



CANCER

Clonal dominance defines metastatic dissemination in pancreatic cancer

I-Lin Ho^{1,2,†*}, Chieh-Yuan Li^{1,2,†‡*}, Fuchen Wang^{3,†}, Li Zhao², Jingjing Liu², Er-Yen Yen^{1,2}, Charles A. Dyke², Rutvi Shah^{1,2}, Zhaoliang Liu², Ali Osman Çetin⁴, Yanshuo Chu², Francesca Citron², Sergio Attanasio², Denise Corti², Faezeh Darbaniyan³, Edoardo Del Poggetto², Sara Loponte², Jintan Liu^{1,2}, Melinda Soeung^{1,2}, Ziheng Chen², Shan Jiang⁵, Hong Jiang², Akira Inoue², Sisi Gao^{2,5}, Angela Deem², Ningping Feng⁵, Haoqiang Ying⁶, Michael Kim⁷, Virginia Giuliani⁴, Giannicola Genovese⁸, Jianhua Zhang², Andrew Futreal², Anirban Maitra⁹, Timothy Heffernan⁵, Linghua Wang², Kim-Anh Do³, Gaetano Gargiulo⁴, Giulio Draetta², Alessandro Carugo^{5,§}, Ruitao Lin³, Andrea Viale^{2,*}

Copyright © 2024 the Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution NonCommercial License 4.0 (CC BY-NC).

Tumors represent ecosystems where subclones compete during tumor growth. While extensively investigated, a comprehensive picture of the interplay of clonal lineages during dissemination is still lacking. Using patient-derived pancreatic cancer cells, we created orthotopically implanted clonal replica tumors to trace clonal dynamics of unperturbed tumor expansion and dissemination. This model revealed the multifaceted nature of tumor growth, with rapid changes in clonal fitness leading to continuous reshuffling of tumor architecture and alternating clonal dominance as a distinct feature of cancer growth. Regarding dissemination, a large fraction of tumor lineages could be found at secondary sites each having distinctive organ growth patterns as well as numerous undescribed behaviors such as abortive colonization. Paired analysis of primary and secondary sites revealed fitness as major contributor to dissemination. From the analysis of pro- and nonmetastatic isogenic subclones, we identified a transcriptomic signature able to identify metastatic cells in human tumors and predict patients' survival.

INTRODUCTION

Cancer is not a static disease. It is more appropriately considered as a complex ecosystem wherein diverse clonal lineages (also called “subclones”) and their environment continually coadapt and coevolve (1, 2). The shifting over time of the subclonal landscape is termed “clonal evolution,” and it is one of the fundamental mechanisms that drive tumor progression and heterogeneity (3, 4). Clonal evolution occurs during all stages of cancer development, from early lesions to end-stage metastatic disease, and it underlies the tumor response to intrinsic or extrinsic perturbations, where tumors adapt to the new environment by selecting for subclones with a survival advantage (3, 4).

Metastasis is a leading cause of cancer death (5). The metastatic cascade model of tumor dissemination describes a stepwise progression of tumors culminating in metastatic disease, and this model has

become universally accepted (6–8). According to the model, metastases are the end point of a linear evolution wherein tumor cells progressively acquire genomic or epigenetic alterations that sustain tumor cell dissemination and survival in distant microenvironments (7). This model is corroborated by genomic studies documenting higher mutational load in secondary lesions compared with primary tumors, although no specific recurrent genes driving the metastatic process have been confirmed (9–12), apart from *KRAS* and *MYC* amplifications in some cases (13–16). The metastatic cascade model delineates probable mechanisms that endow tumor cells with the required features to successfully disperse and manifest a secondary lesion. However, the model fails to account for emerging paradigms in tumor evolution (3, 17) and only few studies account for tumor heterogeneity (18, 19). It also cannot explain evidence of dissemination from early tumor lesions (20–22), which suggests that metastatic potential is established early during tumor progression and that metastases are probably driven by the same genetic alterations that drive tumorigenesis (23).

Analysis of longitudinal patient liquid or tissue biopsy from clinical studies is uncovering characteristic shifts in tumor clonality over time (24–28). However, these data are necessarily limited in the number of samples per patient as well as confounded by intra- and inter-patient variability as well as tumor stage and genetics. In addition, in the clinical setting, longitudinal sampling is nearly universally obtained in a pretreatment/posttreatment setting and does not support investigation of unperturbed tumor evolution. For tumor types that are typically diagnosed at the stage of advanced disease, there are limited clinical data available to study clonal change representation that might have prognostic significance. One such tumor is pancreatic ductal adenocarcinoma (PDAC), which is frequently diagnosed after the invasion of nearby and distant organs has occurred, and fewer than 10% of patients will survive 5 years after diagnosis (29).

¹The University of Texas MD Anderson Cancer Center UTHealth Graduate School of Biomedical Sciences, Houston, TX, USA. ²Department of Genomic Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ³Department of Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁴Max-Delbrück-Center for Molecular Medicine in the Helmholtz Association (MDC), Robert-Rössle-Str. 10, 13125 Berlin, Germany. ⁵TRACTION platform, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁶Department of Cellular and Molecular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁷Department of Surgical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁸Department of Genitourinary Medical Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. ⁹Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA.

*Corresponding author. Email: aviale@mdanderson.org (A.V.); iho@mdanderson.org (I.-L.H.); alexcyli2015@gmail.com (C.-Y.L.)

†These authors contributed equally to this work.

‡Present address: Mission Bio, South San Francisco, CA 94080, USA.

§Present address: IRBM, Pomezia (Rome), Italy.

To study the clonal evolution of PDAC from primary lesion growth to overt disease dissemination, we leveraged Clonal Replica Tumors (CRTs), an innovative *in vivo* platform to generate cohorts of animals bearing human tumors that are essentially identical in their clonal composition (30). Here, we reconstruct in mouse the complex dynamics of PDAC clones during unperturbed tumor expansion over a period of 3 months and capture the metastatic spreading of tumor cells at the organismal level with unparalleled high resolution. Our study goes beyond measuring colonization as a single end point and investigates the metastatic process as intrinsic within the natural evolution of a complex ecosystem in which distinct populations of cells continuously compete for fitness in the pancreas restrained microenvironment. Our results formally demonstrate that the fitness of a subclone within the primary lesion is the major determinant of whether it eventually disseminates and successfully initiates a tumor at secondary sites.

RESULTS

A high-resolution longitudinal model enables the study of unperturbed tumor expansion

To generate a relevant model that captures PDAC diversity and maintains the clinical features of the original tumor as much as possible, we derived the cell line PATC124 from an early passage xenograft (31) established from a patient who underwent pancreaticoduodenectomy for stage IIA pancreatic cancer (T3N0M0). Isolated cells were confirmed to be genetically heterogeneous using single-cell copy number analysis, and copy number variation (CNV) was documented among subclonal populations (Fig. 1A and fig. S1A). When implanted orthotopically in NSG mice, the established cell line PATC124 manifested primary tumors within approximately 30 to 45 days that maintained the histologically features of the parental xenograft tissue. In addition, after a latency of approximately 3 months, the liver in all transplanted NSG animals was colonized with secondary tumors (Fig. 1B); this recapitulated the clinical history of the patient, who succumbed to liver metastasis 17 months after diagnosis.

We previously established a lentiviral barcoding platform to study clonally identical tumors in large cohorts of animals, which we named CRTs (30). Here, we have modified our approach to generate orthotopic CRTs (oCRTs) that we use for longitudinal studies. Briefly, we transduce isolated cells with a high-complexity barcode library at low multiplicity of infection (MOI < 0.1) to ensure one integrant per cell, select the barcoded population of cells using puromycin, and then minimally passage cells *in vitro* to stabilize the barcoded population before expansion and orthotopic injection into the pancreas of recipient NSG animals (30). We then study unperturbed tumor expansion at the organismal level in cohorts ($n = 2$) that are euthanized to collect primary tumor (pancreas), blood (~0.8 ml), whole liver, and lung tissues over a period of 3 months. Last, barcoded cell populations in each tissue are analyzed by next-generation sequencing (NGS) (Fig. 1C).

We compared the subclonal composition of PATC124 cells from all the primary tumors by NGS barcode analysis. Pairwise comparison demonstrated that the proportion of the paired barcodes within each primary tumor was highly correlated through the 10-week time point (Pearson's $R = 0.999$ at day 0, 0.991 at week 4, 0.904 at week 6, and 0.963 at week 10), with decreased correlation observed at the final time point (week 14; $R = 0.727$) (Fig. 1D). We further showed that all PATC124 oCRTs shared a common set of lineages represented by 934

barcodes that were present from the beginning of the experiment (day 0, transplantation) through study end point (week 14) (Fig. 1D). This common set of barcodes represented 93% of the barcoded cells in the inoculum at day 0, and it represented virtually the entire tumor mass of the primary tumors at each following time point (between $R = 0.990$ and $R = 0.998$) until the last time point (week 14; $R = 0.714$). The high representation of tumor mass by a subset of common lineages as well as the high correlation among independent tumors with regard to subclonal composition satisfy the criteria we previously defined for CRTs (30), suggesting that our orthotopic approach, oCRTs, can support studies of clonally identical tumors through at least 10 weeks, after which time barcode correlation begins to diminish and animals become moribund.

To further assess the robustness of PATC124 oCRTs for longitudinal tracing, we performed a series of data simulation and statistical tests (fig. S1B). First, we defined very stringent criteria of similarity using both the differences between the replica vials of injected cells (identical technical replicates) and the differences between clonal lineages that support oCRTs at week 4. We applied these criteria to simulate each oCRT 5000 times and calculated Pearson's correlation for pairwise comparisons of individual subclones and of average Euclidean distance of subclones within whole tumors. This confirmed that the clonal similarity of paired oCRTs is maintained over time, and tumors, although experience a pattern of natural divergence toward the final time point, are still considered identical by statistical tests at week 14 (Fig. 1E and fig. S1, C and D).

In summary, oCRTs progress very similarly over time, but intrinsic (genetic/epigenetic) or extrinsic (microenvironmental) factors in the confined environment of the mouse pancreas may influence individual clonal fitness as the tumor expands, leading to a progressive divergence between individual tumors. However, even when correlation dropped at week 14, tumors appeared minimally divergent within paired animals, with only 0.96% of tumor lineages that presented a discordance between 1 and 5% in terms of representation (Fig. 1F). To validate and strengthen our conclusions, we generated a second and third model of oCRT from other two patient-derived xenograft (PDX) cell lines (fig. S1, E and F): PATC69, a PDX-derived cell line established from a patient who underwent distal pancreatectomy for stage IIA pancreatic cancer (T3N0M0) and eventually succumbed to lung metastasis at 25 months after diagnosis, and PATC108, a PDX cell line established from a patient who underwent pancreaticoduodenectomy for stage IIA pancreatic cancer (T3N0M0) and showed no evidence of secondary disease. Specifically, PATC69 oCRTs are maintained by 229 common clones that represent virtually the entire tumor mass and have a correlation of 0.99 (fig. S1E) and PATC108 oCRTs are maintained by 1240 common clones that represent 98 to 99% of the tumor mass with a correlation above 0.84 (fig. S1F). Thus, our empirically derived data support that oCRTs within a cohort are interchangeable for a period of at least 14 weeks, and we conclude that this platform is suitable for longitudinal studies of subclonal expansion and dissemination.

Complex clonal dynamics characterize unperturbed tumor expansion at the primary site

On the basis of the above results, we used our PATC124 oCRT model to study the unperturbed clonal dynamics of primary tumors. To exhaustively capture clonal behaviors over time, we analyzed common

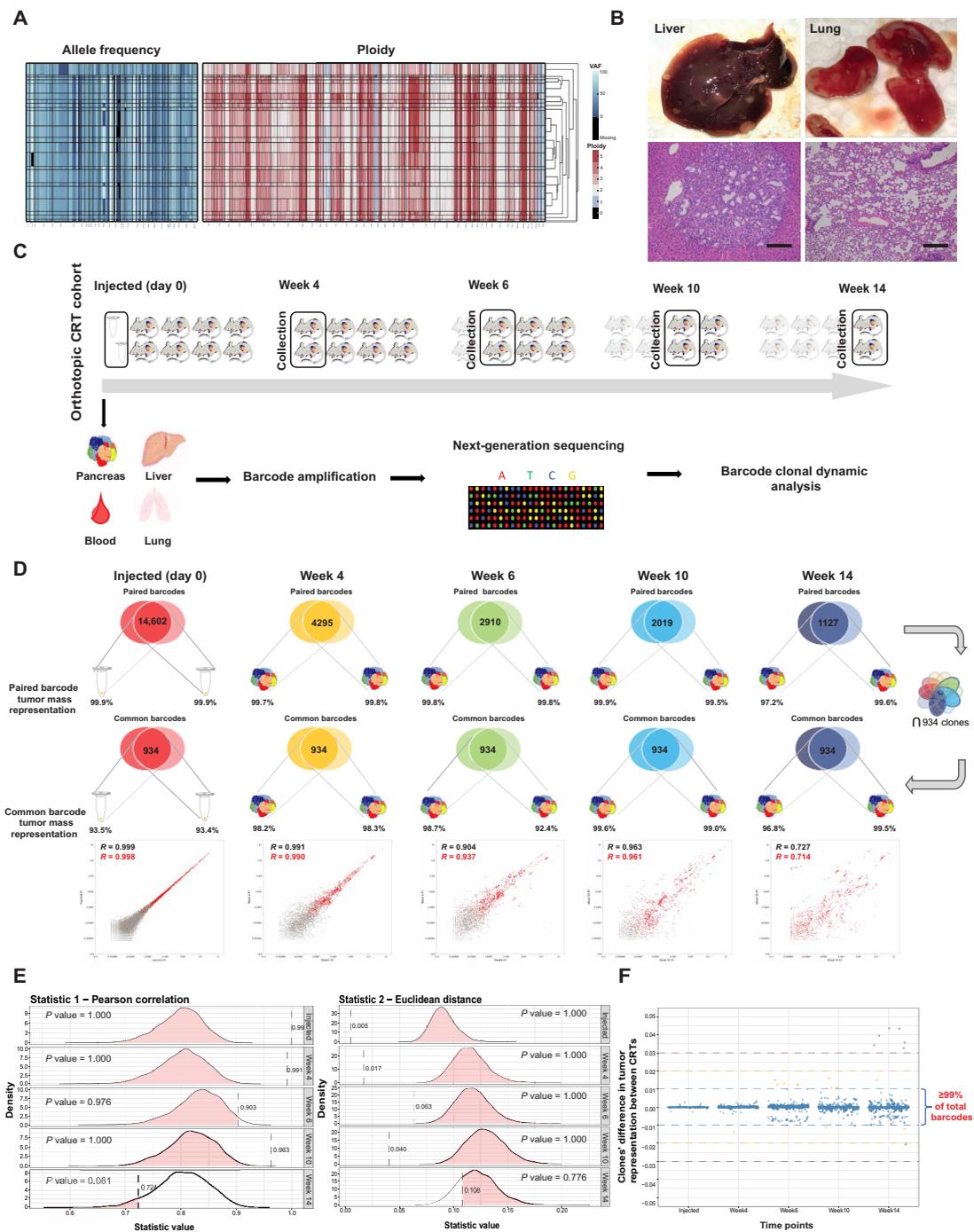


Fig. 1. Quantitative longitudinal characterization of clonal dynamics during unperturbed tumor expansion and dissemination in oCRTs. (A) Single-cell CNV analysis of PATC124 cells through Tapestry platform. The heatmap (left) shows the distribution of 24 clones based on 126 variants and their mean allele frequency. The right heatmap shows the hierarchical clustering of the copy number profiles across the genome. The height of each clone represents their respective fraction in the sample. (B) Metastatic organotropism of PATC124 oCRTs. When moribund conditions were observable, livers and lungs were collected for gross anatomical analysis (top quadrants) and histological analysis (bottom quadrants) (scale bar, 200 μ m). (C) Schematic of the longitudinal PATC124 oCRT study. (D) Clonal correlation among PATC124 oCRTs at different time points. Top, barcoded lineages shared by primary tumors at each time point and their relative representation (%) of tumor mass; middle, set of 934 barcodes common to all tumors and their relative representation (%) of tumor mass; bottom, scatter plots displaying representation of individual lineages derived from $n = 2$ tumors at each time point (red, set of common clones). (E) Resampling bootstrap tests of the 3751 barcodes from $n = 2$ PATC124 tumors at each time point. Pearson's correlation (left) and Euclidean distance (right) tests on simulated data were based on the empirically measured difference between the two tumors at injected and week 4. Distribution curves represent the simulated "true" values, and the dotted lines represent the empirical observed values. If the empirical value falls within and/or above the true distribution (Pearson's correlation) or within and/or below (Euclidean distance), it indicates insufficient evidence to determine that the two samples are different. (F) Differential representation of the 934 common barcodes from $n = 2$ PATC124 tumors at each time point. Each dot represents a unique barcode, and the y axis indicates the difference in percentage of the same barcode in the two tumors at each time point.

barcodes in primary tumors at each time point throughout a 3-month study. Cluster analysis of relative representation within the tumor, normalized to representation in the barcoded cell inoculum at the time of injection, was used to visualize the full spectrum of clonal fitness dynamics during unperturbed tumor expansion (Fig. 2A). This analysis revealed that, contrary to most established theoretical models of tumor evolution (32), many lineages did not progress linearly over time. That is, rather than experiencing a continuous gain or loss of fitness, such as was observed for clusters 1, 7, and 10, many subclonal lineages displayed far more complex fluctuations; for example, clusters 2, 3, 5, and 12 experienced an initial gain of fitness followed by loss of fitness, while cluster 11 initially demonstrated a loss of fitness followed by a loss followed by a pronounced gain of fitness. Some clusters were characterized by even more complex dynamicity, changing their fitness trajectory multiple times during tumor expansion (clusters 4 and 9).

