



CANCER

Power-law growth models explain incidences and sizes of pancreatic cancer precursor lesions and confirm spatial genomic findings

Ashley L. Kiemen^{1,2*}, Pei-Hsun Wu³, Alicia M. Braxton⁴, Toby C. Cornish⁵, Ralph H. Hruban^{1,2}, Laura D. Wood^{1,2}, Denis Wirtz^{3,1,2}, David Zwicker^{6*}

Pancreatic ductal adenocarcinoma is a rare but lethal cancer. Recent evidence suggests that pancreatic intraepithelial neoplasia (PanIN), a microscopic precursor lesion that gives rise to pancreatic cancer, is larger and more prevalent than previously believed. Better understanding of the growth-law dynamics of PanINs may improve our ability to understand how a miniscule fraction makes the transition to invasive cancer. Here, using three-dimensional tissue mapping, we analyzed >1000 PanINs and found that lesion size is distributed according to a power law. Our data suggest that in bulk, PanIN size can be predicted by general growth behavior without consideration for the heterogeneity of the pancreatic microenvironment or an individual's age, history, or lifestyle. Our models suggest that intraductal spread and fusing of lesions drive our observed size distribution. This analysis lays the groundwork for future mathematical modeling efforts integrating PanIN incidence, morphology, and molecular features to understand tumorigenesis and demonstrates the utility of combining experimental measurement with dynamic modeling in understanding tumorigenesis.

INTRODUCTION

Pancreatic ductal adenocarcinoma (PDAC), though rare, is predicted to be the second leading cause of cancer-related deaths in the United States by 2030 (1–3). A major hurdle in confronting this aggressive disease is that there is no effective screening test for PDAC or its precursor lesions (4). As such, PDAC is often diagnosed late when distant metastases are present and few clinical options remain. Only 15% of patients present with localized disease at the time of diagnosis (1). Improved understanding of the early development of pancreatic cancer is a necessary first step to developing effective screening tools. Most PDACs are believed to develop from microscopic precursor lesions called pancreatic intraepithelial neoplasia (PanIN, Fig. 1) (5). Study of PanIN is uniquely complicated due to its small size: PanIN lesions in most cases cannot be seen through noninvasive diagnostic imaging such as computed tomography, magnetic resonance imaging, and endoscopic ultrasound (6). PanINs are primarily studied in surgically resected tissues, and techniques for three-dimensional (3D) mapping of dense tissues at cellular resolution enable quantitative assessment of PanINs and the pancreatic microenvironment in histological images (7–13). Recent works utilizing a large cohort of 3D reconstructed human pancreata revealed that the pancreata of some individuals contain hundreds of PanIN lesions (14).

This number contrasts with the relative rarity of PDAC and suggests that most PanIN lesions will never progress to cancer in a person's lifetime. The mechanism governing this extensive PanIN initiation and growth in human tissues is poorly understood.

The gold standard for understanding the true incidence and morphology of biological structures is direct measurement of 3D structure in human tissues. However, this approach has some limitations. In animal modeling, researchers maintain direct control over disease progression to pair structural metrics with temporal information. Such control does not exist in studies of human disease, where static time points of disease progression are generated through specimens taken from surgical resections—these samples are precious, albeit rarely longitudinal. Thus, while we can construct large cohorts containing structural information from hundreds of PanIN lesions, we cannot directly measure the “age” of these precursors or observe their dynamic behavior. Here, we utilize a cohort containing metrics from >1000 spatially resolved, 3D PanIN lesions mapped from serial histological sections of pancreatic tissues resected from 48 individuals to present potential growth dynamics of PanINs. Some of these samples contain spatially resolved DNA sequencing data describing the somatic mutations of spatially separate PanINs, providing additional information about their history and inter-relation.

Because PanIN growth cannot be observed directly, we use dynamic modeling to predict suitable growth laws by comparing the resulting size distributions to our experimental size data. This approach allows us to identify fundamental processes contributing to growth. In particular, the spatially resolved genomics information suggests that intraductal spread of PanIN lesions, as well as multiple PanIN lesions fusing together to create large, highly branched structures might be important (14). In the following, we first analyze the experimental data in detail and then build successively more complex models to explain the observed lesion size distribution.

