Single-cell annotation with probability estimation using maximal likelihood estimation

To calculate lineage fidelity, the observation-x-reference matrix correlation was performed, as described in the Inferring scRNA cell lineage fidelity using correlation section. For each single cell in the observation data, reference cells with the top 10% correlation score were selected as candidate reference hits. The depth-normalized read count of each gene from these candidate reference hits was used to construct the prior probability matrix (mu_matrix). Gene-specific read depth dispersions were estimated from this prior probability matrix, and the size parameter (size_matrix) was calculated as 1/dispersion. The raw read count of each gene for the observation single cell was treated as the observation vector, with the distribution density following a negative binomial distribution. In a vectorized form, the likelihood estimation was performed as follows: logP_density_mat <- dnbinom(counts_matrix, size = size_matrix, mu = mu_matrix, log = T) and log_likelihoods <- colSums(logP_density_mat).

This calculation provided log-likelihoods indicating how similar the observation single cell is to each of the candidate reference cells. These were then converted into probabilities, assuming that the single cell must match one of these candidate references. The top N ($N = 10$ in this study) reference cells that accumulated to a probability greater than 95% were selected as the inferred matches for the observed single cell. The top reference cell with the maximal probability was considered the best match to the observed single cell.

scRNA lineage fidelity inference from maximal likelihood estimation results

To calculate lineage fidelity, a hierarchical clustering of the reference cell gene expression (log1p-transformed) was constructed and transformed into a phylogenetic tree based on gene expression similarity. The total travel distance on this tree between all pairs of matched reference cells for a given observation single cell was calculated as the lineage fidelity index.

scRNA transformation trajectory from maximal likelihood estimation results

For LUAD, AT2 cells were always assumed to be the ancestor for any cancerous LUAD cells. To plot the minimal transformation trajectory between an observed single cell and its putative ancestor, the expression similarity tree derived from scRNA lineage fidelity inference from maximal likelihood estimation results was used. The most recent common ancestor (MRCA) of the ancestor leaf and the best match cell leaf [as inferred by maximal likelihood estimation (MLE)] was estimated using the ape package (5.7.1). The minimal path joining the ancestor leaf, the MRCA node, and the best match leaf was then drawn. The results were visualized using the ggtree package (3.4.0).

scATAC data analysis

The scATAC fragment files were downloaded from each public dataset and processed into a unified dataset using ArchR (1.0.1), with the exception of the in vitro lung cancer organoid data, which were processed separately. ArchR analysis was performed using default parameters. After integration, cell types were annotated based on promoter region accessibility of manually curated marker genes. For $Cldn4^+$ cell analysis, $Cldn4^+$ cells were extracted and fine-level subclustering was performed. Single-cell replicational age estimation and phylogenetic tree construction based on this replicational age were carried out using EpiTrace (0.0.1.3). Cell evolution pseudotime was inferred in ArchR using Monocle. Global accessibility distributions between peak and nonpeak regions were derived

from the fragment-in-peak (FRIP) parameter in ArchR. TF activity inference was conducted with ArchR using the cisbp database.

DNAm sequencing and microarray

DNAm sequencing and data processing for in-house samples were performed on FFPE tumor samples, following the protocol described by Xiao et al. [125]. DNA was extracted from the samples, bisulfite-converted, subjected to poly-adenylation and 3′ adaptor ligation, and then linearly amplified from the 3′ adaptor before ligating to a 5′ adaptor. The PCR-amplified, adaptor-ligated product underwent biotinylated bait capture before sequencing on an Illumina Novaseq platform. The sequencing data were trimmed using fastp (0.19.4) and then mapped to the GRCh37+decoy genome using bwa-meth (0.2.0). Per-CpG DNAm frequency (beta) was calculated with PileOMeth (0.1.13-3). Publicly available DNAm data from the Illumina methylation array were downloaded as IDAT files and processed using minfi (1.42.0) to generate beta matrices. TFBS DNAm analysis was performed by predicting TFBS with ArchR and cisbp, and all reads covering these TFBS were grouped to estimate the average methylation level at each site.

DNAm-based tissue lineage deconvolution and estimating DNAmDev

NNLS regression was applied to the observation DNAm levels (beta) against the MethAtlas reference panel, specifically for the corresponding differentially methylated loci (DML), using nnls::nnls (1.4), as described in the original paper [126]. The accuracy of the algorithm was first calibrated using publicly available, well-characterized normal control sample data. Including the "normal bladder" reference sample in the reference panel led to erroneous predictions, classifying many nonbladder normal tissue samples as "bladder". Therefore, bladder samples were removed from the reference panel due to their disproportionate influence on NNLS analysis. The final regression result was not normalized, meaning that the sum of the weights could be less than or equal to 1. The "DNAm deviation" (DNAmDev) index is defined as follows: sum(weight(other_epithelial_tissues)) - weight(cell_of_origin)).

Bulk RNA sequencing

The RNA-seq data matrix (counts) was either obtained from in-house samples or downloaded and analyzed using DESeq2. GSEA scores for each cluster were derived using the mSigdbR package's "Hallmark" gene set.