To further characterize the complexity of clonal dynamics, we clustered the lineages into quartiles based on week 14 cumulative clonal abundance (Fig. 2B). Complex dynamics were maintained across all the quartiles, implying that oscillating fitness is independent of the representation subclones have within the tumor (Fig. 2C). When the relative representation of subclones is tracked over time during unperturbed tumor expansion, it clearly demonstrates an alternating clonal dominance (ACD); that is, the relative representation of individual subclones over time can change drastically, and lineages can dominate the tumor volume for a discrete time and then be overtaken by other lineages (Fig. 2C). Some lineages experienced a loss in fitness over time, but they remained detectable through the study end point, albeit at low abundance (Fig. 2C). We captured lineages that were lowly represented during early time points, remaining functionally dormant, but then experienced a burst in fitness (fig. S2A). It appears that ACD is a common behavior during unperturbed clonal expansion, as shown by the other PDX-derived models (fig. S2, B and C). We extended our observation to oCRTs derived from a clinically relevant pancreatic cancer mouse model with *p48-Cre*, *Kras*^{G12D}, and *Trp53*^{R172H} (KPC) (33), demonstrating that ACD is a feature of tumors that also expand into immunocompetent microenvironments (fig. S2D).

This empirical dataset paints a picture of highly dynamic tumor clonal architecture, even in unperturbed tumors, documenting a continuous reshuffling of tumor clonal composition in response to evolutionary pressures of the confined environment. We further confirm a rich reservoir of lowly represented subclones that are persistent throughout tumor expansion and that have the potential to expand and establish dominance and a new equilibrium.

Distinct growth dynamic phenotypes characterize metastatic lineages

To study tumor dissemination, we performed a quantitative longitudinal survey of the PATC124 primary tumor, blood, liver, lung, and ascites at each time point. We used a spike-in control, a “conversion scale” of known cell counts carrying unique barcodes that are distinct from the clonal tracking library to enable conversion of barcode reads into cell count, thus quantifying both clonal complexity (barcode composition) and cell number (Fig. 3, A and B).

Analysis of whole-body blood retrieved by cardiac puncture indicated that circulating tumor cells (CTCs) were rare, on the order of tens or hundreds per milliliter of blood. The number of CTCs was variable and not very consistent across animals and time points, with

a gradual increase in cell number observed at study end point (Fig. 3A). However, a note of caution should be considered when analyzing blood data, as the variability could be attributed to the shedding of tumor DNA in circulation from primary tumors, thus making it difficult to confirm the source.

Similarly to the blood, the number of cells detected in the lungs fluctuated over time, probably owing to the specific organotropism of the cell line model under study. The number of cells in the liver consistently increased across all time points, and at week 14, all animals had a large number of cells detected in the ascitic fluid (Fig. 3A).

As expected, at end point, the barcode composition of the PATC124 primary tumor was richer and more complex than any of the other tissues examined, with 934 distinct lineages detected (Fig. 1D). Although clonal composition was largely reduced compared to the primary tumor in the liver and lungs, we still determined that approximately 47% (438) of the subclones present in the primary tumor had disseminated (fig. S2E). Of these metastatic subclones, ~42% had a specific organotropism (97 and 89 lineages found exclusively in the liver or lung, respectively), while 58% (252) subclones were detected in both organs (fig. S2E).

Next, we further investigated the dissemination dynamics of the PATC69 and PATC108 primary tumors. In PATC69, a metastatic model that develops macroscopic nodes in the lung (fig. S1E), 93% (213) of the 229 barcodes that constitute the set of common barcodes across all primary tumors were detected in the periphery (fig. S3A). Unexpectedly, in PATC108, 30% (373) of the 1240 barcodes that constitute the set of common barcodes were detected in the periphery (fig. S3B), despite patient and mice data demonstrating the model to be nonmetastatic. These findings were unexpected based on the metastatic cascade model, which predicts rare dissemination and tissue engraftment of tumor cells (6, 7), and suggest that, even in the absence of colonization, cells can still shed from primary tumors and be found in the periphery.

Next, we analyzed clonal dynamics of dissemination and colonization. Unlike the lungs and blood, where the number of cells and complexity of barcodes varied over time, in the PATC124 model, seeding to the liver was accomplished early by large number of lineages. The number of cells in the liver increased gradually across time points, while the complexity was relatively stable (Fig. 3, A and B). Cluster analysis of subclone expansion in the liver uncovered at least two major behaviors (Fig. 3C): subclones that were detected early in the liver and then expanded (colonizing subclones) (clusters 2, 3, and 5; 26.4, 8.9, and 8.0%, respectively), and subclones that seeded early and underwent transient expansion before losing fitness and disappearing (abortive subclones) (clusters 1, 4, and 6; 45.8, 8.3, and 2.6%, respectively). Colonizing subclones represented 43% of the clones that seed in the liver clones (151 of 349), and they could undergo either a more constant, gradual expansion (cluster 3), reaching an early plateau (cluster 5) or remain relatively static for several weeks followed by expansion at later time points (cluster 2). Subclones that seeded the liver (~57%) were abortive, indicating that they were proliferative exhausted or failed to adapt to the new environment (Fig. 3C). This abortive colonization phenotype was further exacerbated in the lung, where virtually all detected subclones underwent multiple transient expansions before being counter-selected, failing to generate metastasis (Fig. 3D). To demonstrate that barcode sequencing is detecting real cells, we performed immunofluorescence staining for the major histocompatibility complex using an anti-human leukocyte antigen (HLA) antibody. Lung samples from our PATC124 model sequenced at week

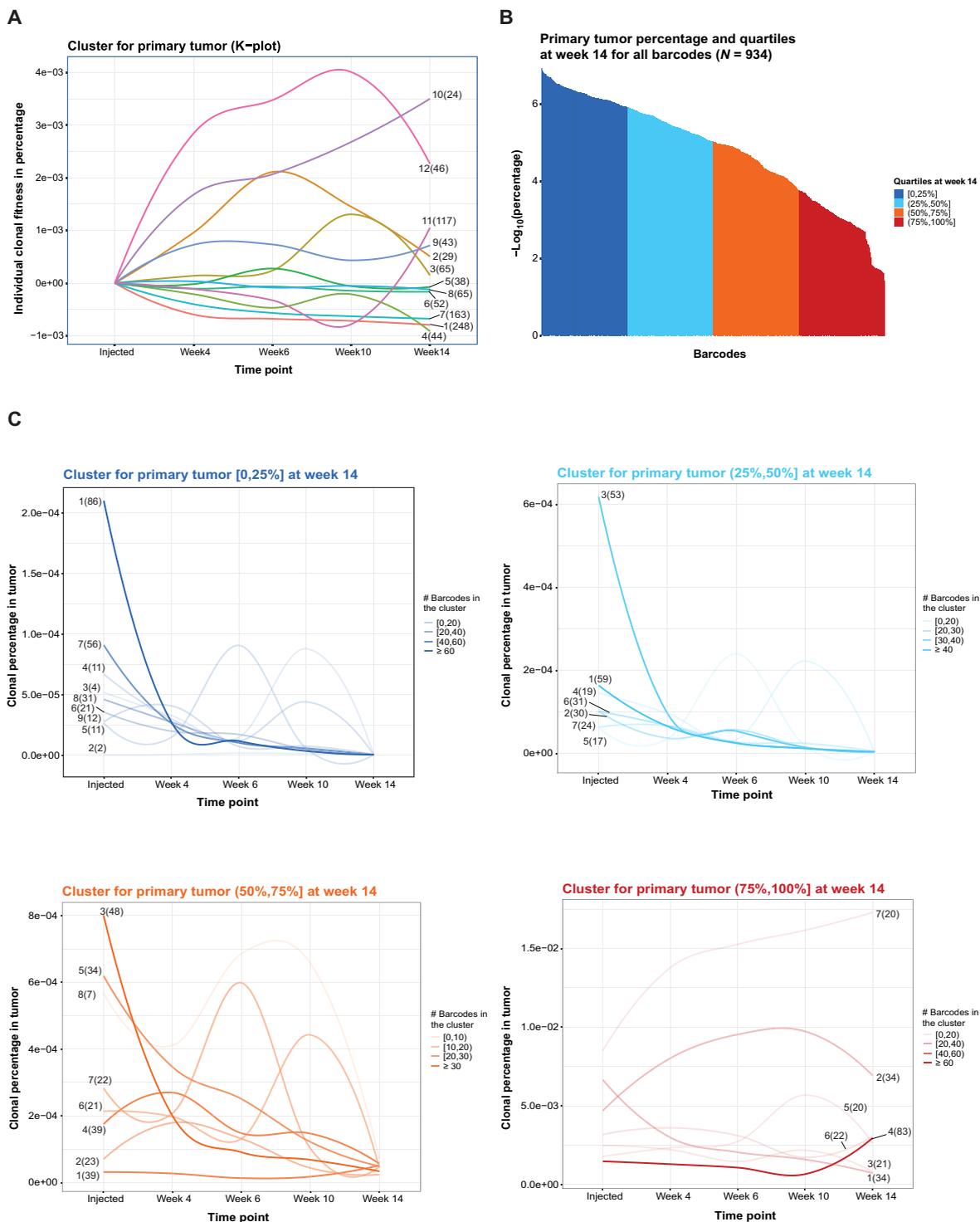


Fig. 2. Longitudinal study of oCRTs shows ACD during tumor expansion. (A) Barcode analysis of $n = 2$ PATC124 tumors per time point was weighted, smoothed, and normalized to injected samples. Barcodes were grouped into 12 discrete dynamic growth patterns (clusters) and plotted as percent of total tumor over time (colored lines). On the right of each line, the numbers outside the brackets indicate the cluster name, and the numbers within the brackets indicate the number of barcodes within the cluster. (B and C) Barcodes of PATC124 were divided into quartiles (0 to 25%, 25 to 50%, 50 to 75%, and 75 to 100%) on the basis of their percent of total tumor mass at week 14 (B), and the clonal dynamic behaviors were clustered and plotted over time (C). The numbers outside brackets indicate the cluster name, and the numbers within brackets indicate the exact number of barcodes within the cluster (darker colors are associated with larger number of barcodes within the cluster).

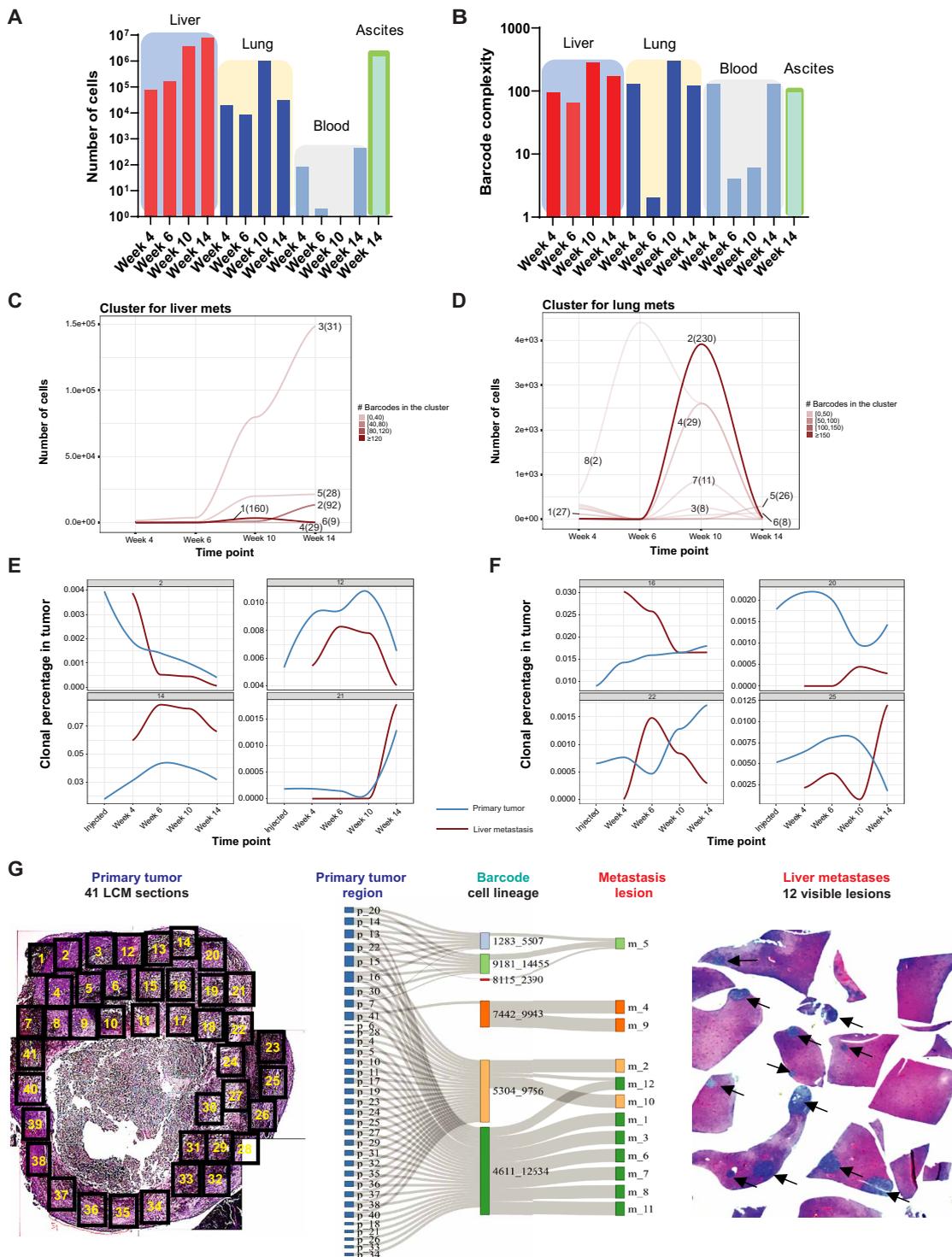


Fig. 3. Longitudinal study of oCRTs reveals complex clonal dynamics at secondary tumor sites. (A) Cumulative cell number of PATC124 at distant sites (liver, lung, blood, and ascites) over time. Cell number at each time point is converted on the basis of spike-in scales. (B) Barcode complexity of PATC124 at distant sites over time. (C and D) Dynamic subclonal growth of PATC124 in liver (C) and lung (D). Cell number at each time point is converted on the basis of spike-in scales. Lineage behavior was grouped into discrete dynamic growth patterns (clusters) over time. The numbers outside the brackets indicate the cluster name, and the numbers within the brackets indicate the exact number of barcodes within the cluster (darker colors are associated with larger number of barcodes within the cluster). (E and F) Dynamic growth of PATC124 subclones in the primary tumor with similar (E) or different (F) trends in the liver. (G) Spatial analysis of subclonal composition of primary tumor and liver metastases in PATC124. Hematoxylin and eosin staining of a section of primary tumor (left) is shown. In the primary tumor, laser capture microdissection (LCM) identified 41 distinct subclonal regions (excluding a necrotic area). Twelve visible metastatic lesions were identified in the liver (right). Each region/lesion was analyzed by NGS to determine subclone composition. Subclones in the primary tumor are mapped to the liver metastases (only clones >10% in metastatic lesions are shown). Line thickness indicates the abundance of the barcode in the metastatic lesion (middle).

14 revealed the presence of small clusters of human cells despite the absence of large metastatic nodes, which is consistent with the sequencing data (fig. S3E). Evaluation of the proliferative index through immunostaining for Ki67 showed positivity of the human tumor cells, although at a much lower extent than tumor cells in the liver, where metastatic cells grow as large nodes (fig. S3E).

Similar behaviors were found in both PATC69 and PATC108 models (fig. S3, C and D). Unexpectedly, on the basis of patient clinical history and tumor tropism in mice, where PATC69 macroscopic metastatic nodes were detected only in the lungs (fig. S1E), the clonal composition of disseminated cells appeared almost identical between lungs and liver, with 97% overlap (fig. S3A). Of note, despite no visible macro-metastases, 63 clones showed expansion in the liver as observed in the lung, although at a lower extent (fig. S3C). When we stained PATC69 liver and lung with anti-HLA and Ki67 antibodies, we clearly observed some regions in the liver to be sparsely populated by disorganized cells positive for HLA that appear different from the more organized and proliferative macro-metastatic nodes found in lungs (fig. S3F), suggesting that cells shed from primary tumors got trapped in the liver as single cells or small clusters and gradually accumulated in considerable numbers due to liver's extensive vasculature and large organ size. Despite the nonselective seeding and different patterns of growth in different organs, these data also demonstrate that a substantial fraction of tumor clones undergo abortive colonization in PATC69 [70% of clones (150 of 213) in liver and 48% of clones (98 of 206) in lung] (fig. S3C).

When examining clonal dynamics in the liver and lung of the non-metastatic PATC108 model, we found behaviors compatible with both clonal expansion and abortive colonization, although the extremely low cell numbers caution against any conclusive interpretation (fig. S3D). Coupled with visual and histological findings showing no visible macroscopic metastasis being detected but a few isolated human cells only (fig. S3G), data suggest that PATC108 cells, although capable of dissemination, fail to adapt to the microenvironments of distal organs and thus cannot form metastasis, which is consistent with the clinical history of the patient whom PATC108 cells originated from.

Because oCRTs enable us to study the same lineages derived from the same primary tumor at distinct sites, focusing on PATC124 model, we next investigated the parallel evolution of subclones at the primary and peripheral sites. By overlaying the dynamic growth curve of a subclone in the primary tumor with its behavior in the liver, we can appreciate that some subclones have relatively similar growth behavior in both environments (Fig. 3E and fig. S2F), whereas some subclones behave quite differently in the two environments (Fig. 3F and fig. S2F). These observations suggest that selective environmental pressures and intrinsic clonal features both contribute to the fitness of a lineage to colonize a distant organ.