RESULTS

PanIN sizes exhibit a broad distribution

The CODA methodology (15) was successfully used to map PanIN lesions in human pancreas tissues (Fig. 2). CODA enables automatic, nonlinear registration of serial histological sections and deep learning segmentation of pancreas microanatomy for quantification of PanIN

¹Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ²Department of Oncology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD, USA. ³Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD, USA. ⁴Department of Comparative Medicine, Medical University of South Carolina, Charleston, SC, USA. ⁵Department of Pathology, University of Colorado School of Medicine, Aurora, CO, USA. ⁶Max Planck Institute for Dynamics and Self-Organization, Göttingen, Germany. *Corresponding author. Email: kiemen@jhmi.edu (A.L.K.); david.zwicker@ds.mpg.de (D.Z.)

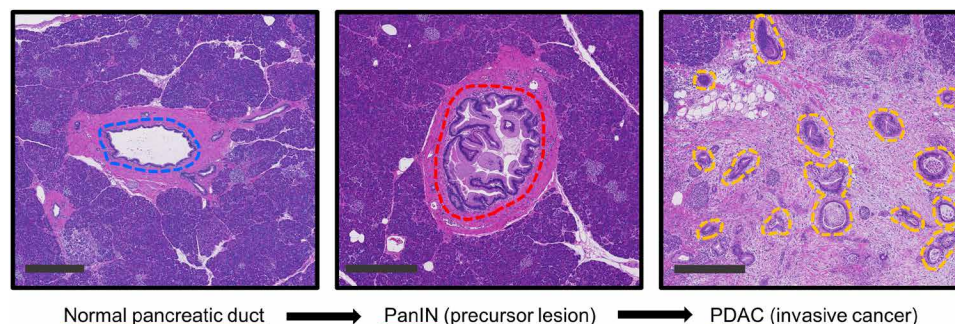


Fig. 1. Pancreatic tumorigenesis as visualized in histological sections. Pancreatic ductal adenocarcinoma (PDAC) develops from histologically recognizable precursor lesions called pancreatic intraepithelial neoplasms (PanINs). Shown here are histological examples of (left) a histologically normal duct, (center) PanIN, and (right) invasive cancer. Scale bar, 0.5 mm.

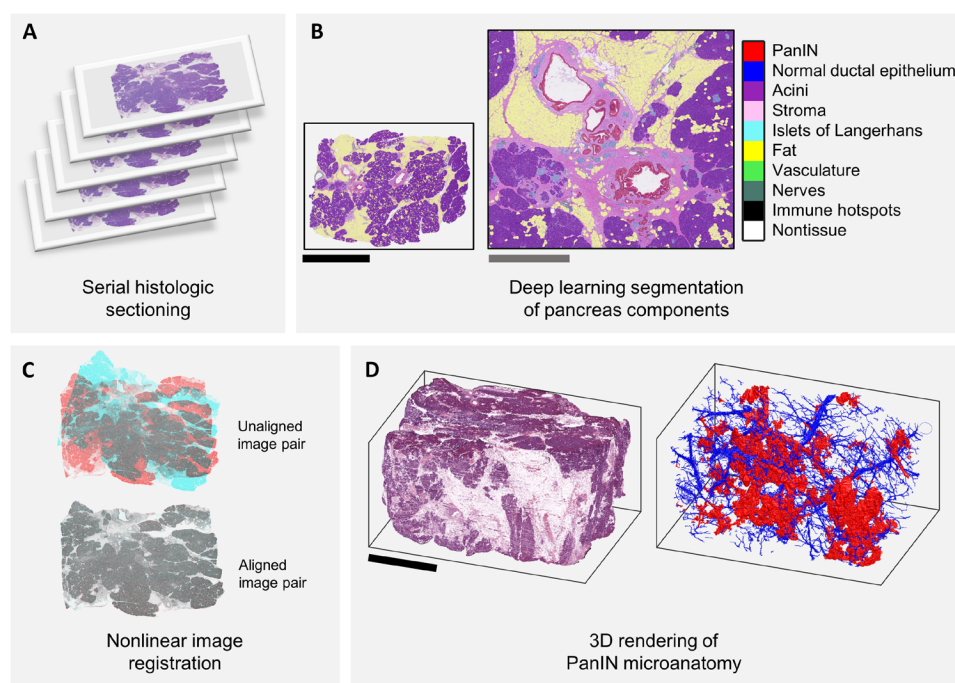


Fig. 2. CODA 3D reconstruction of pancreatic microanatomy. (A) CODA starts with serial histological sectioning of formalin-fixed, paraffin-embedded human pancreas samples. All, or a subset, of sections are stained with hematoxylin and eosin (H&E) and digitized. (B) A deep learning semantic segmentation algorithm was used to segment nine tissue components in the H&E images. (C) A nonlinear image registration algorithm was used to align the serial images into a digital volume. (D) Registered, segmented images were used to create visual and quantifiable maps of the pancreas microanatomy. Scale bars, 1 cm (black) and 2 mm (gray).