Somatic mutation sequencing

Somatic mutation sequencing and data processing for in-house samples were conducted on FFPE tumor samples, following the protocols outlined by Xiao et al. [123] and Xue et al. [127]. Only variants clearly annotated as oncogenic mutations and meeting the following criteria were retained: >0.5% variant allele frequency for short nucleotide variants, >3× copy number for amplifications, ≤1× copy number for deletions, or classical kinase fusions with >1% variant allele frequency. ecDNA amplification was called using regions outside of breakage–fusion–bridge cycle (BFB) area or chromothripsis arms, shorter than 10 million base pairs, harboring oncogenes, and with copy numbers >5×.

Statistics

Multivariate Cox models were employed to analyze OS and PFI. Log-rank tests were used for Kaplan–Meier survival analysis, and 2-sided Wilcoxon tests were used for 2-group comparisons. For box-and-whisker plots of grouped data points, the upper and lower bounds of the boxes represent the 25th and 75th percentiles, respectively, while the median is shown as a horizontal line within the box. The whiskers extend to the minimum and maximum values, defined as the farthest data points within 1.5 times the interquartile range (IQR) from the box bounds. Violin plots represent the empirically estimated density distribution of the data. For enrichment analysis results from Fisher's exact test, the odds ratio is represented as the median point in the graph, with its 95% confidence interval shown as error bars.

Ethical Approval

This study was conducted in accordance with the Declaration of Helsinki and received approval from the Institutional Ethics Committee of Zhongnan Hospital of Wuhan University (2017038-1 and 2024061K). Bioinformatics data involving human subjects were sourced from publicly available databases containing anonymized patient information.

Acknowledgments

Funding: This study was supported by funds from the Chinese Institutes for Medical Research, Beijing (CIMR) to Y.Z., the National Natural Science Foundation of China (82573858 and 82372654), the Fundamental Research Funds for the Central Universities (2042022dx0003), and the Research Fund of Zhongnan Hospital of Wuhan University (RCKY20150102, RLYC2024001001, SWYBK06, and CXPY202555). The funders had no role in the study design, data collection and analysis, decision to publish, or manuscript preparation.

Author contributions: Conceptualization: Y.Z. Algorithm development: W.J. and Y.Z. Biological specimen collection: W.J., K.Q., and L.J. Data collection: W.J. and F.C. Sequencing experiment: W.J. Bioinformatic analysis: Y.X., W.J., and Y.Z. Biological experiment and clinical analysis: Y.X., W.J., K.Q., L.J., and Y.Z. Writing the manuscript: Y.X., L.J., and Y.Z. Funding acquisition: Y.X., L.J., and Y.Z. Supervision of the project: Y.Z. All authors read and approved the final manuscript.

Competing interests: The authors declare that they have no competing interests.

Data Availability

The software packages and their versions used in the study are listed in Table S5. The scLineageFidelity and LineageMLE functions are implemented in R and provided on the GitHub repository (https://github.com/MagpiePKU/scLineageFidelity) together with a comprehensive usage instruction README file. De novo inferred and manually curated gene sets for annotating and scoring single-cell data are provided in Table S6. Publicly available data used in this study can be downloaded from Zenodo (https://zenodo.org/records/5847462, https://zenodo.org/records/7713052), the TCGA official website (https://gdc.cancer.gov), the NCBI website (http://www.ncbi.nlm.nih.gov/), and the MCA website (https://bis.zju.edu.cn/MCA/). In-house generated methylation sequencing data from the Challenge-PANCAN and Challenge-Lung cohorts, used in this

study, have been uploaded to the OMIX database at CNGB, China (https://ngdc.cncb.ac.cn/omix/) with accession number OMIX010221 and will be made publicly available upon publication. Due to local legal requirements, the in-house generated human dataset will only be available in processed data format and requires a case-by-case application through the China Human Genetic Resources Management Office. The remaining data are available within the article and in the Supplementary Materials.