Because of the selective organotropism of the model under study, we further characterized the macroscopic metastatic lesions found in the liver of PATC124 oCRTs at week 14. By coupling NGS barcode detection with laser-capture microdissection, the clonal composition of individual nodes of the metastatic lesion as well as the spatial distribution of the metastatic subclones within the primary tumor were mapped (Fig. 3G). Large liver lesions (10 of 12) were mostly monoclonal in composition, representing the repetitive localization of the same three subclones (Met 2, 10; Met 4, 9; and Met 1, 3, 6 to 8, 11), with fewer (2 of 12) oligoclonal lesions dominated by a single lineage (Met 5, 12) (Fig. 3G). Subclones that sustain 8 of

the 10 large monoclonal liver lesions were diffused within the primary tumor, whereas the remaining 2 (nodes 4 and 9) derived from a single subclone that populated a specific region of the primary tumor (#41). Regarding the polyclonal nodes, Met 12 was sustained by a combination of clones spread throughout the primary tumor, and Met 5 derived from three subclones that were in close proximity within the primary tumor, suggesting that this metastasis originated from a single cluster of cells.

Intrinsic long-term replicative potential is a key factor determining clonal fitness and metastasis outcome

Because multiple nodes in the liver were sustained by single clones that were highly represented in the pancreas tumor (Fig. 3G), we hypothesized that the fitness and abundance of a lineage in the primary tumor may be predictive of successful dissemination to peripheral sites. To test this, we arranged a heatmap of subclones in the PATC124 primary tumor at the last time point (week 14) from low to high abundance, and we annotated the clones that were detected in the tissue outside of the primary tumor applying a cutoff of 200 cells, as used in the clinic to define metastasis (34).

This analysis showed that most of the clones that were highly represented within the primary tumor (high fitness) were detected in the periphery (Fig. 4A). Moreover, subclonal abundance at early time points (weeks 4, 6, and 10) also showed high correlation with dissemination at week 14 (fig. S4A). Consistently, similar correlation was confirmed in both PATC69 and PATC108 models (fig. S4, B and C). This was further supported by comparing plots of percent composition of the primary tumor for all subclones that disseminated versus those that did not, which demonstrated that the subclones that reached the periphery comprised a much larger volume of the primary tumor (Fig. 4B). Although relative fitness of subclones in the primary tumor highly correlated with clonal dissemination potential, we also identified subclones with low abundance in the primary tumor that disseminated, as well as subclones with relatively high abundance that were not found in the periphery (Fig. 4, A and B, and fig. S4, A to C).

We posited that the intertwined relationship between intrinsic fitness in the primary tumor and dissemination potential may be related to the intrinsic ability of any given lineage to propagate, also known as long-term proliferative capability or long-term self-renewal (LTSR) (30, 35). To minimize the confounding effect of clonal abundance in these studies, immediately after puromycin, while barcode composition is highly complex and distribution is normal, we took a small aliquot for NGS analysis and split the remaining barcoded PATC124 cells into two equal parts for two parallel experiments in vivo (Fig. 4C and fig. S4D). One part was transplanted orthotopically to evaluate the intrinsic metastatic potential of independent subclones ("metastatic potential study"; $n = 3$ animals); the other part was assessed in a serial transplantation assay to determine the LTSR of the tumor subclones ["LTSR study"; animals receiving initial cell injection and tumor retransplanted for a total of three subsequent rounds of serial transplantation (F1 to F4, $n = 3$ animals per transplantation)] (fig. S4D). The primary tumor, liver, and lung from animals in the metastatic potential study, as well as the tumors from the fourth transplantation (F4) of the LTSR study, were harvested and submitted for NGS. Barcode analysis of tissues from both studies identified 333 lineages that were both present in either the liver or lungs in the metastatic potential study and present in the F4 primary tumors in the LTSR study (Fig. 4D). This small number of subclones—less than

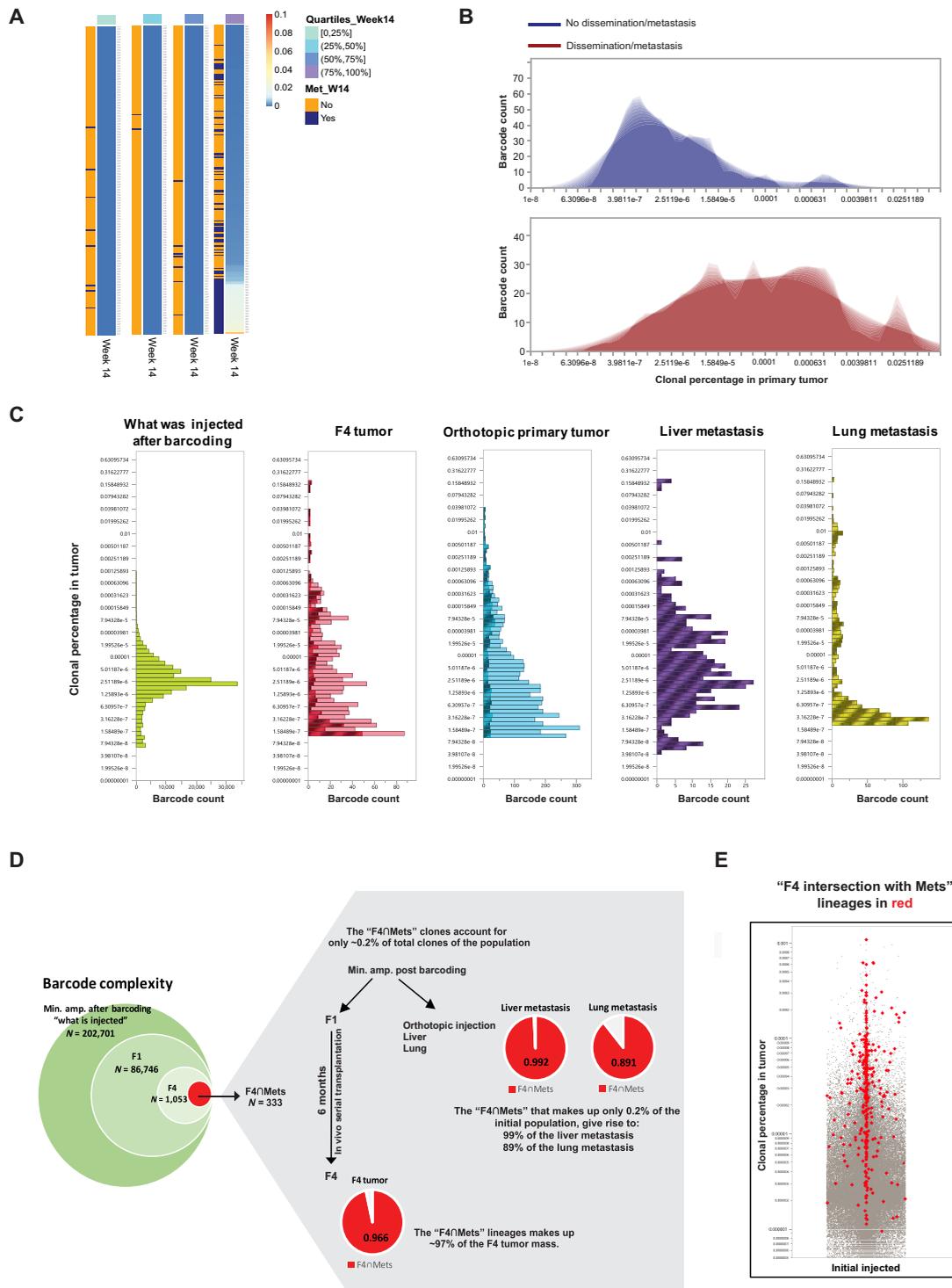


Fig. 4. Intrinsic LTSR is highly correlated with metastatic potential. (A) Heatmap of average subclonal abundance (percentage) in the PATC124 primary tumors with metastasis status (yes, >200 tumor cells/no, <200 tumor cells) at week 14. Quartiles are as defined in Fig. 2. Dissemination of each subclone is indicated on the left of each heatmap in quartiles (yellow, no dissemination; blue, dissemination). **(B)** Average clonal abundance in the PATC124 primary tumors (percentage) at week 14 for metastatic clones (top) and nonmetastatic clones (bottom). **(C)** Barcode distribution of different PATC124 samples in the LTSR study and metastatic potential study. The shaded area represents the distribution of the distribution of 333 lineages that were present both in the F4 tumors in the LTSR study and in the liver and/or lung metastases collected in the metastatic potential study (F4∩Mets). **(D)** Distribution of PATC124 lineages with LTSR and metastatic potential. Left, Venn diagram of shrinking complexity of barcodes from original sample through F4 and for F4∩Mets (red); right, relative tumor mass represented by F4∩Mets in the F4 tumor, liver, and lung metastasis. **(E)** Distribution of 202,701 barcodes (percent of tumor) in the initial sample for the metastatic potential and LTSR studies (gray). The 333 F4∩Mets are shown in red.

0.2% of the implanted subclones (202,701) in the original prepared sample—sustained ~96% of the F4 primary tumors and gave rise to ~99 and ~89% of liver and lung metastases, respectively (Fig. 4D). We confirmed that the 333 clones endowed with high LTSR and high dissemination potential were not overrepresented (relative abundance) in the initial population of barcoded cells (Fig. 4E), as well as in primary tumors from the metastatic potential study and F4 from the LTSR study (Fig. 4C). Together, our data demonstrate that a very small number of subclones is endowed with robust dissemination potential as well as LTSR, and indicate that intrinsic properties of cells rather than solely tumor representation are responsible for dissemination. Of note, only a fraction (~31%) of tumor cells endowed with LTSR are able to disseminate.

NMDA signaling is a key determinant of metastasis

Leveraging the capabilities of the CRT platform to isolate subpopulations of naïve cells of interest based on their behavior *in vivo* (30), we selected six subclones with distinct metastatic potential identified in the longitudinal study of PATC124 for further analysis. Three subclones each were selected, which displayed metastatic behavior (pro-met) or which were not detected in the peripheral tissues (non-met). We performed whole-exome sequencing (WES) for each of the six clones and analyzed their mutational landscape and CNV profiles. All of the subclones harbored mutations in *KRAS* and *TP53*, which are characteristic of the PDAC genomic landscape. Other mutations were also identified in all six lineages, including mutations in the DNA damage checkpoint genes, *ATM* and *ATR* genes. However, overall, the six lineages were characterized by heterogeneous mutational and CNV profiles (fig. S5, A and B), and we were not able to associate a particular genetic trait with the metastatic phenotype. We cannot exclude that this could be due to limitations in the sequencing approach as well as the low number of samples profiled.

To determine whether gene expression patterns may drive metastasis, we applied supervised clustering to bulk transcriptomic analysis of the six subclones, which identified 387 deregulated genes in the pro-met lineages compared to the non-met lineages (Fig. 5A, left). Ingenuity Pathway Analysis (IPA) showed that genes related to glutamatergic synaptic neural transmission, such as the *N*-methyl-D-aspartate (NMDA) receptor and α -amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid (AMPA) receptor, as well as genes involved in neurovascular coupling signaling were enriched in pro-met subclones compared to the non-met subclones (Fig. 5A, right). Because the NMDA signaling has been demonstrated to mediate tumor invasion in models of pancreatic, neuroendocrine, and breast tumors (36–38), in addition to the availability of US Food and Drug Administration (FDA)–approved inhibitor drugs, we prioritized genes in this pathway for functional characterization in the two most distinct pro-met subclones (M-1 and M-2) and two non-met subclones (NM-2 and NM-3). After assessing that all four clones were tumorigenic and engrafted as monoclonal tumors upon orthotopic injection, barring some expected volumetric variability (fig. S5C), their metastatic behavior was validated, and it matched the documented behavior in the longitudinal oCRT study in each case (Fig. 5B). To further corroborate this observation, we mixed the two metastatic clones (M-1 and M-2) and the two nonmetastatic clones (NM-2 and NM-3) in a 1:1:1:1 ratio (25% each) and performed longitudinal dynamic analysis. In the orthotopic tumors, we found that, while the ratio of all four clones was equal at injection, the two metastatic clones soon became dominant clones *in vivo*,

thus demonstrating higher fitness, and outcompete the nonmetastatic clones that had barely expanded (Fig. 5C). When clones are combined, the metastatic patterns in liver and lung confirmed our previous findings that the two metastatic clones are highly metastatic and have preferred organotropism to the liver instead of lung. Although detected in the lung, the number of cells remained two magnitudes lower than those detected in the liver, where metastatic clones formed large nodes (Fig. 5C). On the contrary, the two non-metastatic clones were barely detected in secondary organs (Fig. 5C). Regarding CTCs, our data showed that the metastatic, but not the nonmetastatic, clones could be detected in circulation (Fig. 5C).

We then validated the transcriptomic data by reverse transcription quantitative polymerase chain reaction (PCR), confirming that the NMDA receptors *GRIN2A* (glutamate ionotropic receptor NMDA type subunit 2A), *GRIA2* (glutamate ionotropic receptor AMPA type subunit 2), and *CaMK4* (calmodulin-dependent protein kinase type IV) were significantly up-regulated in the pro-met subclones compared to the non-met subclones (Fig. 5D). In addition, compared to non-met, the pro-met lineages were characterized by a higher basal intracellular calcium concentration (Fig. 5E) as well as a more invasive phenotype (Fig. 5F), and the invasion is glutamate receptor pathway-dependent (fig. S5E), strongly suggesting a functional role for glutamatergic signaling in metastatic subclones.

To test whether inhibition of NMDA receptors would affect the invasive phenotype, we performed an *in vitro* invasion assay with both pro-met subclones treated or not with memantine, a noncompetitive NMDA receptor antagonist that is FDA-approved to treat neurological disorders. Cells grown in medium containing 25, 50, or 100 μ M memantine for 24 hours had statistically significantly fewer invading cells compared to no-memantine controls (Fig. 5G and fig. S5, D and E). To test the effect of memantine on tumor dissemination *in vivo*, we established orthotopic prometastatic bicultural tumors mixing with equal ratio (1:1) M-1 and M-2 clones, and then randomized animals to receive intraperitoneal injection of vehicle or memantine (10 mg/kg, once daily) for 5 days on, 2 days off for 2 months. At the end of drug treatment, the livers were collected and processed for NGS barcode quantification using a spike-in scale control to quantify the number of metastatic cells. Animals that received memantine treatment had fewer metastatic burden compared to animals treated with vehicle (Fig. 5H), indicating that memantine treatment can decrease dissemination of tumor subclones that sustain metastases. However, when we looked at the orthotopic tumor weights from the control and memantine-treated group, the data suggested that memantine treatment did inhibit the growth of primary tumors compared to the control (fig. S5F); thus, we cannot rule out the possibility that memantine had suppressed tumor cell metastasis/dissemination by decreasing the fitness of metastatic clones in the primary tumors, although it inhibits the invasion of tumor cells *in vitro*. Either way, the intrinsic up-regulation of the NMDA pathway may be an actionable biomarker and/or drug target for tumor subclones with a prometastasis phenotype.