size and morphology in 3D space, here performed at a resolution of 12 μm . From each sample, we compiled patient demographic information along with number and size of PanIN lesions per 3D reconstructed surgically resected pancreas sample. Using these data, a range of PanIN sizes and morphologies was found (Fig. 3A). A total of 48 thick slabs of human pancreas tissue were assessed (Fig. 3B). The mean sample volume was 1.83 cm^3 (median: 1.87 cm^3 , range: 0.31 to 3.62 cm^3). Samples contained an average of 21.8 spatially separate PanIN lesions (median: 18.5, range: 4 to 92). PanIN volumes were highly variable within this cohort. The smallest PanIN was $9 \times 10^{-5} \text{ mm}^3$, occupying part of a small, intercalated duct, and the largest PanIN was 24.7 mm^3 , occupying most of the pancreatic ductal system of the sampled region. The average PanIN volume was

0.27 mm^3 (median: 0.01 mm^3). The PanIN structure was similarly highly variable, with small PanIN lesions occupying short regions of single duct branches, and the larger PanIN lesions appearing highly branched, with extension in the pancreatic ducts and into surrounding acinar lobules. Figure 3C displays PanIN densities per sample, calculated as number of PanIN per cubic centimeter of tissue.

The variability of PanIN size is visualized in the histogram shown in Fig. 3D using logarithmic scaling. This representation of the data suggests that PanIN size is distributed according to a power law with a fitted exponent of -1.7 (correlation coefficient 0.96, with 95% confidence intervals 0.89 and 0.98), which implies that PanINs are overwhelmingly small. However, this power law cannot explain the occurrence of the largest PanIN (blue dots in the lower right of