Supplementary Materials

Figs. S1 to S26
Tables S1 to S6

References

1. Gerstberger S, Jiang Q, Ganesh K. Metastasis. *Cell*. 2023;186(8):1564–1579.
2. Priestley P, Baber J, Lolkema MP, Steeghs N, de Bruijn E, Shale C, Duyvesteyn K, Haidari S, van Hoeck A, Onstenk W, et al. Pan-cancer whole-genome analyses of metastatic solid tumours. *Nature*. 2019;575(7781):210–216.
3. Yamamoto N, Yang M, Jiang P, Xu M, Tsuchiya H, Tomita K, Moossa AR, Hoffman RM. Determination of clonality of metastasis by cell-specific color-coded fluorescent-protein imaging. *Cancer Res*. 2003;63(22): 7785–7790.
4. Fidler IJ, Talmadge JE. Evidence that intravenously derived murine pulmonary melanoma metastases can originate from the expansion of a single tumor cell. *Cancer Res*. 1986;46(10):5167–5171.
5. Fidler IJ, Kripke ML. Metastasis results from preexisting variant cells within a malignant tumor. *Science*. 1977;197(4306):893–895.
6. Navin N, Kendall J, Troge J, Andrews P, Rodgers L, McIndoo J, Cook K, Stepansky A, Levy D, Esposito D, et al. Tumour evolution inferred by single-cell sequencing. *Nature*. 2011;472(7341):90–94.
7. Welch DR, Hurst DR. Defining the hallmarks of metastasis. *Cancer Res*. 2019;79(12):3011–3027.
8. Nguyen B, Fong C, Luthra A, Smith SA, DiNatale RG, Nandakumar S, Walch H, Chatila WK, Madupuri R, Kundra R, et al. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell*. 2022;185(3):563–575.e511.
9. Reiter JG, Makohon-Moore AP, Gerold JM, Heyde A, Attiyeh MA, Kohutek ZA, Tokheim CJ, Brown A, DeBlasio RM, Niyazov J, et al. Minimal functional driver gene heterogeneity among untreated metastases. *Science*. 2018;361(6406):1033–1037.
10. Watkins TBK, Lim EL, Petkovic M, Elizalde S, Birkbak NJ, Wilson GA, Moore DA, Gronroos E, Rowan A, Dewhurst SM, et al. Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature*. 2020;587(7832):126–132.
11. Xu Y, Liu K, Li C, Li M, Liu F, Zhou X, Sun M, Ranganathan M, Zhang L, Wang S, et al. The largest Chinese cohort study indicates homologous recombination pathway gene mutations as another major genetic risk factor for colorectal cancer with heterogeneous clinical phenotypes. *Research*. 2023;6:0249.
12. Gavish A, Tyler M, Greenwald AC, Hoefflin R, Simkin D, Tschernichovsky R, Galili Darnell N, Somech E, Barbolin C, Antman T, et al. Hallmarks of transcriptional intratumour heterogeneity across a thousand tumours. *Nature*. 2023;618(7965):598–606.
13. Massague J, Ganesh K. Metastasis-initiating cells and ecosystems. *Cancer Discov*. 2021;11(4):971–994.
14. Rozenblatt-Rosen O, Regev A, Oberdoerffer P, Nawy T, Hupalowska A, Rood JE, Ashenberg O, Cerami E, Coffey RJ, Demir E, et al. The human tumor atlas network: Charting tumor transitions across space and time at single-cell resolution. *Cell*. 2020;181(2):236–249.
15. Li Y, Lu Y, Kang C, Li P, Chen L. Revealing tissue heterogeneity and spatial dark genes from spatially resolved transcriptomics by multiview graph networks. *Research*. 2023;6:0228.
16. Torborg SR, Li Z, Chan JE, Tammela T. Cellular and molecular mechanisms of plasticity in cancer. *Trends Cancer*. 2022;8(9):735–746.
17. Gupta PB, Pastushenko I, Skibinski A, Blanpain C, Kuperwasser C. Phenotypic plasticity: Driver of cancer initiation, progression, and therapy resistance. *Cell Stem Cell*. 2019;24(1):65–78.
18. Karras P, Bordeu I, Pozniak J, Nowosad A, Pazzi C, Van Raemdonck N, Landeloos E, Van Herck Y, Pedri D, Bervoets G, et al. A cellular hierarchy in melanoma uncouples growth and metastasis. *Nature*. 2022;610(7930):190–198.
19. Mzoughi S, Schwarz M, Wang X, Demircioglu D, Ulukaya G, Mohammed K, Zorgati H, Torre D, Tomalin LE, Di Tullio F, et al. Oncofetal reprogramming drives phenotypic plasticity in WNT-dependent colorectal cancer. *Nat Genet*. 2025;57(2):402–412.
20. Moorman A, Benitez EK, Cambulli F, Jiang Q, Mahmoud A, Lumish M, Hartner S, Balkaran S, Bermeo J, Asawa S, et al. Progressive plasticity during colorectal cancer metastasis. *Nature*. 2025;637:947–954.
21. Rousseaux S, Debernardi A, Jacquiau B, Vitte AL, Vesin A, Nagy-Mignotte H, Moro-Sibilot D, Brichon PY, Lantuejoul S, Hainaut P, et al. Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers. *Sci Transl Med*. 2013;5(186):186ra166.
22. Kshitiz AJ, Maziarz JD, Hamidzadeh A, Liang C, Erkenbrack EM, Kim HN, Haeger JD, Pfarrer C, Hoang T, et al. Evolution of placental invasion and cancer metastasis are causally linked. *Nat Ecol Evol*. 2019;3(12):1743–1753.
23. Smith ZD, Shi J, Gu H, Donaghey J, Clement K, Cacchiarelli D, Gnirke A, Michor F, Meissner A. Epigenetic restriction of extraembryonic lineages mirrors the somatic transition to cancer. *Nature*. 2017;549:543–547.
24. Luo Y, Liang H. Developmental-status-aware transcriptional decomposition establishes a cell state panorama of human cancers. *Genome Med*. 2024;16:124.
25. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell*. 2016;166(1):21–45.
26. Yang J, Antin P, Berx G, Blanpain C, Brabletz T, Bronner M, Campbell K, Cano A, Casanova J, Christofori G, et al. Guidelines and definitions for research on epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol*. 2020;21(6):341–352.
27. Diepenbruck M, Christofori G. Epithelial-mesenchymal transition (EMT) and metastasis: Yes, no, maybe? *Curr Opin Cell Biol*. 2016;43:7–13.