Prometastasis signatures preexist in primary tumors and predict patient survival

From our transcriptomic analysis of the pro-met and the non-met isogenic subclones, we established a 200-gene signature associated with the pro-met phenotype. First, we confirmed that the signature can identify a specific cell subset within human tumors. To do so, we matched the metastasis gene signature to the single-cell RNA

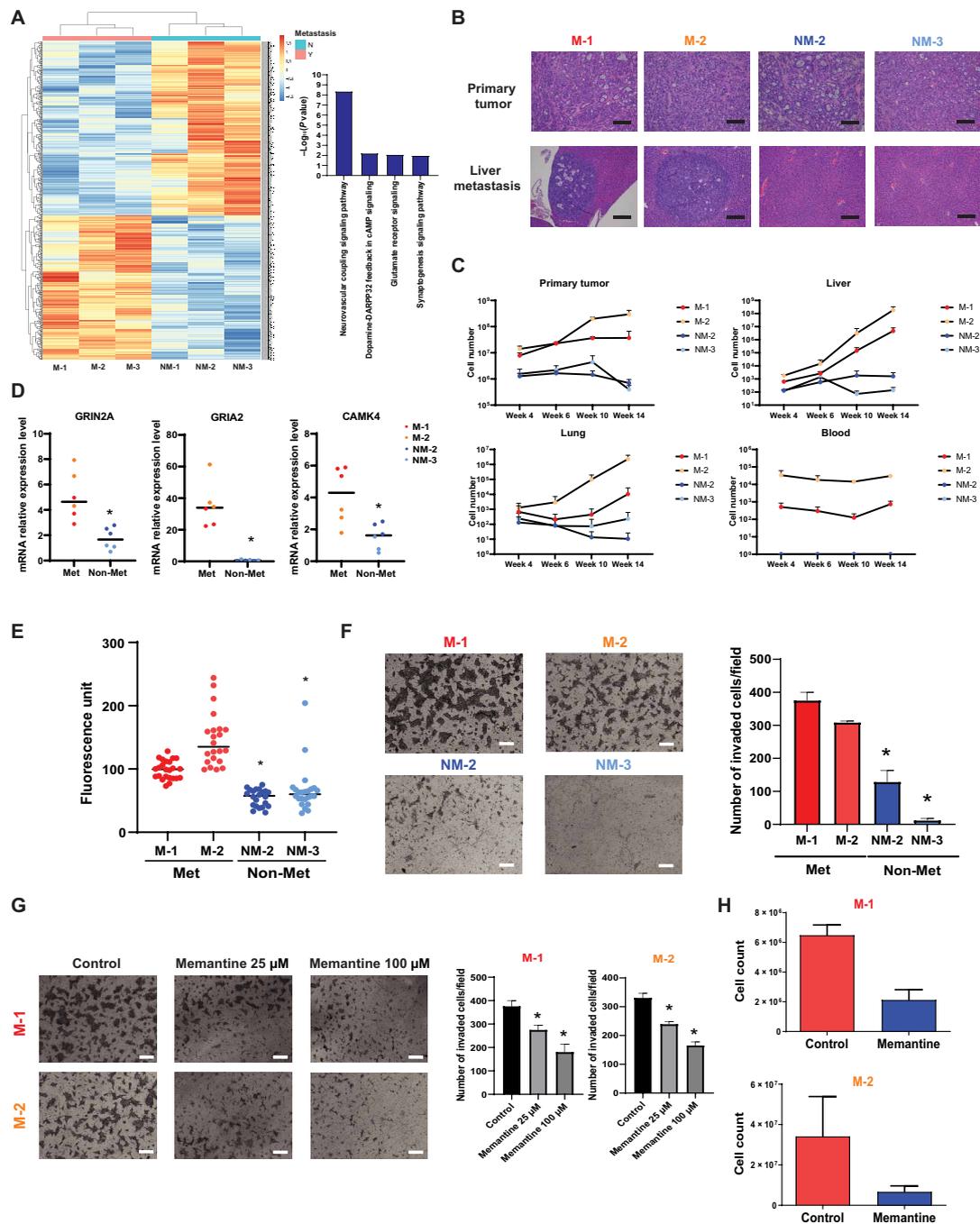


Fig. 5. Transcriptome profiling of metastatic subclones identifies actionable vulnerabilities in metastasis. (A) Heatmap of differential gene expression of isogenic clones with distinct metastatic potential as determined in longitudinal PATC124 oCRT studies (left). IPA analysis of pro-met (M-1 to M-3) and non-met (NM-1 to NM-3) clones identifies enrichment of glutamate receptors and neurovascular coupling signaling pathways in pro-met lineages. (B) Representative histology images from primary tumor and liver metastasis derived from PATC124 monoclonal orthotopic tumors ~2 months after injection. Scale bar, 200 μm . (C) Equal ratio of two metastatic (M-1 and M-2) and two nonmetastatic (NM-2 and NM-3) PATC124 clones was mixed followed by orthotopic transplantation. Primary tumors, livers, lungs, and blood were harvested at different time points, and barcode composition was assessed by sequencing. Cell number at each time point is converted on the basis of spike-in scales. (D) qRT-PCR validation of mRNA expression level of *GRIN2A*, *GRIA2*, and *CAMK4* in two pro-met and two non-met PATC124 subclones. Representative data are shown of two independent experiments. Statistical significance was assessed with Student's *t* test. (E) Basal intracellular calcium concentration of PATC124 clonal lineages selected on the basis of differential metastatic behavior. Statistical significance was assessed using one-way analysis of variance (ANOVA). (F and G) Representative images of invasion assay indicating the basal invasion potential of two pro-met and two non-met PATC124 subclones (F) and the effect of the noncompetitive NMDA antagonist 25 or 100 μM memantine on invasive potential of cells (G). Quantification of invaded cells per field is shown as means \pm SD of two independent experiments. Statistical significance was assessed using one-way ANOVA. Scale bar, 200 μm . (H) Effect of memantine (10 mg/kg) on liver metastasis of two metastatic isolated PATC124 clones. Data are presented as means \pm SD in number of barcoded cells detected in the liver at 11 weeks ($n = 3$). Barcode reads were converted to cell number using a spike-in scale.

sequencing (RNA-seq) data of the parental tumor cell line (PATC124) used to establish oCRTs and found that ~13% of the cells (453 of 3397) are enriched in the metastasis signature, suggesting that the cells that sustain metastasis preexist and can be identified in the primary tumor cell line (fig. S6A). To further demonstrate the relevance of these findings, we tested the ability of the signature in identifying cells that sustain metastasis in publicly available single-cell and spatial RNA-seq datasets of patients with pancreatic cancer (39). While only a small sample size was available for the analysis ($n = 3$), the only tumor that did not present histopathological signs of invasion also did not display any enrichment in metastatic cells at the single-cell level (Fig. 6, A to C, left, and fig. S6, B to D). Only 4.5% of the tumor cells (91 of 2019) have been identified as metastatic (Fig. 6A), against 19.5 and 31.8% of tumors that histopathologically showed lymphovascular invasion (203 of 1039 and 448 of 1408, respectively) (Fig. 6, B and C). Of note, when we used the signature to analyze spatial datasets from the same tumors, we found that metastatic cells formed distinct clusters in the two patients with invasive lesions. In contrast, in the patient with the noninvasive tumor, the limited number of metastatic cells appeared scattered throughout the neoplastic lesion (Fig. 6, A to C, right).

Next, to assess whether the detection of metastatic cells may have prognostic clinical relevance, we integrated RNA-seq data from pancreatic cancer The Cancer Genome Atlas (TCGA) cohort (comprising primary tumor samples, $n = 178$) and the PRISM cohort (which includes metastasis samples, $n = 45$) (40) (total $N = 223$) and used the metastasis signature to segregate patients. We applied multivariate Cox proportional hazard model on patient survival data with pro-met/non-met signatures and age. We found that the pro-met and non-met signatures were associated with significantly poor prognosis (more hazardous) or better prognosis (less hazardous), respectively, with age being insignificant in terms of overall survival (Fig. 6D). To further investigate whether pro-met signature could predict patient survival, we defined each sample as met_dominant or non-met_dominant group based on the enrichment score using single-sample gene set enrichment analysis (ssGSEA) and performed a Kaplan-Meier survival analysis. As reported in Fig. 6E, the met_dominant group (met-up) showed significant worse survival compared to the non-met_dominant group (met-down), confirming the potential of our metastatic signature to identify patients at risk of developing metastatic disease.

These data underscore the clinical significance of our research and illustrate how transcriptional features from isolated clones can be used to predict patient survival by detecting preexisting metastatic cells within primary tumors.

DISCUSSION

Clonal competition is the major process that subtends tumor evolution. As a tumor develops and expands, cell-intrinsic factors determine the fitness of tumor clones to the changing microenvironment and define subclonal tumor composition over time. Therefore, tumors can be viewed as complex ecosystems where clonal lineages compete with one another as they adapt to and coevolve with stromal cells and in response to external perturbations. Tumor evolution science has made progress in deconvoluting tumor progression using multiomics tumor profiling at various stages of the disease; however, the field lacks adequate *in vivo* experimental models of cancer for unperurbed, longitudinal lineage tracing with high resolution and representative of the heterogeneity of human tumors. To address this need

for robust experimental systems to study tumor evolution, we established oCRTs, a high-resolution model of heterogeneous, patient-derived tumor models to quantitatively dissect clonal dynamics during tumor expansion without perturbing the tumor ecosystem.

Over a 3-month longitudinal study, we observed an astonishing and unexpected degree of oscillation in the fitness of subclonal populations. Within the same tumor, we documented subclonal lineages that (i) gradually decreased or increased over time, (ii) transiently increased followed by a gradually decreasing, (iii) were mostly dormant followed by aggressive outgrowth, (iv) displayed a sigmoidal progression of increase-decrease-increase in representation, (v) increased for most of the follow-up period followed by a sudden decrease, and (vi) initially decreased followed by a gradual increase in representation. The net result of these concurrent, complex clonal behaviors is a high degree of ACD where, at any time, certain clonal lineages experience a sudden drop in fitness and are replaced by other clones with relatively higher fitness that expand and overtake the tumor.

This dynamic behavior was observed in our system devoid of any external evolutionary pressure and represents solely behaviors driven by the tumor microenvironment. It is likely that space, oxygen, and nutrient limitations are major drivers of ACD. Our data indicate that, as the tumor expands, subclones compete for resources, and the most-fit lineages expand and dominate the tumor's bulk. At some point, the larger tumor encounters a new stressor in its environment, such as hypoxia, nutrient deprivation, or physical forces (e.g., space limitations), resulting in a shift in the relative fitness and initiating an involution-expansion cycle of the subclones until a new equilibrium is achieved. This process can repeat multiple times as the microenvironmental conditions change and can be confined to specific regions of the tumor only. Notably, we observed ACD even when the tumor size remained relatively constant, as well as at late time points when the tumor decreased in mass due to extensive necrosis. Our studies in human-derived xenografts were conducted in immunodeficient animals; however, when repeated in KPC-derived models, we found similar oscillating clonal dynamics, confirming that ACD is also a property of tumors that expand in intact immune microenvironments.

Our studies showed that a large fraction (30 to 90%) of the lineages that sustain primary tumor growth can be found at a secondary site at some point during tumor progression as well as in tumors that do not develop macro-metastasis. This was an unexpected finding on the basis of the metastatic cascade model, which predicts very high attrition of tumor cells that escape the primary tumor (41, 42). The finding that a large fraction of tumor cells can reach a secondary organ is probably explained by the high sensitivity of our approach, which is able to detect as few as one to three barcoded cells in the entire liver. Although many subclones could seed distant organs, we confirmed that only a few of them were capable of colonization to form a discernable metastatic node. Rather, most of the lineages that reached secondary sites displayed an abortive phenotype, undergoing a transient expansion before disappearing at later time points. This previously undescribed behavior was observed in both the liver and the lung and can be influenced by the intrinsic organotropism of patient-derived models under study. This suggests that tumor cell seeding is probably a generalized nonspecific behavior of disseminated cells, whereas the microenvironment of secondary sites is the critical determinant of the success of tumor cell expansion, as postulated by "seed and soil" theory of metastasization (6). This notion is further supported by the fact that matched barcoded lineages frequently

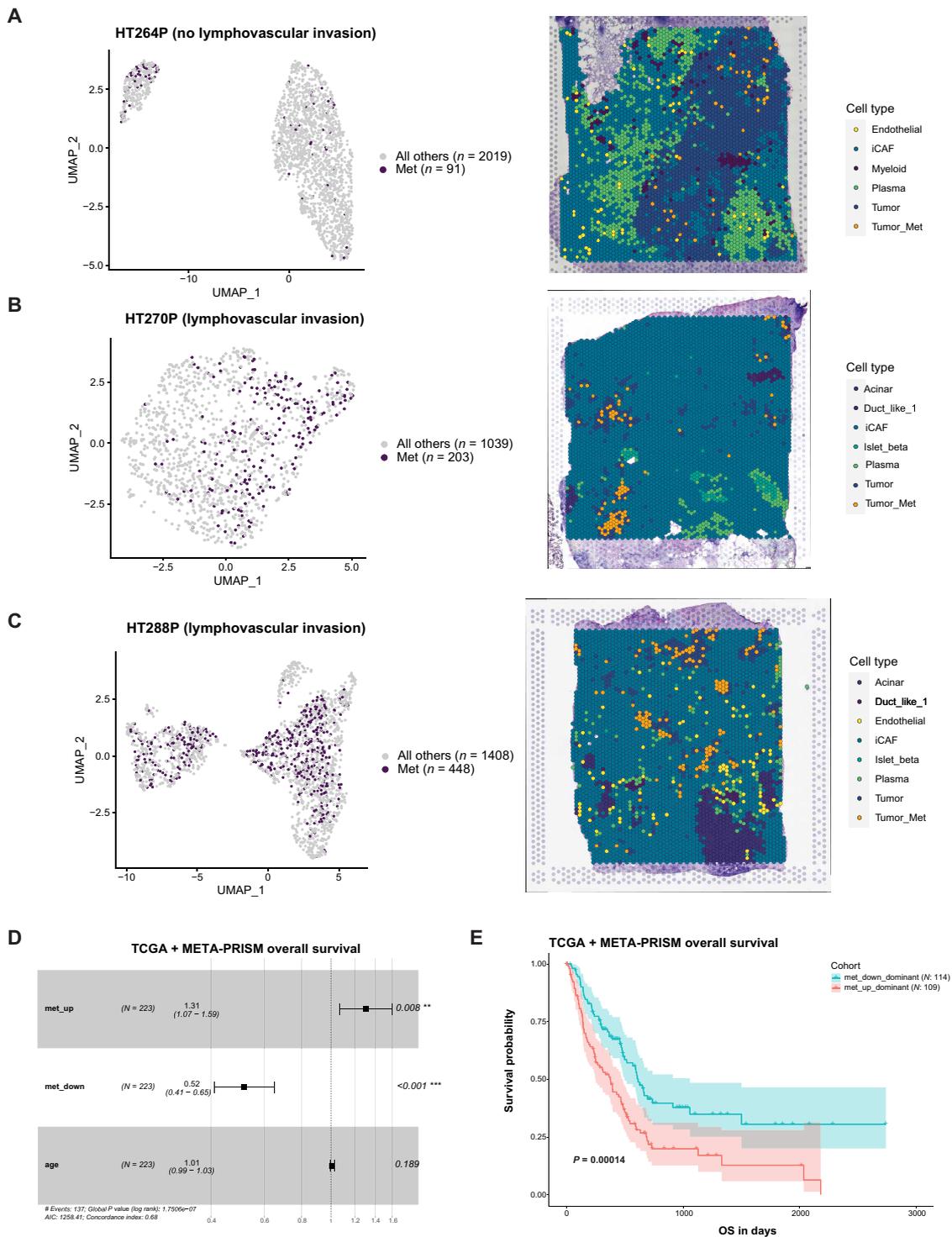


Fig. 6. Molecular signatures associated with metastatic subclones predict survival outcomes for patients with PDAC. (A to C) Single-cell and spatial transcriptomics of the same patient samples without (A) or with [(B) and (C)] lymphovascular invasion. The left panels show Uniform Manifold Approximation and Projection (UMAP) of tumor cells identified from single-cell sequencing as in fig. S5 (B to D). Cells with a significant closer distance ($P < 0.05$) to met-up signature than met-down signature are highlighted as purple color (Met), while others are labeled as “All others.” The right panels show cell typing of spatial transcriptomics to identify tumor spots. Tumor spot predicted to be prometastatic are highlighted with orange color and labeled as Tumor_Met. (D) Forest plot of Cox proportional hazard models of overall survival for met-up and met-down signatures in the human PDAC cohort [primary tumor (TCGA) + metastasis (PRISM)]. (E) Human PDAC cohort [primary tumor (TCGA) + metastasis (PRISM)] was dichotomized into met_up_dominant and met_down_dominant groups, and survival was displayed as Kaplan-Meier plots.

displayed markedly differential fitness at the primary tumor site versus in the periphery. In support of findings recently reported by other authors (43), we found that metastases may originate as polyclonal and can subsequently evolve as monoclonal/oligoclonal, when one or few clones become dominant probably due to selection imposed by the microenvironment.

We noted an interesting correlation between the outgrowth metastases and ACD in the paired primary tumor. Specifically, the first notable shift in clonal dominance was usually observed at week 6, which is also when we detected expansion of metastatic subclones in the liver. Although strictly hypothetical at this time, we are intrigued by the idea that the coincident changes in growth at the two sites might be related to microenvironmental changes in the primary tumor. For example, events that trigger ACD and clonal reshuffling in the primary tumor could mobilize tumor cells, or they could promote diffusible signals that could be received and responded to by tumor cells at secondary sites.

We also determined that the abundance of a lineage within the primary tumor is highly correlated with metastatic dissemination. This finding was not unexpected (44), and multiple evidences have been previously reported (15, 45, 46), but the resolution of our oCRT platform also allowed us to determine that there are many exceptions to this rule. We observed multiple subclones that were highly represented in the primary tumor that were not found outside the primary site, as well as lowly represented subclones that were able to expand at secondary sites. These exceptions indicate that cell-intrinsic factors also contribute to the metastatic phenotype; what these cell-intrinsic factors are and whether they are always present in metastatic subclones that were also highly abundant remain to be determined. In this vein, further characterization of metastatic subclones revealed that they are endowed with LTSR, which represents an unbiased demonstration of a cell-intrinsic property that could influence metastatic potential, independently from abundance in the primary tumor.

In summary, the oCRT approach represents a technological advancement in modeling human disease and revealed a new level of subclonal complexity intrinsic within tumors. oCRTs enable longitudinal studies to quantify the subclonal heterogeneity over time, revealing unexpectedly heterogeneous growth dynamics and frequent ACD. This study has also emphasized the extent of functional diversity that occurs naturally during tumor expansion and demonstrates that tumor evolution can be driven by cell-intrinsic properties that promote fitness of some lineages in harsh conditions to ensure tumor survival.

METHODS

Generation of CRTs

Patient-derived samples

Tumor-derived samples were obtained at MD Anderson Cancer center with consent from (i) a 62-year-old Caucasian male patient (PATC124), (ii) a 64-year-old Caucasian male patient (PATC69), and (iii) a 71-year-old male patient (PATC108) who underwent pancreaticoduodenectomy for T3N0M0 pancreatic cancer. Approval of all patient-derived sample collection methods can be found in Institutional Review Board–approved protocol LAB07-0854 chaired by J. Fleming [University of Texas MD Anderson Cancer Center (UTMDACC)].