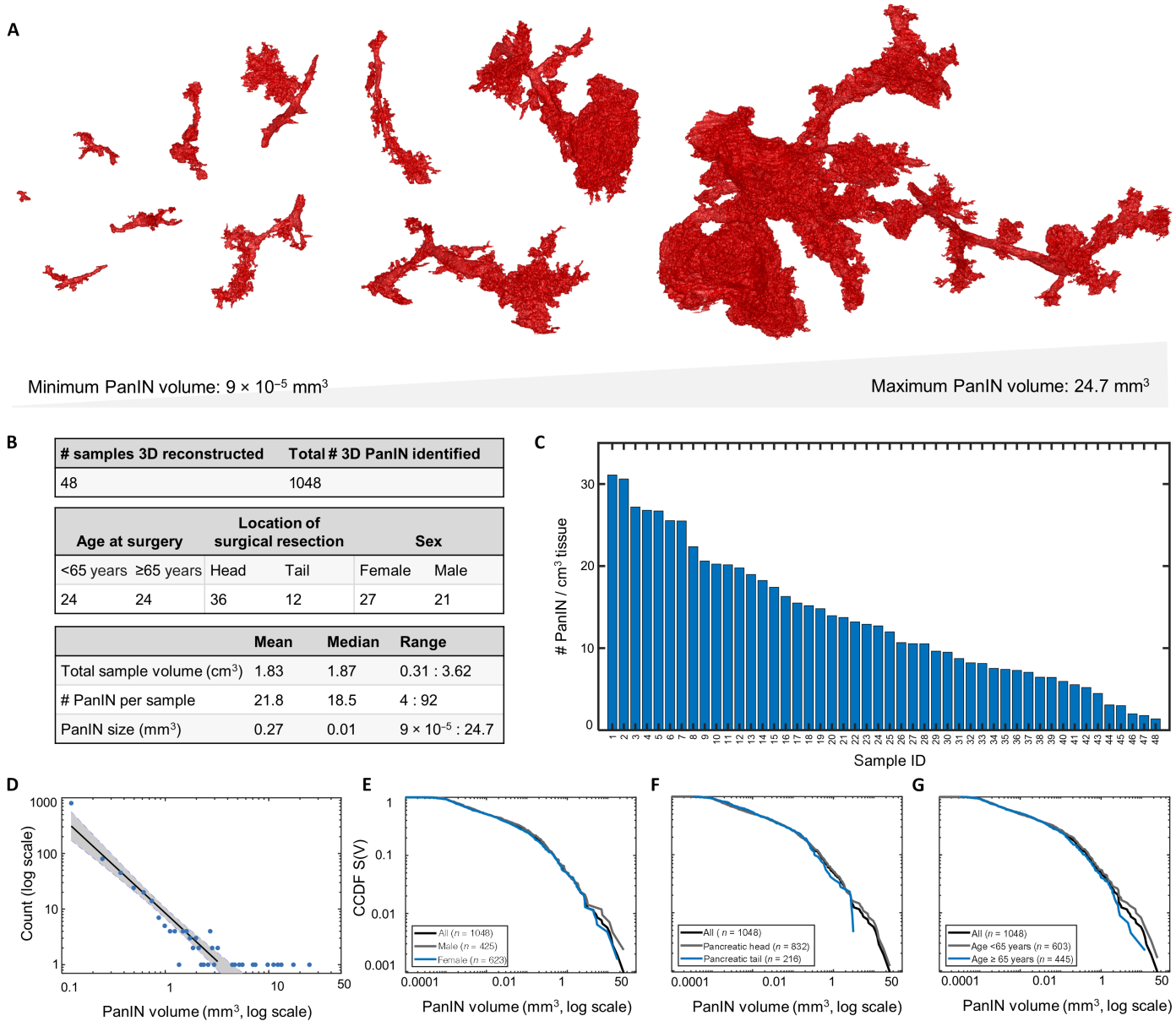


Fig. 3. Observed PanIN sizes exhibit a broad distribution. (A) PanINs were found in a range of volumes, with a minimum PanIN volume of $9 \times 10^{-5} \text{ mm}^3$ and a maximum PanIN volume of 24.7 mm^3 . (B) Tables displaying total number of pancreas samples reconstructed, number of PanIN found, patient demographics, and detailed 3D sample information. (C) Bar graph displaying number of PanINs identified per cubic centimeter of pancreas tissue for 48 grossly normal slabs of human tissue. Minimum of 1.4 PanIN per cubic centimeter of tissue and maximum of 31.1 PanIN per cubic centimeter of tissue. (D) Histogram of PanIN volumes, plotted at logarithmic scale. (E to G) Cumulative distribution function of PanIN volumes, displaying number of PanIN identified per cubic centimeter of pancreas tissue compared across age, location of surgical resection, and sex. All nonsignificant ($P > 0.05$).

Fig. 3D). We hypothesize that this poor fit at large volumes may be due to the merging of multiple PanIN lesions into exceptionally large structures, as suggested by our previous genetic work (14). A precise quantification of very large PanIN is challenging due to the limited number sampled, and this information is further concealed because histograms generally rely on binning of the data. To circumvent this problem, we instead represent the data using a complementary cumulative distribution function (CCDF), $S(V)$, which gives the fraction of observed PanINs with a volume larger than V (see Fig. 3, E to G). The precise shape of $S(V)$ carries more information

about the distribution of PanIN size than the histogram, because it does not require binning. For instance, it reveals that PanINs below $V_{\min} = 0.001 \text{ mm}^3$ are rarely detected, so we disregarded data below this size in our analysis. PanINs larger than this size must have been initiated sometime in the past and grown over the course of the days, months, or years leading up to the time of pancreatic surgical resection. Consequently, the size distribution described by the CCDF contains information about the history of the sample. Using the CCDF, we compared PanIN density across three demographic factors to show that no significant difference in PanIN content exists

as a function of patient age (Fig. 3E), sex (Fig. 3F), or location of surgical resection within this cohort (Fig. 3G), consistent with previous findings (14). Instead, we hypothesize that the size distribution is shaped by the way in which PanINs grow.

A power-law growth law explains size distribution qualitatively

PanIN growth is a complex, poorly understood process, which is likely affected by the pancreatic microenvironment (interactions of cells harboring somatic mutations with stromal cells and pancreatic digestive enzymes) and an individual's age, family history, and lifestyle. However, our comparative analysis suggests that, in this cohort, age (Fig. 3E), sex (Fig. 3F), and the location of surgical resection (tissue taken from the pancreatic head versus tail, Fig. 3G) do not significantly affect the size or incidence of PanINs. It is thus plausible that the overall features of PanIN size distribution are less sensitive to such details and are rather shaped by general growth behavior. For instance, PanINs could grow according to their present size (quantified by their volume V), proportionally to their surface area (which scales as $V^{2/3}$ if the shape varies little), or only along the inner lining of the pancreatic ducts in which they are, by definition, contained. In the last case, growth would be independent of the volume V , which can be captured as a growth rate proportional to V^0 . Together, the PanIN growth rate of all these alternatives can be summarized by a power law