28. Pastushenko I, Mauri F, Song Y, de Cock F, Meeusen B, Swedlund B, Impens F, Van Haver D, Opitz M, Thery M, et al. Fat1 deletion promotes hybrid EMT state, tumour stemness and metastasis. *Nature*. 2021;589(7842):448–455.
29. Youssef KK, Narwade N, Arcas A, Marquez-Galera A, Jimenez-Castano R, Lopez-Blau C, Fazilaty H, Garcia-Gutierrez D, Cano A, Galceran J, et al. Two distinct epithelial-to-mesenchymal transition programs control invasion and inflammation in segregated tumor cell populations. *Nat Cancer*. 2024;5(11):1660–1680.
30. Simeonov KP, Byrns CN, Clark ML, Norgard RJ, Martin B, Stanger BZ, Shendure J, McKenna A, Lengner CJ. Single-cell lineage tracing of metastatic cancer reveals selection of hybrid EMT states. *Cancer Cell*. 2021;39(8):1150–1162.e9.
31. Pastushenko I, Brisebarre A, Sifrim A, Fioramonti M, Revenco T, Boumahdi S, Van Keymeulen A, Brown D, Moers V, Lemaire S, et al. Identification of the tumour transition states occurring during EMT. *Nature*. 2018;556(7702):463–468.
32. Zhang Y, Donaher JL, Das S, Li X, Reinhardt F, Krall JA, Lambert AW, Thiru P, Keys HR, Khan M, et al. Genome-wide CRISPR screen identifies PRC2 and KMT2D-COMPASS as regulators of distinct EMT trajectories that contribute differentially to metastasis. *Nat Cell Biol*. 2022;24:554–564.
33. Luond F, Sugiyama N, Bill R, Bornes L, Hager C, Tang F, Santacroce N, Beisel C, Ivanek R, Burglin T, et al. Distinct contributions of partial and full EMT to breast cancer malignancy. *Dev Cell*. 2021;56(23):3203–3221.e11.
34. Xu X, Rock JR, Lu Y, Futtner C, Schwab B, Guinney J, Hogan BL, Onaitis MW. Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc Natl Acad Sci USA*. 2012;109(13):4910–4915.
35. Cully M, Downward J. SnapShot: Ras signaling. *Cell*. 2008;133(7):1292.
36. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, Dimitriadoy S, Liu DL, Kantheti HS, Saghafinia S, et al. Oncogenic signaling pathways in the cancer genome atlas. *Cell*. 2018;173(2):321–337.e10.
37. Desai TJ, Brownfield DG, Krasnow MA. Alveolar progenitor and stem cells in lung development, renewal and cancer. *Nature*. 2014;507(7491):190–194.
38. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell*. 2012;150(6):1107–1120.
39. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature*. 2014;511(7511):543–550.
40. Campbell JD, Alexandrov A, Kim J, Wala J, Berger AH, Pedamallu CS, Shukla SA, Guo G, Brooks AN, Murray BA, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat Genet*. 2016;48(6):607–616.
41. Marjanovic ND, Hofree M, Chan JE, Canner D, Wu K, Trakala M, Hartmann GG, Smith OC, Kim JY, Evans KV, et al. Emergence of a high-plasticity cell state during lung cancer evolution. *Cancer Cell*. 2020;38(2):229–246.e13.
42. Orstad G, Fort G, Parnell TJ, Jones A, Stubben C, Lohman B, Gillis KL, Orellana W, Tariq R, Klingbeil O, et al. FoxA1 and FoxA2 control growth and cellular identity in NKX2-1-positive lung adenocarcinoma. *Dev Cell*. 2022;57(15):1866–1882.e10.
43. Yang D, Jones MG, Naranjo S, Rideout WM III, Min KHJ, Ho R, Wu W, Replogle JM, Page JL, Quinn JJ, et al. Lineage tracing reveals the phylodynamics, plasticity, and paths of tumor evolution. *Cell*. 2022;185(11):1905–1923.e25.
44. Little DR, Lynch AM, Yan Y, Akiyama H, Kimura S, Chen J. Differential chromatin binding of the lung lineage transcription factor NKX2-1 resolves opposing murine alveolar cell fates in vivo. *Nat Commun*. 2021;12(1):2509.
45. Concepcion CP, Ma S, LaFave LM, Bhutkar A, Liu M, DeAngelo LP, Kim JY, Del Priore I, Schoenfeld AJ, Miller M, et al. Smarca4 inactivation promotes lineage-specific transformation and early metastatic features in the lung. *Cancer Discov*. 2022;12(2):562–585.
46. LaFave LM, Kartha VK, Ma S, Meli K, Del Priore I, Lareau C, Naranjo S, Westcott PMK, Duarte FM, Sankar V, et al. Epigenomic state transitions characterize tumor progression in mouse lung adenocarcinoma. *Cancer Cell*. 2020;38(2):212–228.e13.
47. Kaiser AM, Gatto A, Hanson KJ, Zhao RL, Raj N, Ozawa MG, Seoane JA, Bieging-Rolett KT, Wang M, Li I, et al. p53 governs an AT1 differentiation programme in lung cancer suppression. *Nature*. 2023;619(7971):851–859.
48. Pierce SE, Granja JM, Corces MR, Brady JJ, Tsai MK, Pierce AB, Tang R, Chu P, Feldser DM, Chang HY, et al. LKB1 inactivation modulates chromatin accessibility to drive metastatic progression. *Nat Cell Biol*. 2021;23(8):915–924.
49. Sun X, Perl AK, Li R, Bell SM, Sajti E, Kalinichenko VV, Kalin TV, Misra RS, Deshmukh H, Clair G, et al. A census of the lung: CellCards from LungMAP. *Dev Cell*. 2022;57(1):112–145.e2.
50. Xiao Y, Jin W, Ju L, Fu J, Wang G, Yu M, Chen F, Qian K, Wang X, Zhang Y. Tracking single-cell evolution using clock-like chromatin accessibility loci. *Nat Biotechnol*. 2024;43:784–798.
51. Martins LR, Sieverling L, Michelhans M, Schiller C, Erkut C, Grunewald TGP, Triana S, Frohling S, Velten L, Glimm H, et al. Single-cell division tracing and transcriptomics reveal cell types and differentiation paths in the regenerating lung. *Nat Commun*. 2024;15(1):2246.
52. Juul NH, Yoon JK, Martinez MC, Rishi N, Kazadaeva YI, Morri M, Neff NF, Trope WL, Shrager JB, Sinha R, et al. KRAS(G12D) drives lepidic adenocarcinoma through stem-cell reprogramming. *Nature*. 2023;619:860–867.
53. Dost AFM, Moye AL, Vedaie M, Tran LM, Fung E, Heinze D, Villacorta-Martin C, Huang J, Hekman R, Kwan JH, et al. Organoids model transcriptional hallmarks of oncogenic KRAS activation in lung epithelial progenitor cells. *Cell Stem Cell*. 2020;27(4):663–678.e8.
54. Moye AL, Dost AF, Ietswaart R, Sengupta S, Ya V, Aluya C, Fahey CG, Louie SM, Paschini M, Kim CF. Early-stage lung cancer is driven by a transitional cell state dependent on a KRAS-ITGA3-SRC axis. *EMBO J*. 2024;43(14):2843–2861.
55. DeTomaso D, Yosef N. Hotspot identifies informative gene modules across modalities of single-cell genomics. *Cell Syst*. 2021;12(5):446–456.e9.
56. Winslow MM, Dayton TL, Verhaak RG, Kim-Kiselak C, Snyder EL, Feldser DM, Hubbard DD, DuPage MJ, Whittaker CA, Hoersch S, et al. Suppression of lung adenocarcinoma progression by Nkx2-1. *Nature*. 2011;473(7345):101–104.
57. Li CM, Gocheva V, Oudin MJ, Bhutkar A, Wang SY, Date SR, Ng SR, Whittaker CA, Bronson RT, Snyder EL, et al.