PDX cell isolation and culture

Early passage xenografts (F1) from primary human PDAC (31, 47) were harvested and processed as previously described (47). Briefly, tumors were minced into very small pieces under sterile conditions followed by mechanical dissociation with the Human Tumor Dissociation Kit (Miltenyi Biotec) and enzymatic digestion. After digestion, single isolated cells were seeded at high confluency on collagen IV–coated plates (Corning) in Dulbecco's modified Eagle's medium (DMEM)/F12 medium (Gibco) supplemented with 10% fetal bovine serum (FBS) (Gibco), 1% bovine serum albumin (Fisher Scientific), 0.5 μ M hydrocortisone (Sigma-Aldrich), 10 mM HEPES (Invitrogen), cholera toxin (100 ng/ml, Sigma-Aldrich), insulin-transferrin-selenium (5 ml/liter, BD), and penicillin-streptomycin (100 μ g/ml, Gibco). To remove stromal cell contamination, brief trypsinization (0.25% trypsin-EDTA, Gibco) was periodically performed to eliminate fibroblasts. The purity of the PDX-derived human cell culture was confirmed over time by flow cytometry through the evaluation of HLA-ABC and mouse H-2Kd histocompatibility complex antigens (Becton Dickinson). Final isolated human cells were maintained in culture for a maximum of couple of passages before being switched to DMEM plus 10% FBS and transduced with the barcode library.

KPC pancreatic cancer mouse model

Mice were housed in a pathogen-free facility at the UTMDACC. All manipulations were performed under the Institutional Animal Care and Use Committee–approved protocol (00001843-RN01). *p48-Cre* mice (48), *LSL-Kras^{G12D}* (The Jackson Laboratory, no. 008179), and *Trp53^{R172H}* (provided by G. Lozano, UTMDACC) were crossed to generate the KPC pancreatic cancer mouse model. Spontaneous pancreatic tumors were entirely harvested and digested, and KPC cells were derived. In brief, entire tumors were minced into small pieces with sterile blades and incubated at 37°C for 45 min with collagenase IV (Gibco)–dispase II (Roche), 2 mg/ml for enzymatic digestion. Cells were then centrifuged and further digested with 0.25% trypsin (Gibco) for 5 min at 37°C to obtain a single-cell suspension. After being maintained and expanded in DMEM (Gibco) supplemented with 10% FBS (Gibco) and penicillin-streptomycin (100 μ g/ml, Gibco), tumor cells were enriched with anti-EpCAM antibody via flow cytometry sorter to establish KPC cells.

Barcoding and generation of CRTs

The CloneTracker 50M Lentiviral Barcode Library was purchased from Collecta. Established PDX-derived tumor cells or KPC cells at 70% confluence were suspended in medium containing a very low MOI (<0.2) of barcode library viral particles and infected overnight in the presence of polybrene (8 μ g/ml) in 245-mm square plates (Corning). After 48 hours, the cells were detached; infection was confirmed, and its efficiency was evaluated by flow cytometry [percentage of red fluorescent protein (RFP)–positive cells]. The cells were then plated in media with puromycin at an optimal concentration to eliminate uninfected cells. When confluent, barcoded cells were detached, washed, and plated in two 245-mm square plates (passage 1). The medium was changed every 2 to 3 days, and the cells were allowed to become confluent. When approximately 80% confluency was reached, the passage 1 plates were detached, washed, and pooled before being split into two equal parts: one stored for NGS for barcode composition, and the other part was plated in two 245-mm square plates (passage 2). When the cells became confluent again, passage 2 cells were detached, washed, pooled, and split into two equal parts: one cryopreserved in BamBanker (Wako Chemicals), and the other part was

plated in two 245-mm square plates (passage 3). The procedure was repeated, with “odd passages” being snap-frozen for sequencing, and all “even passages” being cryopreserved for reculturing and high-throughput clone isolation.

Quantitative scale

To perform quantitative analysis in metastasis, a five-log cell spike in control—a conversion scale of known cell counts carrying unique barcodes distinct from the clonal tracking library—was added to metastatic samples before DNA extraction (see Table 1 for unique barcode sequences). In brief, 293T cells were infected individually with lentiviral particles containing unique barcodes that are not present in the lineage tracing library but share the same plasmid backbone (pRSI16) at very low MOI (<0.1). After confirmation and evaluation of infection by flow cytometry (percentage of RFP-positive cells), infected cells were selected in puromycin and expanded. Serial dilutions of each unique barcoded cells were performed and mixed with each other and subsequently aliquoted in a multitude of 1.5-ml vials to generate a “scale.” For quantification of metastases at distal organs, each scale contained barcoded cells represented as follows: BC.3063.3418 1×10^1 , BC.4451.4842 1×10^2 , BC.4858.5013 1×10^3 , BC.5022.5430 1×10^4 , BC.5993.6943 1×10^5 . After sequencing, reads generated by unique barcoded cells of the scale could be used to normalize reads generated by samples, thus enabling conversion of reads to more interpretable numbers of cells.

In vivo transplantation and treatments

All studies were performed in *NOD.Cg-Prkdc^{scid} Il2rgt^{m1Wjl}/SzJ* (NSG) (PDX cell lines) or C57BL/6J (KPC cell line) mice purchased from The Jackson Laboratory.

Longitudinal oCRT cohort generation

Fifty mice were injected orthotopically with cells derived from stabilized cultures. In brief, a 1- to 1.5-cm incision in the upper left abdominal quadrant was made to expose the pancreas of recipient mice. Cells (3×10^6 per mouse) were suspended in 1:1 DMEM/Matrigel (BD) followed by injection into the pancreas. After injection, the mouse abdominal wall was closed with absorbable sutures (Ethicon), while the skin was closed with metal clips (Fine Science Tools). The mice were then sacrificed at indicated time points, and harvested primary tumors and tissues were processed for next-generation sequencing. For KPC experiments, 1.5×10^6 cells per mouse were orthotopically injected.

Orthotopic LTSR experiments

After barcoding, passage 4 cells were injected subcutaneously or orthotopically in three NSG mice. For subcutaneous tumors, 5×10^6 cells were injected. When tumors reached 1 cm in size (~1.5 months), the tumors were collected, pooled, and digested to single cells.

One-half of the recovered cells was reinjected in a new round of three animals, while the other half was used for sequencing. Tumors were passaged for three times in vivo. For orthotopic tumors, 3×10^6 cells were injected. Primary tumors and distant organs were collected after 3 months of transplantation followed by sequencing for barcode information.

Isogenic single clone in vivo tumorigenicity assay

Isogenic single clones were transplanted orthotopically with 2×10^6 cells. Primary tumors and distant organs were harvested 2 months after transplantation for histology.

In vivo tumor reconstitution assay for memantine treatment

Two metastatic clones were mixed equally with the same number of cells. A total of 3 million cells were then orthotopically transplanted into the pancreas of recipient mice (NSG). After injection, the tumors were allowed to expand and grow for 3 weeks before treatment. For memantine treatment, mice received memantine (10 mg/kg) daily via intraperitoneal injection for the 3 weeks after transplantation and followed by a 5 days on/2 days off cycle for another 8 weeks. The livers were then harvested and submitted to NGS for barcode composition.

In vivo tumor reconstitution assay for longitudinal dynamics

Two metastatic clones and two nonmetastatic clones were mixed equally (1:1) with the same number of cells. A total of 3×10^6 cells were then orthotopically transplanted into the pancreas of recipient mice (NSG). After injection, the tumors were sacrificed at different time points, and primary tumors, livers, lungs, and blood were harvested and submitted to NGS for barcode composition.

CRT sample processing and analysis

Genomic DNA extraction and PCR for NGS library production

Genomic DNA extraction and PCR for NGS library production were adapted from a published protocol with minor modifications (47). In brief, scale for cell number quantification was added to distant organ sample before processing. For cell pellets from in vitro samples, DNA was extracted with DNeasy Blood & Tissue Kits (Qiagen) according to the manufacturer’s protocol. For tissues, frozen tumors/tissues were minced and suspended in Buffer P1 (Qiagen, 1 ml of buffer/100 mg of tumor) supplemented with ribonuclease A (100 µg/ml, Thermo Fisher Scientific) followed by disassociation with a gentleMACS homogenizer (Miltenyi Biotec). Samples were then transferred to a 15-ml polypropylene tube (Falcon) and lysed by adding $1/20$ volume of 10% SDS (Promega) and $1/20$ volume of proteinase K (Qiagen) followed by incubation at 56°C for 20 min. Genomic DNA was sheared by passing the lysate 10 to 15 times through a 22-gauge syringe needle followed by phenol:chloroform:isoamyl alcohol (25:24:1, pH 8.0, Sigma-Aldrich) and chloroform:isoamyl alcohol (24:1, Sigma-Aldrich) purification. The final aqueous phase was transferred to a new tube, and 0.1 volume of 3 M NaOAc (Sigma-Aldrich) and 0.8 volume of isopropanol (Fisher Scientific) were added to precipitate genomic DNA. DNA was pelleted by centrifuging at 14,000 rpm for 20 min and subsequently washed once in 70% ethanol (Fisher Scientific) and centrifuged again for 5 min at 14,000 rpm. The DNA pellet was air-dried and dissolved overnight in UltraPure distilled water (Thermo Fisher Scientific), and final DNA concentration was assessed by NanoDrop 2000 (Thermo Fisher Scientific) quantification.

For the generation of NGS libraries, barcodes were amplified with Titanium Taq DNA polymerase (Clontech-Takara) through two rounds of PCR. In brief, the first PCRs were performed for 16 cycles

Table 1. Sequences of scale barcodes.

Name	Sequence
BC.3063.3418	CATGCACACAGTACACTGTTCCGGTGTGTGCAACTGAC
BC.4451.4842	CAACGTTGTGACTGTGACTTCCGGTGTGTCAGTACTGTGAC
BC.4858.5013	GGTGCACAGTACTGACTGTTCCGGTGTGGTCAACTGTGAC
BC.5022.5430	GTCATGGTCAACTGTGACTTCCGGTGTGCAACTGTGTGACTG
BC.5993.6943	TGGTGTGCATGTGACTGTTCCGCACATGACCACTGTGTGAC

with 13K_R2 (5'-AGTAGCGTGAAGAGCAGAGAA-3') and FHTS3 (5'-TCGGATTCAAGCAAAGACGGCATA-3'). The second PCRs were performed for 12 cycles with P5_NR2 (5'-AATGATACGGCGA-CCACCGAGACGAGCACCGACAACAACGCAGA-3') and Gx1_Bp (5'-TCAAGCAGAAGACGGCATAACGAAGACA-3'). PCR amplification products were analyzed by agarose gel electrophoresis (2.5%, Lonza) for the expected 279-base pair (bp) size and then extracted from the agarose gel with the QIAquick Gel Purification kit (Qiagen). The amount of purified PCR product was quantified using the High Sensitivity D1000 ScreenTape and Agilent 4200 TapeStation system (Agilent Technologies). Barcode representation was measured by NGS on an Illumina HiSeq2000 with 13K_Seq (5'-AGAGGTTT-AGAGTTCTACAGTCCGAA-3') as the sequencing primer.

Tissue multiregional sequencing via LCM

Paired primary tumor and liver from the same animal were harvested, fixed, and embedded followed by tissue sectioning. Six-micrometer slices were mounted on membrane slides (MMI) for hematoxylin and eosin staining to identify regions of interest. For the primary tumor, 41 regions were captured via laser capture microdissection (LCM) (MMI Cellcut) with necrotic area excluded. For the liver, 12 visible metastatic lesions were captured via LCM. Each individual region or lesion was collected in an isolation cap (MMI) and then transferred to a 96-well PCR plate. For each well, 16 μ l of lysis buffer [10 mM tris-HCl (pH 7.5), 10 mM NaCl, 10 mM MgCl₂, 0.19% NP-40, and 2 μ g of proteinase K] was added and incubated at 50°C for 3 hours to extract genomic DNA. The plate was then incubated at 95°C for 30 min to inactivate proteinase K. For the mixture containing Titanium Taq DNA polymerase (Clontech-Takara), individual forward primer with unique index Ion-A-BC-GexBP1 (5'-CCATCTCATCCCTGCGTGTCTCCGACT-CAGXXXXXXXXXXGATTCAAGCAGAAGACGGCATA-GAAGACAG-3'), X bases refers to Ion Xpress 96 adaptor barcode sequence), and a common reverse primer GexSeqS-IonP1 (5'-CCA-CTACGCCTCCGCTTCTCTCTATGGCAGTCGGTGATT-GAGGTTTCAGAGTTCTACAGTC-3'), both at 600 nM final concentration, were added to each well and amplified with the following cycling profile: 95°C 3 min, 6 cycles of 95°C 30 s, 56°C 10 s, 72°C 10 s, followed by 30 cycles of 95°C 30 s, 62°C 10 s, 72°C 10 s, and then final extension at 72°C for 2 min and 4°C hold. Ten microliters of PCR products from each well were pooled together (total 960 μ l), and 500 μ l of the pooled product was used for size selection with 1X AMPure XP (Beckman Coulter). The final library product was quantified with the High Sensitivity D1000 ScreenTape and Agilent 4200 TapeStation system (Agilent Technologies) followed by sequencing with the Ion Torrent Proton NGS platform (Thermo Fisher Scientific).

Barcode data alignment

Barcode sequencing data were processed and aligned through a custom pipeline. Cutadapt (<https://doi.org/10.14806/ej.17.1.200>) was used to trim adaptor sequences from the reads. After trimming, we used Bowtie (<https://doi.org/10.1186/gb-2009-10-3-r25>) to align the reads to the barcode library while allowing one mismatch. SAMtools (<https://doi.org/10.1093/bioinformatics/btp352>) was then used to extract read counts of the aligned barcodes.

Conversion from barcode read to cell number

For quantitative analysis and to normalize between tumor samples, we spiked in the conversion scale of known cell numbers containing a unique set of barcodes during sample processing. For each clonal tracking barcode, we divided the reads by one of the scale's read

numbers and multiplied the value by the cell number corresponding to that specific scale

Estimated cell number = (Barcode read \times scale cell number spiked in)/(scale read)

*The Conversion Scale spike-in equals to 10,000 cells for liver, lung, blood, and ascites samples.

Longitudinal CRT data analysis

Data cleaning and filtering criteria

Missing data were removed, and technical errors were adjusted.

For the resampling bootstrap test: Barcodes with nonzero reads at both injection and week 4 in each mouse were first selected. Among those, barcodes that contributed to the cumulative abundance percentage of 99.9% at any time point were kept, leading to a total of $N = 3751$ barcodes for the following analysis.

For barcode dynamic clustering analysis in primary tumors: A set of common barcodes present in all primary tumors across all time points was selected ($N = 934$) for the following analysis.

For barcode dynamic clustering analysis in the metastatic tumors: Within the $N = 934$ barcodes derived from primary tumors, barcodes with nonzero reads in the liver or lung of any mouse at any time point were kept, yielding $N = 349$ barcodes for liver metastases and $N = 341$ barcodes for lung metastases.

Resampling bootstrap test

A resampling bootstrap test was performed to demonstrate whether tumor growth patterns for different mice were similar (i.e., homogeneous) among different observation stages. We used the early-stage data observed at injection and week 4 as the benchmark to generate pseudo-data to test the difference between the two mice at each time point. The specific steps of the resampling bootstrap test are given below:

Step 1. Calculate the difference of percentages of each barcode between two mice.

The logit transformation of the percentage of each barcode for each mouse at each time point was first calculated. A value of 0.5 was added if the number of reads equaled 0 before we performed logit transformation on the percentage. At any time point t , for each barcode $i = 1, \dots, N$, the absolute difference of logit transformed percentages, d_{ti} , between two mice (mouse A and mouse B) was calculated, i.e.,

$$d_{ti} = \text{abs}[\text{logit}(p_{ti}^A) - \text{logit}(p_{ti}^B)]$$

where p_{ti}^j is the abundance percentage of barcode i in mouse j at time t , $i = 1, \dots, N$, $j = \text{mouse A or B}$, $t = \text{week 0, 4, 6, 10, or 14}$.

Step 2. Determine the parameters of the null distribution based on data from the injection and week 4 time points.

We used the observations from the early stage time points (injection and week 4) as references and determined the parameters associated with the null distribution. At each time point, a total of $N = 3751$ absolute differences were obtained and the pooled data, including absolute differences at injection and week 4, were coalesced ($N = 7502$). On the basis of the empirical distribution of the pooled data, we divided the sample into three subgroups: (i) small differences ($\leq 20\%$ quantile), (ii) median differences ($> 20\%$, $\leq 80\%$ quantile), and (iii) large differences ($> 80\%$ quantile). Assuming that $\text{abs}(d_{ti})$ followed a normal distribution with mean μ_g and variance σ_g^2 in each group $g = 1, 2, 3$, we obtained the corresponding moment estimators, denoted by $\hat{\mu}_g$ and $\hat{\sigma}_g^2$, respectively.

Step 3. Simulate the pseudo samples under the null hypothesis.

At any time point t , for each barcode $i = 1, \dots, N$, we generate a total of $K = 5000$ pseudo sample for mice A and mice B:

$$\text{logit}(\tilde{p}_{ii}^A) = \text{logit}(\bar{p}_{ii}) + \frac{\epsilon_{ii}}{2}, \text{logit}(\tilde{p}_{ii}^B) = \text{logit}(\bar{p}_{ii}) - \frac{\epsilon_{ii}}{2}$$

where $\text{logit}(\bar{p}_{ii}) = [\text{logit}(p_{ii}^A) + \text{logit}(p_{ii}^B)]/2$ and $\epsilon_{ii} \sim w_1 N(\hat{\mu}_1, \hat{\sigma}_1^2) + w_2 N(\hat{\mu}_2, \hat{\sigma}_2^2) + w_3 N(\hat{\mu}_3, \hat{\sigma}_3^2)$, with $(w_1, w_2, w_3) = (0.2, 0.6, 0.2)$.

Step 4. Calculate the P value using the resampling data.

We used two statistics as the similarity measurements to test the similarity between the two mice at the same time point. The two test statistics were:

1) Statistic 1: $T1$ was the Pearson correlation coefficient between p_{ii}^A and p_{ii}^B , $i = 1, \dots, N$. If $T1$ was close to 1, mice A and B were similar.