$$\partial_t V = k V_{\min} \left(\frac{V}{V_{\min}} \right)^\alpha \quad (1)$$

which quantifies how the PanIN volume V changes over time t . Here, $V_{\min} = 0.001 \text{ mm}^3$ is the cutoff volume, k quantifies the growth rate, and α denotes the growth exponent distinguishing different modes of growth; $\alpha = 1, 2, 0$ correspond to the three alternative modes discussed above, but in principle all values of α are permissible. Figure 4 (A and B) visualizes the strong influence of the exponent α on PanIN volume as a function of time.

The observed size distribution depends not only on how individual lesions grow but also on when they were initiated. To capture this, we describe a collection of many lesions with individual volumes V_i . We start without any PanIN at $t = 0$ and add new PanIN

with an initiation rate density j , which quantifies the number of lesions initiated per year per cubic centimeter of pancreas tissue. Consequently, the PanIN count N for a sample of volume V_s evolves as

$$\partial_t N = j V_s \quad (2)$$

Together, Eqs. 1 and 2 describe how the PanIN count N and the volumes V_i of each individual PanIN evolve over time. In particular, we can then quantify the CCDF $S(V)$ at the sample age T .

To explore suitable growth exponents α , we start by analyzing the simplest scenario where PanINs are initiated at a constant rate j , and each PanIN grows independently according to Eq. 1. In this scenario, the PanIN count N grows linearly with time, $N(t) = j V_s t$. The predicted size distribution of PanIN volumes V after a finite time T retains the strong dependence on α (see Fig. 4, C and D). We next compare the predictions of the power-law growth model to the observed distribution $S(V)$. Figure 4E shows two fits of this model involving either the entire range of data (green line) or only small PanINs (blue line). This shows that the power-law growth model explains the distribution of smaller PanIN lesions reasonably well, but cannot account for the entire size distribution. This might be expected, because larger PanINs may not simply grow, but may also merge with other PanINs, which is not reflected in the current model. Nevertheless, the fit of the model to smaller PanINs suggests that PanINs grow proportionally to their volume or even more rapidly because the model with $\alpha > 1$ best explains the data. In contrast, the deviation of the distributions for large volumes stems from (i) many more small PanINs than our simple model predicts or (ii) more exceedingly large PanINs than our model predicts. Consequently, not only variability in PanIN initiation but also seeding of new PanIN and merging of older PanIN lesions could explain these deviations. We will show that these scenarios are all plausible, but lead to very different dynamics, which could be discriminated experimentally.

Growth law predicts PanIN initiation times

A core assumption of the first analysis above was that the PanIN initiation rate j was constant in time. We cannot directly test this assumption using our data, which were collected from individuals who underwent surgical resection for a pancreatic abnormality such as cancer (which primarily occurs in older populations). However, it is generally accepted that PanINs are more common in older