Foxa2 and Cdx2 cooperate with Nkx2-1 to inhibit lung adenocarcinoma metastasis. *Genes Dev*. 2015;29(17):1850–1862.
58. Hwang DH, Sholl LM, Rojas-Rudilla V, Hall DL, Shivdasani P, Garcia EP, MacConaill LE, Vivero M, Hornick JL, Kuo FC, et al. KRAS and NKX2-1 mutations in invasive mucinous adenocarcinoma of the lung. *J Thorac Oncol*. 2016;11(4):496–503.
59. Thompson EW, Paik S, Brunner N, Sommers CL, Zugmaier G, Clarke R, Shima TB, Torri J, Donahue S, Lippman ME, et al. Association of increased basement membrane invasiveness with absence of estrogen receptor and expression of vimentin in human breast cancer cell lines. *J Cell Physiol*. 1992;150(3):534–544.
60. Oft M, Akhurst RJ, Balmain A. Metastasis is driven by sequential elevation of H-ras and Smad2 levels. *Nat Cell Biol*. 2002;4(7):487–494.
61. Janda E, Lehmann K, Killisch I, Jechlinger M, Herzig M, Downward J, Beug H, Grunert S. Ras and TGFβ cooperatively regulate epithelial cell plasticity and metastasis: Dissection of Ras signaling pathways. *J Cell Biol*. 2002;156(2):299–313.
62. Oft M, Heider KH, Beug H. TGFβ signaling is necessary for carcinoma cell invasiveness and metastasis. *Curr Biol*. 1998;8(23):1243–1252.
63. Yang J, Mani SA, Donaher JL, Ramaswamy S, Itzykson RA, Come C, Savagner P, Gitelman I, Richardson A, Weinberg RA. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell*. 2004;117(7):927–939.
64. Thiery JP. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer*. 2002;2(6):442–454.
65. Laughney AM, Hu J, Campbell NR, Bakhoum SF, Setty M, Lavallee VP, Xie Y, Masilionis I, Carr AJ, Kottapalli S, et al. Regenerative lineages and immune-mediated pruning in lung cancer metastasis. *Nat Med*. 2020;26:259–269.
66. Sharma A, Bleriot C, Currenti J, Ginhoux F. Oncofetal reprogramming in tumour development and progression. *Nat Rev Cancer*. 2022;22(10):593–602.
67. Yu J, Yan Y, Li S, Xu Y, Parolia A, Rizvi S, Wang W, Zhai Y, Xiao R, Li X, et al. Progestogen-driven B7-H4 contributes to onco-fetal immune tolerance. *Cell*. 2024;187(17):4713–4732.e19.
68. Malta TM, Sokolov A, Gentles AJ, Burzykowski T, Poisson L, Weinstein JN, Kaminska B, Huelsken J, Omberg L, Gevaert O, et al. Machine learning identifies stemness features associated with oncogenic dedifferentiation. *Cell*. 2018;173(2):338–354e15.
69. Kulesa PM, Kasemeier-Kulesa JC, Teddy JM, Margaryan NV, Seftor EA, Seftor RE, Hendrix MJ. Reprogramming metastatic melanoma cells to assume a neural crest cell-like phenotype in an embryonic microenvironment. *Proc Natl Acad Sci USA*. 2006;103(10):3752–3757.
70. Kasemeier-Kulesa JC, Romine MH, Morrison JA, Bailey CM, Welch DR, Kulesa PM. NGF reprograms metastatic melanoma to a bipotent glial-melanocyte neural crest-like precursor. *Biol Open*. 2018;7(1):Article bio030817.
71. Raica M, Cimpean AM, Ribatti D. The role of podoplanin in tumor progression and metastasis. *Anticancer Res*. 2008;28(5B):2997–3006.
72. Marchini C, Montani M, Konstantinidou G, Orru R, Mannucci S, Ramadori G, Gabrielli F, Baruzzi A, Berton G, Merigo F, et al. Mesenchymal/stromal gene expression signature relates to basal-like breast cancers, identifies bone metastasis and predicts resistance to therapies. *PLOS ONE*. 2010;5(11):Article e14131.
73. Wang R, Zhang P, Wang J, Ma L, Weigao E, Suo S, Jiang M, Li J, Chen H, Sun H, et al. Construction of a cross-species cell landscape at single-cell level. *Nucleic Acids Res*. 2023;51(2):501–516.
74. Li J, Dang SM, Sengupta S, Schurmann P, Dost AFM, Moye AL, Trovero MF, Ahmed S, Paschini M, Bhetariya PJ, et al. Organoid modeling reveals the tumorigenic potential of the alveolar progenitor cell state. *EMBO J*. 2025;44(6):1804–1828.
75. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, Ilagan F, Kuo AH, Hsieh RW, Cai S, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Science*. 2020;367(6476):405–411.
76. Gerstung M, Jolly C, Leshchiner I, Dentro SC, Gonzalez S, Rosebrock D, Mitchell TJ, Rubanova Y, Anur P, Yu K, et al. The evolutionary history of 2,658 cancers. *Nature*. 2020;578(7793):122–128.
77. Butler G, Amend SR, Venditti C, Pienta KJ. Punctuational evolution is pervasive in distal site metastatic colonization. *Proc Biol Sci*. 2025;292(2039):20242850.
78. Greaves M, Maley CC. Clonal evolution in cancer. *Nature*. 2012;481:306–313.
79. Spain L, Coulton A, Lobon I, Rowan A, Schnidrig D, Shepherd STC, Shum B, Byrne F, Goicoechea M, Piperni E, et al. Late-stage metastatic melanoma emerges through a diversity of evolutionary pathways. *Cancer Discov*. 2023;13(6):1364–1385.
80. Frankell AM, Dietzen M, Al Bakir M, Lim EL, Karasaki T, Ward S, Veeriah S, Colliver E, Huebner A, Bunkum A, et al. The evolution of lung cancer and impact of subclonal selection in TRACERx. *Nature*. 2023;616(7957):525–533.
81. Baca SC, Prandi D, Lawrence MS, Mosquera JM, Romanel A, Drier Y, Park K, Kitabayashi N, MacDonald TY, Ghandi M, et al. Punctuated evolution of prostate cancer genomes. *Cell*. 2013;153(3):666–677.
82. Gao R, Davis A, McDonald TO, Sei E, Shi X, Wang Y, Tsai PC, Casasent A, Waters J, Zhang H, et al. Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat Genet*. 2016;48(10):1119–1130.
83. Field MG, Durante MA, Anbunathan H, Cai LZ, Decatur CL, Bowcock AM, Kurtenbach S, Harbour JW. Punctuated evolution of canonical genomic aberrations in uveal melanoma. *Nat Commun*. 2018;9:116.
84. Reiter JG, Hung WT, Lee IH, Nagpal S, Giunta P, Degner S, Liu G, Wassenaar ECE, Jeck WR, Taylor MS, et al. Lymph node metastases develop through a wider evolutionary bottleneck than distant metastases. *Nat Genet*. 2020;52:692–700.
85. Winkler J, Tan W, Diadhiou CM, McGinnis CS, Abbasi A, Hasnain S, Durney S, Atamaniuc E, Superville D, Awni L, et al. Single-cell analysis of breast cancer metastasis reveals epithelial-mesenchymal plasticity signatures associated with poor outcomes. *J Clin Invest*. 2024;134(17):Article e164227.
86. Deng S, Gong H, Zhang D, Zhang M, He X. A statistical method for quantifying progenitor cells reveals incipient cell fate commitments. *Nat Methods*. 2024;21:597–608.
87. Perelli L, Zhang L, Mangiameli S, Giannese F, Mahadevan KK, Peng F, Citron F, Khan H, Le C, Gurreri E, et al. Evolutionary fingerprints of epithelial-to-mesenchymal transition. *Nature*. 2025;640:1083–1092.
88. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, Lee JI, Suh YL, Ku BM, Eum HH, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming

of metastatic lung adenocarcinoma. *Nat Commun*. 2020;11(1):2285.
89. Hao Y, Hao S, Andersen-Nissen E, Mauck WM III, Zheng S, Butler A, Lee MJ, Wilk AJ, Darby C, Zager M, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184:3573–3587.e29.
90. Liu YM, Ge JY, Chen YF, Liu T, Chen L, Liu CC, Ma D, Chen YY, Cai YW, Xu YY, et al. Combined single-cell and spatial transcriptomics reveal the metabolic evolvement of breast cancer during early dissemination. *Adv Sci*. 2023;10(6):Article e2205395.
91. Hsu WH, LaBella KA, Lin Y, Xu P, Lee R, Hsieh CE, Yang L, Zhou A, Blecher JM, Wu CJ, et al. Oncogenic KRAS drives lipofibrogenesis to promote angiogenesis and colon cancer progression. *Cancer Discov*. 2023;13(12):2652–2673.
92. Hirz T, Mei S, Sarkar H, Kfoury Y, Wu S, Verhoeven BM, Subtelny AO, Zlatev DV, Wszolek MW, Salari K, et al. Dissecting the immune suppressive human prostate tumor microenvironment via integrated single-cell and spatial transcriptomic analyses. *Nat Commun*. 2023;14(1):663.
93. Kfoury Y, Baryawno N, Severe N, Mei S, Gustafsson K, Hirz T, Brouse T, Scadden EW, Igolkina AA, Kokkaliaris K, et al. Human prostate cancer bone metastases have an actionable immunosuppressive microenvironment. *Cancer Cell*. 2021;39(11):1464–1478.e8.
94. Jiang H, Yu D, Yang P, Guo R, Kong M, Gao Y, Yu X, Lu X, Fan X. Revealing the transcriptional heterogeneity of organ-specific metastasis in human gastric cancer using single-cell RNA sequencing. *Clin Transl Med*. 2022;12(2):Article e730.
95. Yu Z, Lv Y, Su C, Lu W, Zhang R, Li J, Guo B, Yan H, Liu D, Yang Z, et al. Integrative single-cell analysis reveals transcriptional and epigenetic regulatory features of clear cell renal cell carcinoma. *Cancer Res*. 2023;83(5):700–719.
96. Zhang M, Zhai W, Miao J, Cheng X, Luo W, Song W, Wang J, Gao WQ. Single cell analysis reveals intra-tumour heterogeneity, microenvironment and potential diagnosis markers for clear cell renal cell carcinoma. *Clin Transl Med*. 2022;12(5):Article e713.
97. Gonzalez H, Mei W, Robles I, Hagerling C, Allen BM, Hauge Okholm TL, Nanjaraj A, Verbeek T, Kalavacherla S, van Gogh M, et al. Cellular architecture of human brain metastases. *Cell*. 2022;185:729–745.e20.
98. Lu Y, Yang A, Quan C, Pan Y, Zhang H, Li Y, Gao C, Lu H, Wang X, Cao P, et al. A single-cell atlas of the multicellular ecosystem of primary and metastatic hepatocellular carcinoma. *Nat Commun*. 2022;13(1):4594.
99. Choi JH, Lee BS, Jang JY, Lee YS, Kim HJ, Roh J, Shin YS, Woo HG, Kim CH. Single-cell transcriptome profiling of the stepwise progression of head and neck cancer. *Nat Commun*. 2023;14:1055.
100. Durante MA, Rodriguez DA, Kurtenbach S, Kuznetsov JN, Sanchez MI, Decatur CL, Snyder H, Feun LG, Livingstone AS, Harbour JW. Single-cell analysis reveals new evolutionary complexity in uveal melanoma. *Nat Commun*. 2020;11(1):496.
101. Moss J, Magenheim J, Neiman D, Zemmour H, Loyfer N, Korach A, Samet Y, Maoz M, Druid H, Arner P, et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat Commun*. 2018;9(1):5068.
102. Travis WD. Pathology of lung cancer. *Clin Chest Med*. 2011;32:669–692.