2) Statistic 2: $T2$ was the average Euclidean distance between p_{ii}^A and p_{ii}^B , $i = 1, \dots, N$. If $T2$ was close to 0, mice A and B were similar.

The corresponding P values were then obtained on the basis of the observed data and the simulated data obtained from step 3.

Dynamic clustering

To analyze the barcode dynamic over time, the method of Clustering Large Applications (CLARA) (49) was applied to cluster the slopes of barcodes between each two consecutive time points. The final number of clusters was determined on the basis of the total within the sum of square (WSS) (50).

K-plot

To adjust the effect of injection and make all barcodes more comparable, we “moved” all barcodes to have the same start point at injection. Each line represented a locally weighted smoothing (lowess) line (51) of each cluster.

High-throughput clonal isolation

The high-throughput clone isolation workflow was adapted from a published protocol [Seth *et al.* (30)].

Deconvolution of barcode complexity from cell culture to NGS

Established cell cultures were deconvoluted at lower clonal complexity by seeding 5×10^3 per well barcoded cells in 96-well plates. Once the wells reached ~80% confluency, the cells were detached with 30 μ l of 0.25% trypsin-EDTA. One-third of the cells (10 μ l) were transferred to the corresponding well position on a 96-well PCR “sister” replica plate for further NGS library processing. The remaining cells were mixed with dimethyl sulfoxide (DMSO)–FBS (20:80) freezing solution and frozen at -80°C . Frozen plates were later used for flow-activated sorting to isolate single clones.

To detect the barcode composition of each well on the stored frozen plates, 4.7 μ l of lysis buffer [10 mM tris-HCl (pH 7.5), 10 mM NaCl, 10 mM MgCl_2 , 0.19% NP-40, and 2 μ g of proteinase K] was added to each well of the sister replica plates containing 10 μ l of cells-trypsin. The replica plate was then incubated at 50°C for 40 min and then at 95°C for 15 min to inactivate proteinase K. A mixture containing Titanium Taq DNA polymerase (Clontech-Takara), individual forward primer with unique index Ion-A-BC-GexBP1 (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXGATTCAAGCAGAAGACG-GCATAACGAAGACAG-3'), X bases refers to Ion Xpress 96 adaptor barcode sequence), and a common reverse primer GexSeqS-IonP1 (5'-CCACTACGCCTCCGCTTCTCTCTATGGGCAGTCGGT-GATTGAGGTTTCAGAGTTCTACAGTC-3'), both at 600 nM final

concentration, were added to each well and amplified with the following cycling profile: 95°C 3 min, 6 cycles of 95°C 30 s, 56°C 10 s, 72°C 10 s, followed by 22 cycles of 95°C 30 s, 62°C 10 s, 72°C 10 s, and then final extension at 72°C for 2 min and 4°C hold. Ten microliters of PCR products from each well was pooled together (total 960 μ l). Five hundred microliters of the pooled product was concentrated 10-fold with one time 1.8 \times AMPure XP purification (Beckman Coulter) by eluting with 50 μ l of elution buffer. The 50- μ l concentrated DNA was loaded on a 2% agarose gel for band purification. Expected bands (size ~185 bp) were cut, and DNA was purified using QIAquick gel extraction kit (Qiagen). Following gel purification, DNA was quantified with the High Sensitivity D1000 ScreenTape and Agilent 4200 TapeStation system (Agilent Technologies). The library was then sequenced with the Ion Torrent Proton NGS platform (Thermo Fisher Scientific).

Isolation of single barcoded cells, flow-activated cell sorting, and positional sequencing

After identification of wells with higher representation of clones of interest, previously frozen plates were kept on ice, and 180 μ l of 37°C warm DMEM per well was added to thaw wells of interest. The thawed cells were then transferred to another 96-well plate and recovered in the 37°C incubator. After 2 days, the cells were detached and resuspended in DMEM/FBS containing SYTOX Green (Thermo Fisher Scientific) to sort live single cells in 96-well plates containing DMEM supplemented with 20% FBS and 1% penicillin-streptomycin. Single-cell colonies were then left to grow with periodic monitoring. When the wells reached ~50% confluency, the cells were detached and split into two plates using the approach described in the previous section. After lysing, cells were prepared for PCR according to the manual of Titanium Taq DNA polymerase (Clontech-Takara) using individual forward primer with unique index Ion-A-BC-GexBP1 (5'-CCATCTCATCCCTGCGTGTCTCCGACTCAGXXXXXXXXXXGATTCAAGCAGAAGACG-GCATAACGAAGACAG-3'), with X bases referring to Ion Xpress 96 adaptor barcode sequence (Thermo Fisher Scientific, catalog 4474517), and a common unique indexed reverse primer GexSeqS-bc-IonP1 (5'-CCACTACGCCTCCGCTTCTCTCTATGGGCAGTCGGTGATYYYYYYYYCGAG-GTTCAGAGTTCTACAGTC-3', with Y referring to individual “plate index” (Table 1), both at 600 nM final concentration and added to each well. Wells were then amplified with the cycling profile of 95°C 3 min, 7 cycles of 95°C 30 s, 56°C 10 s, 72°C 10 s, followed by 23 cycles of 95°C 30 s, 62°C 10 s, 72°C 10 s, and then final extension at 72°C for 2 min and 4°C hold. Each well now contained amplified barcodes with NGS-compatible adaptors with unique well indexes, as well as unique plate indexes. All the PCR products were pooled from each well and each plate, and 50 μ l was loaded on a 2% agarose gel for band purification. The expected band size ~190 bp was cut, and DNA was purified using QIAquick gel extraction kit (Qiagen). Following gel purification, DNA was quantified with the High Sensitivity D1000 ScreenTape and Agilent 4200 TapeStation system (Agilent Technologies). The library was then sequenced with the Ion Torrent Proton NGS platform (Thermo Fisher Scientific).

Barcode analysis of Ion Torrent data

Fastq were demultiplexed for row and column barcodes using Ion Torrent NGS software. Upon considering differences in the sequencing platform and the nature of clonal sample, a highly stringent analysis strategy was used. Subsequent analysis was performed using custom R scripts, developed using Bioconductor packages (Biostrings and ShortRead) (52) (<https://bioconductor.org/packages/>

release/bioc/html/Biostrings.html). Briefly, we extracted reads that follow the pattern “5′-TCAAGCAGAAGACGGCATAACGAAGACAGTTCGNNNNNNNNNNNNNNNNNNTTTC-GNNNNNNNNNNNNNNNNNNNTTCGGACTGTAGAACTCTGAACCTCRYYYYYYYY-3′,” where Ns represent barcode 2 and barcode 1, respectively, and Ys represent the plate barcode (see Table 2). Barcodes of each read were then compared with the library (CellTracker 50M Lentiviral Double-Barcoded Library, Cellesta), allowing for mismatches. Reads from where both barcodes aligned uniquely with the sequences in the library with minimum hamming distance were used for downstream analysis. The reads were then separated into different plates using plate-specific indexes. A read number cutoff of 100 reads was applied to remove false-positive barcode ID from each of the 96 samples (well index). Wells containing clones of interest represented as the dominant fraction of reads were chosen for further expansion. In case the same barcode was detected in multiple wells, wells with the maximum percent representations were chosen for downstream expansion and isolation. It is possible, although exceptional, that more than one unique barcode was associated with each sample due to a negligible percentage of double barcoding and the accuracy of flow-activated single-cell sorting.

Expansion of isolated clonal cultures and Sanger barcode validation

After identification of wells of interest, wells were thawed as previously described and allowed to expand in tissue culture plates for further barcode Sanger validation. Briefly, after expansion, a fraction of clonal cells was used to extract genomic DNA using DNeasy Blood & Tissue Kits (Qiagen) according to the manufacturer’s protocol. For PCR amplification, GexBP1 (5′-TCAAGCAGAAGACGGCATAACGAAGACA-3′) and SangerNR2 (5′-ACGAGCACCGACAACAACGCAGA-3′) were used as forward and reverse primers, at final concentration 600 nM, and followed cycling condition 95°C 3 min, 38 cycles of 95°C 30 s, 60°C 20 s, 72°C 10 s, followed by final extension at 72°C for 2 min and 4°C hold. PCR products were then loaded on a 2% agarose gel, and the expected band at 242 bp was cut followed by DNA purification using the QIAquick gel extraction kit (Qiagen) for Sanger sequencing. SangerNR2 was used as the sequencing primer (5′-ACGAGCACCGACAACAACGCAGA-3′).

Omics analysis

WES of isolated clones

Genomic DNA from six isolated clones as well as parental cells of the PDX cell line was extracted and subjected to WES. Exome capture was performed on 500 ng of DNA per sample based on the Kapa Hyper Prep using the Agilent SureSelect Human All Exon kit according to the manufacturer’s instructions. WES was performed on the

Illumina HiSeq 2500 sequencing platform. Pair-end reads in FASTQ format were generated from BCL raw data using Illumina CASAVA. The reads were aligned to the hg19 reference genome using BWA (53). Duplicate reads were removed using Picard (<http://broadinstitute.github.io/picard/>), and local realignments were performed using GATK (54).

Mutation calling and CNV identification

A pooled common normal was used as control for germline mutations and polymorphism variants. Somatic single-nucleotide variants (SNVs) were identified using MuTect (55), and small insertions and deletions (InDels) were identified using Pindel (56). A series of post-calling filtering were applied for somatic mutations including (i) total read count in tumor sample ≥ 20 , (ii) total read count in germline sample ≥ 10 , (iii) VAF (variant allele frequency) ≥ 0.02 in tumor sample and ≤ 0.02 in matched normal sample, and (iv) a population frequency threshold of 0.5% was used to filter out common variants in the databases of dbSNP129 (57), 1000 Genome Projects (58), Exome Aggregation Consortium (59), and ESP6500 (60). To understand the potential functional consequence of detected variants, we annotated them using Annovar (61) and dbNSFP (62), and compared them with those from dbSNP (63), ClinVar (64), COSMIC (65), and TCGA databases. CNVs were identified using an in-house algorithm named ExomeCN. The copy number \log_2 ratios of tumor versus matched normal were calculated across the entire capture regions and then subjected to segmentation using CBS (66). A cutoff of \log_2 ratio ≤ -0.4 was applied to identify copy losses, and \log_2 ratio ≥ 0.4 was applied to identify copy gains.

RNA-seq analysis

Total RNA from six isolated clones of the PDX1 cell line was extracted using RNeasy Mini Kit (Qiagen) following the manufacturer’s instructions. Quantity and quality were measured using the RNA ScreenTape Analysis (Agilent Technologies). Raw sequencing data were converted to fastq format and aligned to hg19 reference genome using the Spliced Transcripts Alignment to a Reference (STAR) algorithm (67). HTSeq-count was used to generate the raw read counts for each gene (68). DESeq2 was used for data processing, normalization, and differential expression analysis following standard procedures (69). The differentially expressed genes (DEGs) were selected by the criteria of the $\log_2(\text{foldchange}) \leq -1$ or ≥ 1 , and the cutoff of the adjusted P value was 0.05. IPA was conducted with default parameters (<https://qiagenbioinformatics.com/products/ingenuity-pathway-analysis/>). In brief, the DEGs were obtained by comparing the promoter clones versus the non-met clones using DESeq2. The genes were then mapped to IPA database with either Ensembl gene ID or HUGO gene symbols. The mapped DEGs and corresponding \log_2 Foldchange, P values, and adjusted P values were used as input for IPA core

Table 2. Plate indexes.

GExSeqS-bc1onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATYYYYYYYTGAGGTTTCAGAGTTCTACAGTC-3′
GExSeqS-bc1onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTATATCCACGAGGTTTCAGAGTTCTACAGTC-3′
GExSeqS-bc3onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTATCCATACGAGGTTTCAGAGTTCTACAGTC-3′
GExSeqS-bc5onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTCCATATACGAGGTTTCAGAGTTCTACAGTC-3′
GExSeqS-bc7onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTATAGGTACGAGGTTTCAGAGTTCTACAGTC-3′
GExSeqS-bc9onP1	5′-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTAGGTATACGAGGTTTCAGAGTTCTACAGTC-3′

analysis to determine the canonical pathway enrichment. Metastasis and nonmetastasis gene signatures were selected using the top 100 most DEGs.

Single-cell CNV analysis through MissionBio Tapestry platform

Cells were resuspended in Tapestry cell buffer for single-cell suspension at concentration of 4000 cells per microliter. Cell encapsulations, lysis, barcoding, and library preparation were performed with Tapestry V3 workflow (Mission Bio, San Francisco, CA, USA). The sequencing was performed on Illumina's NGS platform. The genome-wide CNV panel was designed by Mission Bio targeting single-nucleotide polymorphisms (SNPs) evenly spaced out across unmasked regions of the human genome, approximately 10 million bp apart. A minimum of 10 amplicons were used for smaller chromosomes despite a smaller gap between amplicons, when possible. Candidate SNPs were selected by filtering the dbSNP 151 database for synonymous mutations with a minor allele frequency > 0.35 within regions of moderate GC content (28 to 62%). The FASTQ file was processed using the Tapestry Pipeline (<https://portal.missionbio.com/>), which outputs an h5 file. The h5 file contains the allele frequency (AF) calls for all cells for each position on all amplicons and the read counts to each cell and amplicon. A total of 708 amplicons and 2217 cells were used for SNP and CNV analysis. The 493 variants obtained had more than 50% of the cells with DP >10, GQ >30, AF for WT calls <5, AF for HOM calls >95, AF for HET calls >30, and at least 1% of the cells mutated. The copy number for each clone was estimated using a reference cell line (Spiked-in Raji) as a reference and assuming that Chromosome X has a mean copy number of 1 in all cells. The copy number estimation was performed using the "compute_ploidy" function in the Mission Bio Mosaic package.

Single-cell CNV analysis through 10x Genomics platform

About 4000 cells were captured followed by library preparation using the 10x Genomics Chromium Single-Cell DNA Reagent Kit according to the manufacturer's protocol. The library was then sequenced on the Illumina NovaSeq S2 platform. Downstream analysis on single-cell CNV was performed according to 10x Genomics' instruction (<https://support.10xgenomics.com/single-cell-dna/software/overview/welcome>). In brief, sample demultiplexing, barcode processing, read alignment, copy number estimation, and hierarchical clustering were performed through Cell Ranger (10x Genomics). The visualization of intratumoral heterogeneity and hierarchical clustering was acquired through Loupe scDNA Browser (10x Genomics).

Single-cell RNA-seq of PDX cell lines

Five thousand cells of PDX1 parental cells were captured followed by library preparation using 10x Genomics Chromium single cell 3' V3 chemistry according to the manufacturer's protocol. The library was then sequenced on the Illumina NovaSeq S2 platform. Downstream analysis on the resulting single-cell sequencing data was conducted using the R package Seurat (70). Raw data had 4642 cells with 21,808 genes. The dataset was filtered with the following criteria: number of genes greater than 200 (removing low-quality cells) and less than 6700 (removing possible doublets or multiplets); reading count less than 60,000 (removing possible doublets or multiplets); percent of mitochondria less than 5% (removing low-quality or dead cells). The cleaned dataset resulted in a total of 3397 cells. Homemade metastasis gene signatures were obtained from bulk RNA-seq analysis. The top 100 up-regulated genes (met-up) and top 100 down-regulated genes (met-down) were selected as the two gene lists. The R package AUCell (71) was used to identify cells with the 100 met-up and 100 met-down

metastasis signatures (i.e., cells enriched with the genes in these two signatures). Together, 639 cells were identified as having either the 100 met-up, 100 met-down, or both signatures (correspondingly, 2758 cells were not enriched in these two signatures). Among the 639 cells, specifically, "met" was defined as cells enriched in 100 met-up but not in 100 met-down ($n = 453$). A UMAP (Uniform Manifold Approximation and Projection) plot was generated for cells with metastasis signatures only.

Human pancreatic cancer spatial transcriptomic analysis

Single-cell sequencing and spatial transcriptomics data of three pancreatic cancer patient samples (39) were downloaded from HTAN Data Coordinating Center Data Portal: <https://data.humantumoratlas.org/>. Spatial transcriptomics data were normalized with the "SCTransform" function of R Seurat package and scaled. The "variable.feature.n" parameter was set to the number of total features to ensure that the normalized and scaled matrix included enough genes for gene signature analysis. The nearest template prediction (NTP) algorithm from R package "CMScaller" was used to analyze the gene expression matrix obtained from spatial transcriptomics analysis (72, 73). The 100 met-up and 100 met-down genes were used as the templates. The "nPerm" parameter of the "ntp" function was set to 1000 for P value estimation. If the prediction of a spot shows that its gene expression is closer to the met-up gene template than the met-down templates with a P value <0.05, the spot was assigned as Met. Otherwise, they were assigned as "All others." The cell types of the spots were inferred from single-cell sequencing data of the same patient samples that were filtered, normalized with "SCTransformation," and clustered as described in the published data (39). Cell types of single cells were assigned and were used to annotate the spot-level cell types with the "FindTransferAnchors" and "TransferData" functions from R Seurat package. The distance to the met-up and met-down templates of single cells was also calculated with the NTP algorithm as described above.

Human pancreatic cancer survival analysis

ssGSEA (74) method from the GSVA (75) (v1.48.3) R package was used via "gsva" function with "method = 'ssgsea'" and "ssgsea.norm = TRUE" on transcripts per million values for all TCGA + META-PRISM samples in the same run. After calculating gene set enrichment scores for samples, gene set enrichment scores were normalized for standard normal distribution (mean = 0, SD = 1). Cox proportional hazard model was fit with overall survival of all samples (TCGA + META-PRISM, $n = 223$) using standard normalized enrichment scores. Then, all samples were dichotomized into two groups, met_up_dominant and met_down_dominant, using standard normalized gene set enrichment scores (If met_up_enrichment_score > met_down_enrichment_score, then this sample is met_up_dominant, and vice versa). Last, overall survival of met_up_dominant and met_down_dominant groups was compared with Kaplan-Meier plots using log-rank test. All survival analysis was performed in R v4.3.1 using "survival" and "survminer" packages.