103. Zuccato JA, Mamatjan Y, Nassiri F, Ajisebutu A, Liu JC, Muazzam A, Singh O, Zhang W, Voisin M, Mirhadi S, et al. Prediction of brain metastasis development with DNA methylation signatures. *Nat Med*. 2025;31(1):116–125.
104. Li S, Yang M, Teng S, Lin K, Wang Y, Zhang Y, Guo W, Wang D. Chromatin accessibility dynamics in colorectal cancer liver metastasis: Uncovering the liver tropism at single cell resolution. *Pharmacol Res*. 2023;195:Article 106896.
105. Gao GF, Parker JS, Reynolds SM, Silva TC, Wang LB, Zhou W, Akbani R, Bailey M, Balu S, Berman BP, et al. Before and after: Comparison of legacy and harmonized TCGA genomic data commons' data. *Cell Syst*. 2019;9:24–34.e10.
106. Kim H, Nguyen NP, Turner K, Wu S, Gujar AD, Luebeck J, Liu J, Deshpande V, Rajkumar U, Namburi S, et al. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nat Genet*. 2020;52(9):891–897.
107. Baek S, Sung E, Kim G, Hong MH, Lee CY, Shim HS, Park SY, Kim HR, Lee I. Single-cell multi-omics reveals tumor microenvironment factors underlying poor immunotherapy responses in ALK-positive lung cancer. *Cancer Commun*. 2025;45(4):422–427.
108. Serra RW, Fang M, Park SM, Hutchinson L, Green MR. A KRAS-directed transcriptional silencing pathway that mediates the CpG island methylator phenotype. *eLife*. 2014;3:Article e02313.
109. Su J, Huang YH, Cui X, Wang X, Zhang X, Lei Y, Xu J, Lin X, Chen K, Lv J, et al. Homeobox oncogene activation by pan-cancer DNA hypermethylation. *Genome Biol*. 2018;19(1):108.
110. Van Tongelen A, Loriot A, De Smet C. Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes. *Cancer Lett*. 2017;396:130–137.
111. Gazin C, Wajapeyee N, Gobeil S, Virbasius CM, Green MR. An elaborate pathway required for Ras-mediated epigenetic silencing. *Nature*. 2007;449(7165):1073–1077.
112. Maynard A, McCoach CE, Rotow JK, Harris L, Haderk F, Kerr DL, Yu EA, Schenk EL, Tan W, Zee A, et al. Therapy-induced evolution of human lung cancer revealed by single-cell RNA sequencing. *Cell*. 2020;182(5):1232–1251.e22.
113. Chan ASL, Zhu H, Narita M, Cassidy LD, Young ARJ, Bermejo-Rodriguez C, Janowska AT, Chen HC, Gough S, Oshimori N, et al. Titration of RAS alters senescent state and influences tumour initiation. *Nature*. 2024;633:678–685.
114. Han G, Sinjab A, Rahal Z, Lynch AM, Treekitkarnmongkol W, Liu Y, Serrano AG, Feng J, Liang K, Khan K, et al. An atlas of epithelial cell states and plasticity in lung adenocarcinoma. *Nature*. 2024;627(8004):656–663.
115. Ganesan K, Xu C, Wu S, Sui Y, Du B, Zhang J, Gao F, Chen J, Tang H. Ononin inhibits tumor bone metastasis and osteoclastogenesis by targeting mitogen-activated protein kinase pathway in breast cancer. *Research*. 2024;7:0553.
116. Wu YL, Tsuboi M, He J, John T, Grohe C, Majem M, Goldman JW, Laktionov K, Kim SW, Kato T, et al. Osimertinib in resected EGFR-mutated non-small-cell lung cancer. *N Engl J Med*. 2020;383(18):1711–1723.
117. Wu YL, Dziadziuszko R, Ahn JS, Barlesi F, Nishio M, Lee DH, Lee JS, Zhong W, Horinouchi H, Mao W, et al. Alectinib in resected ALK-positive non-small-cell lung cancer. *N Engl J Med*. 2024;390(14):1265–1276.
118. Long GV, Hauschild A, Santinami M, Atkinson V, Mandala M, Chiarion-Sileni V, Larkin J, Nyakas M,

Dutriaux C, Haydon A, et al. Adjuvant dabrafenib plus trametinib in stage III BRAF-mutated melanoma. *N Engl J Med*. 2017;377(19):1813–1823.
119. Long GV, Hauschild A, Santinami M, Kirkwood JM, Atkinson V, Mandala M, Merelli B, Sileni VC, Nyakas M, Haydon A, et al. Final results for adjuvant dabrafenib plus trametinib in stage III melanoma. *N Engl J Med*. 2024;391(18):1709–1720.
120. Awada A, Colomer R, Inoue K, Bondarenko I, Badwe RA, Demetriou G, Lee SC, Mehta AO, Kim SB, Bachelot T, et al. Neratinib plus paclitaxel vs trastuzumab plus paclitaxel in previously untreated metastatic ERBB2-positive breast cancer: The NEfERT-T randomized clinical trial. *JAMA Oncol*. 2016;2(12):1557–1564.
121. Saura C, Oliveira M, Feng YH, Dai MS, Chen SW, Hurvitz SA, Kim SB, Moy B, Delaloge S, Gradishar W, et al. Neratinib plus capecitabine versus lapatinib plus capecitabine in HER2-positive metastatic breast cancer previously treated with ≥ 2 HER2-directed regimens: Phase III NALA trial. *J Clin Oncol*. 2020;38(27):3138–3149.
122. Chalabi Hajkarim M, May M, Amin AD, Jamison J, Tagore S, D'Souza E, Caprio L, Walsh ZH, Fan JL, Zhang M, et al. Cellular states associated with metastatic organotropism and survival in patients with pancreatic ductal adenocarcinoma. *Nat Genet*. 2025;57:2728–2742.
123. Xiao Y, Jin W, Qian K, Ju L, Wang G, Wu K, Cao R, Chang L, Xu Z, Luo J, et al. Integrative single cell atlas revealed intratumoral heterogeneity generation from an adaptive epigenetic cell state in human bladder urothelial carcinoma. *Adv Sci*. 2024;11(24):Article e2308438.
124. Li M, Ang KS, Teo B, Rom U, Nguyen MN, Maurer-Stroh S, Chen J. Rediscovering publicly available single-cell data with the DISCO platform. *Nucleic Acids Res*. 2025;53(D1):D932–D938.
125. Xiao Y, Wang X, Weng H, Ding Z, Qian K, Jin W, Lu S, Ju L, He Z, Wang G, et al. Ultrasensitive tumour-agnostic non-invasive detection of colorectal cancer recurrence using ctDNA methylation. *Clin Transl Med*. 2022;12(9):Article e1015.
126. Zhu T, Liu J, Beck S, Pan S, Capper D, Lechner M, Thirlwell C, Breeze CE, Teschendorff AE. A pan-tissue DNA methylation atlas enables in silico decomposition of human tissue methylomes at cell-type resolution. *Nat Methods*. 2022;19(3):296–306.
127. Xue L, He W, Zhang Y, Wang Z, Chen H, Chen Z, Zhu W, Liu D, Jia H, Jiang Y, et al. Origins of biallelic inactivation of NF2 in neurofibromatosis type 2. *Neuro-Oncology*. 2022;24(6):903–913.