Hematoxylin and eosin staining

Collected tissues were fixed overnight with buffered paraformaldehyde (PFA) followed by 70% ethanol and embedded in paraffin using Leica ASP300S processor. Tissue blocks were sectioned with a microtome (Leica RM2235). Slides were then deparaffined and stained with hematoxylin and eosin. Images were captured with a Nikon DS-Fi1 digital camera using a wide-field Nikon Eclipse-Ci microscope.

Invasion assay

Invasion assays were performed with Corning BioCoat Matrigel Invasion Chambers according to the manufacturer's instructions. In brief, cells were starved with serum-free medium for 24 hours. Cells (1×10^5 to 2×10^5) in 0.5-ml serum-free medium were added to the inserts, and medium plus 10% FBS were added to the wells as chemoattractant with or without NMDAR agonist glutamate or antagonist memantine. After 24 hours, the cells were fixed with 4% PFA followed by 0.5% crystal violet staining. Images were captured with a wide-field Nikon Eclipse-Ni microscope. The number of invaded cells was determined using ImageJ software.

Reverse transcription and quantitative real-time PCR

Total RNA from each clone was extracted with Qiagen RNeasy Mini Kit according to the manufacturer's instructions. One microgram of RNA was used to generate cDNA using Thermo Fisher Scientific SuperScript VILO Master Mix. Ten nanograms of cDNA was used for quantitative PCR using Applied Biosystems PowerUp SYBR Green PCR Master Mix with gene-specific primers (see Table 3). Relative gene expression level was determined by the comparative CT ($\Delta\Delta\text{CT}$) method and normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH).

Intracellular Ca^{2+} assay

Intracellular Ca^{2+} concentration was determined by the Enzo FLUOFORTE Calcium Assay Kit according to the manufacturer's instructions. In brief, 8×10^4 cells per well were plated in a 96-well plate for overnight. The cells were then stained with FLUOFORTE Dye for 1 hour at room temperature followed by fluorescence detection at $\text{Ex} = 490 \text{ nm}/\text{Em} = 525 \text{ nm}$ using the PHERAStar HTS microplate reader (BMG Labtech).

MTT cell viability assay

The MTT assay (Sigma-Aldrich) was performed according to the manufacturer's instructions. In brief, 5×10^3 cells per well were plated in a 96-well plate for overnight. The cells were then treated with different concentrations of memantine for 48 hours followed by incubation with complete medium containing MTT (0.4 mg/ml) at 37°C for 4 hours. The reduced MTT crystals were dissolved in DMSO (Sigma-Aldrich), and the absorbance was detected at 570 nm using the PHERAStar HTS microplate reader (BMG Labtech).

Statistical analysis

In vitro and in vivo data are presented as the means \pm SD. Statistical analyses were performed using a two-tailed unpaired Student's *t* test

or one-way analysis of variance (ANOVA) after the evaluation of variance. No statistical methods were applied to predetermine sample size.

Supplementary Materials

This PDF file includes:

Figs. S1 to S6

REFERENCES AND NOTES

- G. H. Heppner, Tumor heterogeneity. *Cancer Res.* **44**, 2259–2265 (1984).
- S. Loponte, S. Lovisa, A. K. Deem, A. Carugo, A. Viale, The many facets of tumor heterogeneity: Is metabolism lagging behind? *Cancers* **11**, 1574 (2019).
- M. Gerlinger, N. McGranahan, S. M. Dewhurst, R. A. Burrell, I. Tomlinson, C. Swanton, Cancer: Evolution within a lifetime. *Annu. Rev. Genet.* **48**, 215–236 (2014).
- N. McGranahan, C. Swanton, Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* **27**, 15–26 (2015).
- C. L. Chaffer, R. A. Weinberg, A perspective on cancer cell metastasis. *Science* **331**, 1559–1564 (2011).
- I. J. Fidler, The pathogenesis of cancer metastasis: The 'seed and soil' hypothesis revisited. *Nat. Rev. Cancer* **3**, 453–458 (2003).
- S. Valastyan, R. A. Weinberg, Tumor metastasis: Molecular insights and evolving paradigms. *Cell* **147**, 275–292 (2011).
- F. G. Giancotti, Mechanisms governing metastatic dormancy and reactivation. *Cell* **155**, 750–764 (2013).
- P. J. Campbell, S. Yachida, L. J. Mudie, P. J. Stephens, E. D. Pleasance, L. A. Stebbings, L. A. Morsberger, C. Latimer, S. McLaren, M. L. Lin, D. J. McBride, I. Varela, S. A. Nik-Zainal, C. Leroy, M. Jia, A. Menzies, A. P. Butler, J. W. Teague, C. A. Griffin, J. Burton, H. Swerdlow, M. A. Quail, M. R. Stratton, C. Iacobuzio-Donahue, P. A. Futreal, The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* **467**, 1109–1113 (2010).
- S. Jones, W. D. Chen, G. Parmigiani, F. Diehl, N. Beerwinkel, T. Antal, A. Traulsen, M. A. Nowak, C. Siegel, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, J. Willis, S. D. Markowitz, Comparative lesion sequencing provides insights into tumor evolution. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 4283–4288 (2008).
- S. Yachida, S. Jones, I. Bozic, T. Antal, R. Leary, B. Fu, M. Kamiyama, R. H. Hruban, J. R. Eshleman, M. A. Nowak, V. E. Velculescu, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* **467**, 1114–1117 (2010).
- R. Yaeger, W. K. Chatila, M. D. Lipsyc, J. F. Hechtman, A. Cercek, F. Sanchez-Vega, G. Jayakumar, S. Middha, A. Zehir, M. T. A. Donoghue, D. You, A. Viale, N. Kemeny, N. H. Segal, Z. K. Stadler, A. M. Varghese, R. Kundra, J. Gao, A. Syed, D. M. Hyman, E. Vakiani, N. Rosen, B. S. Taylor, M. Ladanyi, M. F. Berger, D. B. Solit, J. Shia, L. Saltz, N. Schultz, Clinical sequencing defines the genomic landscape of metastatic colorectal cancer. *Cancer Cell* **33**, 125–136.e3 (2018).
- R. Klotz, A. Thomas, T. Teng, S. M. Han, O. Iriando, L. Li, S. Restrepo-Vassalli, A. Wang, N. Izadian, M. MacKay, B. S. Moon, K. J. Liu, S. K. Ganesan, G. Lee, D. S. Kang, C. S. Walmesley, C. Pinto, M. F. Press, W. Lu, J. Lu, D. Juric, A. Bardia, J. Hicks, B. Salhia, F. Attenello, A. D. Smith, M. Yu, Circulating tumor cells exhibit metastatic tropism and reveal brain metastasis drivers. *Cancer Discov.* **10**, 86–103 (2020).
- S. Mueller, T. Engleitner, R. Maresch, M. Zukowska, S. Lange, T. Kaltenbacher, B. Konukiewicz, R. Öllinger, M. Zwiebel, A. Strong, H. Y. Yen, R. Banerjee, S. Louzada, B. Fu, B. Seidler, J. Götzfried, K. Schuck, Z. Hassan, A. Arbeiter, N. Schönhuber, S. Klein, C. Veltkamp, M. Friedrich, L. Rad, M. Barenboim, C. Ziegenhain, J. Hess, O. M. Dovey, S. Eser, S. Parekh, F. Constantino-Casas, J. de la Rosa, M. I. Sierra, M. Fraga, J. Mayerle, G. Klöppel, J. Cadiñanos, P. Liu, G. Vassiliou, W. Weichert, K. Steiger, W. Enard, R. M. Schmid, F. Yang, K. Unger, G. Schneider, I. Varela, A. Bradley, D. Saur, R. Rad, Evolutionary routes and KRAS dosage define pancreatic cancer phenotypes. *Nature* **554**, 62–68 (2018).
- R. Maddipati, R. J. Norgard, T. Baslan, K. S. Rathi, A. Zhang, A. Saeid, T. Higashihara, F. Wu, A. Kumar, V. Annamalai, S. Bhattacharya, P. Raman, C. A. Adkisson, J. R. Pitarresi, M. D. Wengyn, T. Yamazoe, J. Li, D. Balli, M. J. LaRiviere, T. V. C. Ngo, I. W. Folkert, I. D. Millstein, J. Bermeo, E. L. Carpenter, J. C. McAuliffe, M. H. Oktay, R. A. Brekken, S. B. Lowe, C. A. Iacobuzio-Donahue, F. Notta, B. Z. Stanger, MYC levels regulate metastatic heterogeneity in pancreatic adenocarcinoma. *Cancer Discov.* **12**, 542–561 (2022).
- M. Chan-Seng-Yue, J. C. Kim, G. W. Wilson, K. Ng, E. F. Figueroa, G. M. O'Kane, A. A. Connor, R. E. Denroche, R. C. Grant, J. McLeod, J. M. Wilson, G. H. Jang, A. Zhang, A. Dodd, S. B. Liang, A. Borgida, D. Chadwick, S. Kalimuthu, I. Lungu, J. M. S. Bartlett, P. M. Krzyzanowski, V. Sandhu, H. Tiriak, F. E. M. Froeling, J. M. Karasinska, J. T. Topham, D. J. Renouf, D. F. Schaeffer, S. J. M. Jones, M. A. Marra, J. Laskin, R. Chetty, L. D. Stein, G. Zogopoulos, B. Haibe-Kains, P. J. Campbell, D. A. Tuveson, J. J. Knox, S. E. Fischer, S. Gallinger, F. Notta, Transcription phenotypes of pancreatic cancer are driven by genomic events during tumor evolution. *Nat. Genet.* **52**, 231–240 (2020).

Table 3. Primers for quantitative real-time PCR.

Gene	Primer sequence
CAMK4_FW	GCCTCGTCCCGATTACTG
CAMK4_RV	TCCCTTCTGTTGCATCTGT
GRIN2A_FW	TCATGCAGGATTATGACTGGCA
GRIN2A_RV	TGTGGTCTTGACGAAGCTGAT
GRIA2_FW	CATTGAGATGAGACCCGACCT
GRIA2_RV	GGTATGCAAACCTGTCCCATGA

17. R. Gao, A. Davis, T. O. McDonald, E. Sei, X. Shi, Y. Wang, P. C. Tsai, A. Casasent, J. Waters, H. Zhang, F. Meric-Bernstam, F. Michor, N. E. Navin, Punctuated copy number evolution and clonal stasis in triple-negative breast cancer. *Nat. Genet.* **48**, 1119–1130 (2016).
18. I. J. Fidler, M. L. Kripke, Metastasis results from preexisting variant cells within a malignant tumor. *Science* **197**, 893–895 (1977).
19. K. Yamashina, G. H. Heppner, Correlation of frequency of induced mutation and metastatic potential in tumor cell lines from a single mouse mammary tumor. *Cancer Res.* **45**, 4015–4019 (1985).
20. S. Nagrath, L. V. Sequist, S. Maheswaran, D. W. Bell, D. Irimia, L. Ullkus, M. R. Smith, E. L. Kwak, S. Digumarthy, A. Muzikansky, P. Ryan, U. J. Balis, R. G. Tompkins, D. A. Haber, M. Toner, Isolation of rare circulating tumour cells in cancer patients by microchip technology. *Nature* **450**, 1235–1239 (2007).
21. Y. Husemann, J. B. Geigl, F. Schubert, P. Musiani, M. Meyer, E. Burghart, G. Forni, R. Eils, T. Fehm, G. Riethmüller, C. A. Klein, Systemic spread is an early step in breast cancer. *Cancer Cell* **13**, 58–68 (2008).
22. A. D. Rhim, E. T. Mirek, N. M. Aiello, A. Maitra, J. M. Bailey, F. McAllister, M. Reichert, G. L. Beatty, A. K. Rustgi, R. H. Vonderheide, S. D. Leach, B. Z. Stanger, EMT and dissemination precede pancreatic tumor formation. *Cell* **148**, 349–361 (2012).
23. J. G. Reiter, A. P. Makohon-Moore, J. M. Gerold, A. Heyde, M. A. Attiye, Z. A. Kohutek, C. J. Tokheim, A. Brown, R. M. DeBlasio, J. Niyazov, A. Zucker, R. Karchin, K. W. Kinzler, C. A. Iacobuzio-Donahue, B. Vogelstein, M. A. Nowak, Minimal functional driver gene heterogeneity among untreated metastases. *Science* **361**, 1033–1037 (2018).
24. L. Ding, T. J. Ley, D. E. Larson, C. A. Miller, D. C. Koboldt, J. S. Welch, J. K. Ritchey, M. A. Young, T. Lamprecht, M. D. McLellan, J. F. McMichael, J. W. Wallis, C. Lu, D. Shen, C. C. Harris, D. J. Dooling, R. S. Fulton, L. L. Fulton, K. Chen, H. Schmidt, J. Kalicki-Veizer, V. J. Magrini, L. Cook, S. D. McGrath, T. L. Vickery, M. C. Wendt, S. Heath, M. A. Watson, D. C. Link, M. H. Tomasson, W. D. Shannon, J. E. Payton, S. Kulkarni, P. Westervelt, M. J. Walter, T. A. Graubert, E. R. Mardis, R. K. Wilson, J. F. DiPersio, Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* **481**, 506–510 (2012).
25. M. Jamal-Hanjani, G. A. Wilson, N. McGranahan, N. J. Birkbak, T. B. K. Watkins, S. Veeriah, S. Shafi, D. H. Johnson, R. Mitter, R. Rosenthal, M. Salm, S. Horswell, M. Escudero, N. Matthews, A. Rowan, T. Chambers, D. A. Moore, S. Turajlic, H. Xu, S. M. Lee, M. D. Forster, T. Ahmad, C. T. Hiley, C. Abbosh, M. Falzon, E. Borg, T. Marafioti, D. Lawrence, M. Hayward, S. Kolvekar, N. Panagiotopoulos, S. M. Janes, R. Thakrar, A. Ahmed, F. Blackhall, Y. Summers, R. Shah, L. Joseph, A. M. Quinn, P. A. Crosbie, B. Naidu, G. Middleton, G. Langman, S. Trotter, M. Nicolson, H. Remmen, K. Kerr, M. Chetty, L. Gomersall, D. A. Fennell, A. Nakas, S. Rathinam, G. Anand, S. Khan, P. Russell, V. Ezhil, B. Ismail, M. Irvin-Sellers, V. Prakash, J. F. Lester, M. Kornaszewska, R. Attanoos, H. Adams, H. Davies, S. Dentre, P. Taniere, B. O'Sullivan, H. L. Lowe, J. A. Hartley, N. Iles, H. Bell, Y. Ngai, J. A. Shaw, J. Herrero, Z. Szallasi, R. F. Schwarz, A. Stewart, S. A. Quezada, J. le Quesne, P. van Loo, C. Dive, A. Hackshaw, C. Swanton, Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* **376**, 2109–2121 (2017).
26. A. Schuh, J. Becq, S. Humphray, A. Alexa, A. Burns, R. Clifford, S. M. Feller, R. Grocock, S. Henderson, I. Khrebtukova, Z. Kingsbury, S. Luo, D. McBride, L. Murray, T. Menju, A. Timbs, M. Ross, J. Taylor, D. Bentley, Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. *Blood* **120**, 4191–4196 (2012).
27. L. Sivapalan, G. J. Thorn, E. Gadaleta, H. M. Kocher, H. Ross-Adams, C. Chelala, Longitudinal profiling of circulating tumour DNA for tracking tumour dynamics in pancreatic cancer. *BMC Cancer* **22**, 369 (2022).
28. S. Turajlic, H. Xu, K. Litchfield, A. Rowan, T. Chambers, J. I. Lopez, D. Nicol, T. O'Brien, J. Larkin, S. Horswell, M. Stares, L. Au, M. Jamal-Hanjani, B. Challacombe, A. Chandra, S. Hazell, C. Eichler-Jonsson, A. Soulati, S. Chowdhury, S. Rudman, J. Lynch, A. Fernando, G. Stamp, E. Nye, F. Jabbar, L. Spain, S. Lall, R. Guarch, M. Falzon, I. Proctor, L. Pickering, M. Gore, T. B. K. Watkins, S. Ward, A. Stewart, R. DiNatale, M. F. Becerra, E. Reznik, J. J. Hsieh, T. A. Richmond, G. F. Mayhew, S. M. Hill, C. McNally, C. Jones, H. Rosenbaum, S. Stanislaw, D. L. Burgess, N. R. Alexander, C. Swanton, PEACE, TRACERx Renal Consortium, Tracking cancer evolution reveals constrained routes to metastases: TRACERx renal. *Cell* **173**, 581–594.e12 (2018).
29. H. Ying, P. Dey, W. Yao, A. C. Kimmelman, G. F. Draetta, A. Maitra, R. A. DePinho, Genetics and biology of pancreatic ductal adenocarcinoma. *Genes Dev.* **30**, 355–385 (2016).
30. S. Seth, C. Y. Li, I. L. Ho, D. Corti, S. Lopoente, L. Sapio, E. del Poggetto, E. Y. Yen, F. S. Robinson, M. Peoples, T. Karpintins, A. K. Deem, T. Kumar, X. Song, S. Jiang, Y. Kang, J. Fleming, M. Kim, J. Zhang, A. Maitra, T. P. Heffernan, V. Giuliani, G. Genovese, A. Futreal, G. F. Draetta, A. Carugo, A. Viale, Pre-existing functional heterogeneity of tumorigenic compartment as the origin of chemoresistance in pancreatic tumors. *Cell Rep.* **26**, 1518–1532.e9 (2019).
31. M. P. Kim, D. B. Evans, H. Wang, J. L. Abbruzzese, J. B. Fleming, G. E. Gallick, Generation of orthotopic and heterotopic human pancreatic cancer xenografts in immunodeficient mice. *Nat. Protoc.* **4**, 1670–1680 (2009).
32. A. Davis, R. Gao, N. Navin, Tumor evolution: Linear, branching, neutral or punctuated? *Biochim. Biophys. Acta Rev. Cancer* **1867**, 151–161 (2017).
33. J. W. Lee, C. A. Komar, F. Bengsch, K. Graham, G. L. Beatty, Genetically engineered mouse models of pancreatic cancer: The KPC Model (*LSL-Kras^{G12D/+};LSL-Trp53^{R172H/+};Pdx-1-Cre*), its variants, and their application in immuno-oncology drug discovery. *Curr. Protoc. Pharmacol.* **73**, 14.39.1–14.39.20 (2016).
34. S. K. Apple, Sentinel lymph node in breast cancer: Review article from a pathologist's point of view. *J. Pathol. Transl. Med.* **50**, 83–95 (2016).
35. A. Viale, P. Pellicci, Regulation of self-renewing division in normal and leukaemia stem cells, in *Cell Cycle Deregulation in Cancer*, G. H. Enders, Ed. (Springer, 2010), pp. 109–125.
36. L. Li, D. Hanahan, Hijacking the neuronal NMDAR signaling circuit to promote tumor growth and invasion. *Cell* **153**, 86–100 (2013).
37. Q. Zeng, I. P. Michael, P. Zhang, S. Saghafinia, G. Knott, W. Jiao, B. D. McCabe, J. A. Galván, H. P. C. Robinson, I. Zlobec, G. Ciriello, D. Hanahan, Synaptic proximity enables NMDAR signalling to promote brain metastasis. *Nature* **573**, 526–531 (2019).
38. L. Li, Q. Zeng, A. Bhutkar, J. A. Galván, E. Karamitopoulou, D. Noordermeer, M. W. Peng, A. Piersigilli, A. Perren, I. Zlobec, H. Robinson, M. L. Iroula-Arispe, D. Hanahan, GKAP acts as a genetic modulator of NMDAR signaling to govern invasive tumor growth. *Cancer Cell* **33**, 736–751.e5 (2018).
39. D. Cui Zhou, R. G. Jayasinghe, S. Chen, J. M. Herndon, M. D. Iglesia, P. Navale, M. C. Wendt, W. Maravan, K. Sato, E. Storr, C. K. Mo, J. Liu, A. N. Southard-Smith, Y. Wu, N. Naser al Deen, J. M. Baer, R. S. Fulton, M. A. Wyczalkowski, R. Liu, C. C. Fronick, L. A. Fulton, A. Shinkle, L. Thammavong, H. Zhu, H. Sun, L. B. Wang, Y. Li, C. Zuo, J. F. McMichael, S. R. Davies, E. L. Appelbaum, K. J. Robbins, S. E. Chasnoff, X. Yang, A. N. Reeb, C. Oh, M. Serasanambati, P. Lal, R. Varghese, J. R. Mashl, J. Ponce, N. V. Terekhanova, L. Yao, F. Wang, L. Chen, M. Schnaubelt, R. J. H. Lu, J. K. Schwarz, S. V. Puram, A. H. Kim, S. K. Song, K. I. Shoghi, K. S. Lau, T. Ju, K. Chen, D. Chatterjee, W. G. Hawkins, H. Zhang, S. Achilefu, M. G. Chheda, S. T. Oh, W. E. Gillanders, F. Chen, D. G. DeNardo, R. C. Fields, L. Ding, Spatially restricted drivers and transcriptional cell populations cooperate with the microenvironment in untreated and chemo-resistant pancreatic cancer. *Nat. Genet.* **54**, 1390–1405 (2022).
40. Y. Pradat, J. Viot, A. A. Yurchenko, K. Gunbin, L. Cerbone, M. Deloger, G. Grisay, L. Verlingue, V. Scott, I. Padioulet, L. Panunzi, S. Michiels, A. Hollebecq, G. Jules-Clément, L. Mezquita, A. Lainé, Y. Loriot, B. Besse, L. Friboulet, F. André, P. H. Couronné, D. Gautheret, S. I. Nikolaev, Integrative pan-cancer genomic and transcriptomic analyses of refractory metastatic cancer. *Cancer Discov.* **13**, 1116–1143 (2023).
41. I. J. Fidler, Metastasis: Quantitative analysis of distribution and fate of tumor emboli labeled with 125 I-5-iodo-2'-deoxyuridine. *J. Natl. Cancer Inst.* **45**, 773–782 (1970).
42. K. J. Luzzi, I. C. MacDonald, E. E. Schmidt, N. Kerkvliet, V. L. Morris, A. F. Chambers, A. C. Groom, Multistep nature of metastatic inefficiency: Dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. *Am. J. Pathol.* **153**, 865–873 (1998).
43. R. Maddipati, B. Z. Stanger, Pancreatic cancer metastases harbor evidence of polyclonality. *Cancer Discov.* **5**, 1086–1097 (2015).
44. R. S. Kerbel, C. Waghorne, B. Korczak, A. Lagarde, M. L. Breitman, Clonal dominance of primary tumours by metastatic cells: Genetic analysis and biological implications. *Cancer Surv.* **7**, 597–629 (1988).
45. J. Berthelet, V. C. Wimmer, H. J. Whitfield, A. Serrano, T. Boudier, S. Mangiola, M. Merdas, F. el-Saafin, D. Baloyan, J. Wilcox, S. Wilcox, A. C. Parslow, A. T. Papenfuss, B. Yeo, M. Ernst, B. Pal, R. L. Anderson, M. J. Davis, K. L. Rogers, F. Hollande, D. Merino, The site of breast cancer metastases dictates their clonal composition and reversible transcriptomic profile. *Sci. Adv.* **7**, eabf4408 (2021).
46. D. G. McFadden, T. Papagiannakopoulos, A. Taylor-Weiner, C. Stewart, S. L. Carter, K. Cibulskis, A. Bhutkar, A. McKenna, A. Dooley, A. Vernon, C. Sougnez, S. Malmstrom, M. Heimann, J. Park, F. Chen, A. F. Farago, T. Dayton, E. Shefler, S. Gabriel, G. Getz, T. Jacks, Genetic and clonal dissection of murine small cell lung carcinoma progression by genome sequencing. *Cell* **156**, 1298–1311 (2014).
47. A. Carugo, G. Genovese, S. Seth, L. Nezi, J. L. Rose, D. Bossi, A. Cicalese, P. K. Shah, A. Viale, P. F. Pettazzoni, K. C. Akdemir, C. A. Bristow, F. S. Robinson, J. Tepper, N. Sanchez, S. Gupta, M. R. Estecio, V. Giuliani, G. I. Dellino, L. Riva, W. Yao, M. E. di Francesco, T. Green, C. D'Alesio, D. Corti, Y. Kang, P. Jones, H. Wang, J. B. Fleming, A. Maitra, P. G. Pellicci, L. Chin, R. A. DePinho, L. Lanfrancone, T. P. Heffernan, G. F. Draetta, In vivo functional platform targeting patient-derived xenografts identifies WDR5-Myc association as a critical determinant of pancreatic cancer. *Cell Rep.* **16**, 133–147 (2016).
48. E. Del Poggetto, I.-L. Ho, C. Balestrieri, E.-Y. Yen, S. Zhang, F. Citron, R. Shah, D. Corti, G. R. Diaferia, C.-Y. Li, S. Lopoente, F. Carbone, Y. Hayakawa, G. Valenti, S. Jiang, L. Sapio, H. Jiang, P. Dey, S. Gao, A. K. Deem, S. Rose-John, W. Yao, H. Ying, A. D. Rhim, G. Genovese, T. P. Heffernan, A. Maitra, T. C. Wang, L. Wang, G. F. Draetta, A. Carugo, G. Natoli, A. Viale, Epithelial memory of inflammation limits tissue damage while promoting pancreatic tumorigenesis. *Science* **373**, eabj0486 (2021).
49. Q. Zhang, I. Couloigner, A new and efficient k-medoid algorithm for spatial clustering, in *Computational Science and Its Applications – ICCSA 2005*, O. Gervasi, M. L. Gavrilova, V. Kumar, A. Laguna, H. P. Lee, Y. Mun, D. Taniar, C. J. K. Tan, Eds. (Springer, 2005), pp. 181–189.
50. S. Dudoit, J. Fridlyand, T. P. Speed, Comparison of discrimination methods for the classification of tumors using gene expression data. *J. Am. Stat. Assoc.* **97**, 77–87 (2002).

51. W. S. Cleveland, S. J. Devlin, Locally weighted regression: An approach to regression analysis by local fitting. *J. Am. Stat. Assoc.* **83**, 596–610 (1988).
52. M. Morgan, S. Anders, M. Lawrence, P. Abouyou, H. Pagès, R. Gentleman, ShortRead: A bioconductor package for input, quality assessment and exploration of high-throughput sequencing data. *Bioinformatics* **25**, 2607–2608 (2009).
53. H. Li, R. Durbin, Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
54. A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytzky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, M. A. DePristo, The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
55. K. Cibulskis, M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander, G. Getz, Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219 (2013).
56. K. Ye, M. H. Schulz, Q. Long, R. Apweiler, Z. Ning, Pindel: A pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* **25**, 2865–2871 (2009).
57. R. Sachidanandam, D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Wiley, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E. R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander, D. Altshuler, International SNP Map Working Group, A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928–933 (2001).
58. C. Genomes Project, 1000 Genomes Project Consortium, G. R. Abecasis, A. Auton, L. D. Brooks, M. A. De Pristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth, G. A. McVean, An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
59. M. Lek, K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. DeFlaux, M. De Pristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H.-H. Won, D. Yu, D. M. Altshuler, D. Ardissono, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. M. Carroll, M. I. McCarthy, D. M. Govern, R. M. Pherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur, Exome Aggregation Consortium, Analysis of protein-coding genetic variation in 60,706 humans. *Nature* **536**, 285–291 (2016).
60. W. Fu, T. D. O'Connor, G. Jun, H. M. Kang, G. Abecasis, S. M. Leal, S. Gabriel, M. J. Rieder, D. Altshuler, J. Shendure, D. A. Nickerson, M. J. Bamshad, NHLBI Exome Sequencing Project, J. M. Akey, Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
61. K. Wang, M. Li, H. Hakonarson, ANNOVAR: Functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
62. X. Liu, C. Li, C. Mou, Y. Dong, Y. Tu, dbNSFP v4: A comprehensive database of transcript-specific functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Genome Med.* **12**, 103 (2020).
63. E. M. Smigielski, K. Sirotkin, M. Ward, S. T. Sherry, dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Res.* **28**, 352–355 (2000).
64. M. J. Landrum, J. M. Lee, G. R. Riley, W. Jang, W. S. Rubinstein, D. M. Church, D. R. Maglott, ClinVar: Public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res.* **42**, D980–D985 (2014).
65. J. G. Tate, S. Bamford, H. C. Jubb, Z. Sondka, D. M. Beare, N. Bindal, H. Boutselakis, C. G. Cole, C. Creatore, E. Dawson, P. Fish, B. Harsha, C. Hathaway, S. C. Jupe, C. Y. Kok, K. Noble, L. Ponting, C. C. Ramshaw, C. E. Rye, H. E. Speedy, R. Stefancsik, S. L. Thompson, S. Wang, S. Ward, P. J. Campbell, S. A. Forbes, COSMIC: The Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* **47**, D941–D947 (2019).
66. A. B. Olshen, E. S. Venkatraman, R. Lucito, M. Wigler, Circular binary segmentation for the analysis of array-based DNA copy number data. *Biostatistics* **5**, 557–572 (2004).
67. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
68. S. Anders, P. T. Pyl, W. Huber, HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
69. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
70. T. Stuart, A. Butler, P. Hoffman, C. Hafemeister, E. Papalexi, W. M. Mauck III, Y. Hao, M. Stoeckius, P. Smibert, R. Satija, Comprehensive integration of single-cell data. *Cell* **177**, 1888–1902.e21 (2019).
71. S. Aibar, C. B. González-Blas, T. Moerman, V. A. Huynh-Thu, H. Imrichova, G. Hulselmans, F. Rambow, J. C. Marine, P. Geurts, J. Aerts, J. van den Oord, Z. K. Atak, J. Wouters, S. Aerts, SCENIC: Single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
72. Y. Hoshida, Nearest template prediction: A single-sample-based flexible class prediction with confidence assessment. *PLOS ONE* **5**, e15543 (2010).
73. P. W. Eide, J. Bruun, R. A. Lothe, A. Sveen, CMScaller: An R package for consensus molecular subtyping of colorectal cancer pre-clinical models. *Sci. Rep.* **7**, 16618 (2017).
74. D. A. Barbie, P. Tamayo, J. S. Boehm, S. Y. Kim, S. E. Moody, I. F. Dunn, A. C. Schinzel, P. Sandy, E. Meylan, C. Scholl, S. Fröhling, E. M. Chan, M. L. Sos, K. Michel, C. Mermel, S. J. Silver, B. A. Weir, J. H. Reiling, Q. Sheng, P. B. Gupta, R. C. Wadlow, H. Ie, S. Hoersch, B. S. Wittner, S. Ramaswamy, D. M. Livingston, D. M. Sabatini, M. Meyerson, R. K. Thomas, E. S. Lander, J. P. Mesirov, D. E. Root, D. G. Gilliland, T. Jacks, W. C. Hahn, Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112 (2009).
75. S. Hanzelmann, R. Castelo, J. Guinney, GSVA: Gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14**, 7 (2013).
76. A. Colaprico, T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, H. Noushmehr, TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucleic Acids Res.* **44**, e71 (2016).

Acknowledgments: We thank R. Nguyen and J. Martinez in the Department of Genomic Medicine at MDCC for laboratory management, H. Risteski for continuous funding assistance, the MDACC Advanced Technology Genomics Core (ATGC) funded by National Cancer Institute (NCI) grant CA016672 (ATGC), the MDACC Flow Cytometry and Cellular Imaging Core Facility supported by grant NCI#P30CA16672, S. Parikh and M. A. Layton at Mission Bio R&D for the single-cell genome-wide copy number work, and G. Lozano for providing *Tp53^{R172H}* mice. **Funding:** This study was supported by Cancer Research and Prevention Institute of Texas (CPRI) grant RP230373, the V Foundation (V2020-018), NIH/NCI SPORE in Gastrointestinal Cancer grant P50CA221707, NIH/NCI grant R01CA258917 to A.V., UT MD Anderson Cancer Center Pancreatic Cancer Moon Shot and Pancreatic Cancer Action Network (PanCAN) to G.D., NIH/NCI P01 CA117969-12 to A.V. and G.D., Sheikh Khalifa Foundation to A.M., AIRC and the European Union's Horizon 2020 research and innovation program under Marie Skłodowska-Curie grant agreement no. 800924 to F.C., MD Anderson UTHealth GSBS Pauline Altman-Goldstein Foundation Discovery Fellowship to I.-L.H., CPRIT Training Award (RP210028) to E.-Y.Y., and Helmholtz (VH-NG-1153) and ERC (714922) to G.Ga. A.O.Ç. was supported by iNAMES—MDC-Weizmann Helmholtz International Research School for Imaging and Data Science from the NANO to the MESO. **Author contributions:** C.-Y.L., I.-L.H., and A.V. designed the studies, interpreted the data, and wrote the manuscript; F.W., R.L., F.D., and K.-A.D. performed mathematical and statistical analysis and assisted with data discussion; L.Z., Jingjing Liu, Y.C., Z.L., and A.O.Ç. analyzed and interpreted computational data; C.-Y.L., I.-L.H., E.-Y.Y., F.C., R.S., S.A., C.A.D., D.C., E.D.P., S.L., Jintan Liu, M.S., Z.C., H.J., and A.I. performed the experiments; N.F., A.M., H.J., C.-Y.L., K.-A.D., R.L., I.-L.H., A.I., H.Y., M.K., V.G., G.Ge, J.Z., A.F., T.H., L.W., G.Ga., G.D., A.C., and A.V. contributed essential reagent resources, assisted with data discussion and interpretation, or supervised and administered parts of the project; S.J. assisted with animal colony maintenance; A.D. and S.G. edited the manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** The PDX cell lines as well as resources and reagents can be provided pending scientific review and a completed material transfer agreement. Requests should be submitted to lead contact, A.V. (aviale@mdanderson.org). The scRNA-seq, scCNV-seq, clonal bulk RNA-seq, WES data, and custom codes have been deposited in the European Genome-Phenome Archive (EGA) (<https://ega-archive.org>) under study ID EGAS00001006358. RNA-seq and clinical data of TCGA-PAAD (pancreatic adenocarcinoma) cohort containing 178 primary tumor samples were downloaded via TCGAbiolinks (76). TPM (transcripts per million) expression values and survival data were used for downstream analysis. META-PRISM (40) data were downloaded from Gustave Roussy's cBioPortal (https://cbioportal.gustaveroussy.fr/study/summary?id=metaprisms_2023) and Gustave Roussy's nextcloud server (<https://nextcloud.gustaveroussy.fr/s/JPAraPnKQbtDMP>). TPM expression values and survival data of metastatic tissue samples from 45 META-PRISM PAAD patients were used for downstream analysis. All data needed to evaluate the conclusions in the paper are present in the paper and/or the Supplementary Materials.

Submitted 15 July 2022
 Accepted 8 February 2024
 Published 13 March 2024
 10.1126/sciadv.add9342