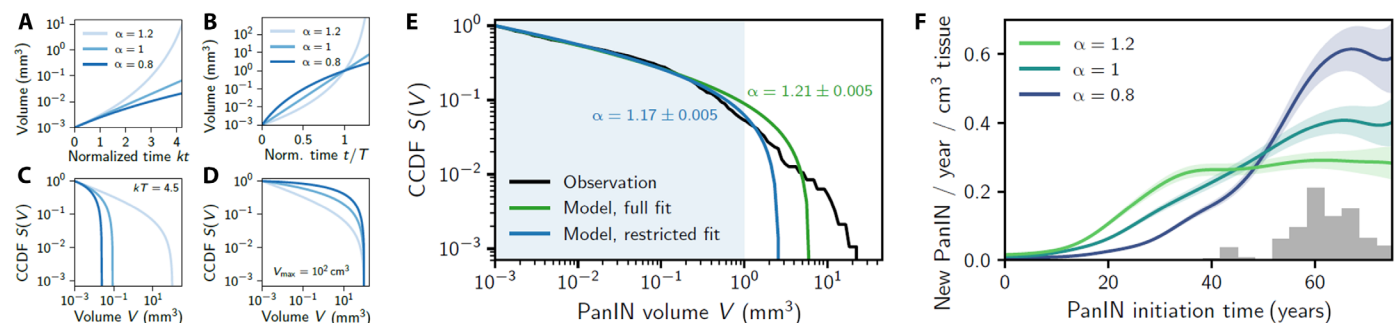


Fig. 4. Simple growth PanIN growth model explains size distribution qualitatively. (A) PanIN volume V as a function of time t predicted by the power-law growth model given by Eq. 1 for various growth exponents α and identical growth rate k . (B) $V(t)$ reaching the same volume at $t = T$ for various α . (C) CCDF $S(V)$ of PanIN volumes predicted by the power-law growth model for a given growth rate k and various α . (D) $S(V)$ with identical maximal volume $V_{\max} = 100 \text{ mm}^3$ for various α . (E) Comparison of observed (black line; same data as Fig. 3F) and predicted (green and blue lines) size distributions $S(V)$. Parameters α and $V_{\max}$ of the power-law growth model were obtained by fitting over all volumes (green data, $\chi^2 = 0.054$) or over the indicated range (blue data, $\chi^2 = 0.018$). (F) Smoothed PanIN initiation rate density j as a function of age inferred using the power-law growth model and the observed PanIN sizes for various α . Shaded area indicates confidence interval of width $j / \sqrt{N}$, where N is the number of PanINs for that year. The samples' ages are summarized by the gray histogram. (A to F) Additional parameters: $V_{\min} = 0.001 \text{ mm}^3$.

individuals and that the somatic genetic events that give rise to PanINs accumulate as we age (16–18). To quantify the PanIN initiation rate indirectly, we use Eq. 1 to predict when a PanIN measured at volume V must have been initiated (with a volume $V_{\min}$) relative to the age T of the sample. For simplicity, we use the same growth rate k for all PanINs, chosen minimally such that no PanINs are older than the age of the patient at the time of pancreatic resection. Together, this allows us to predict the initiation rate density j (the number of PanINs initiated in a given year per cubic centimeter of pancreas tissue) as a function of time. Figure 4F shows that a fairly constant initiation rate density requires super-exponential growth ($\alpha \approx 1.2$, green data), consistent with our result above. In contrast, exponential ($\alpha \approx 1$, teal data) or sub-exponential ($\alpha < 1$, violet data) growth requires strongly increasing initiation rates, e.g., new PanINs must appear more frequently in older samples. To get a deeper insight into the connection between initiation rate density j and growth exponent α , we next discuss two concrete realizations that can cause these different behaviors.

PanIN seeding could explain increasing initiation rates

Increased PanIN initiation rates could potentially be explained by seeding, where some neoplastic cells detach from a PanIN lesion, travel within the lumen of the duct, and initiate a new PanIN that is physically separate from the parent PanIN lesion (see Fig. 5A). Experimental evidence confirms the possibility of intraductal spread, as DNA sequencing has shown that adjacent, spatially separate PanIN sometimes harbor a similar profile of somatic mutations (14, 19). To see whether this explanation is feasible, we extend the

power-law growth model given by Eqs. 1 and 2 to include seeding. For simplicity, we assume that the volume of a PanIN does not change when it seeds a new one, essentially assuming $V \gg V_{\min}$. Seeding can then be captured by the modified initiation rate density

$$j(t) = j_0 + \frac{a}{V_s} \sum_{i=1}^{N(t)} \left[\frac{V_i(t)}{V_{\min}} \right]^\gamma \quad (3)$$

where j_0 is a constant de novo initiation rate density, a quantifies the strength of seeding from each of the N existing PanINs of volumes $\{V_i\}$, and γ is an exponent describing how the seeding depends on the size of the parent PanIN: A constant rate corresponds to $\gamma = 0$, whereas $\gamma = 1$ implies seeding proportional to the volume of the PanIN, and fractional values describe scenarios between these two extremes. Note that the second term in Eq. 3 is proportional to the PanIN count N , causing an autocatalytic increase in the number of PanIN, similar to how metastasis can themselves metastasize and drastically increase the number of metastatic foci.

We simulate a population of PanINs for various choices of the five parameters (k , α , j_0 , a , and γ) of the PanIN seeding model to compare the resulting size distribution to the measured data. Note that two of the five parameters, namely, α and γ , distinguish qualitatively different scenarios, whereas the other three parameters determine the quantitative behavior. To capture this, we analyze the model for various pairs (α, γ) and determine the remaining parameters using a fit to the experimental data. Using χ^2 to quantify the goodness of fit, we can then judge which pair (α, γ) provides the best description of the experimental data. Figure 5B shows that the

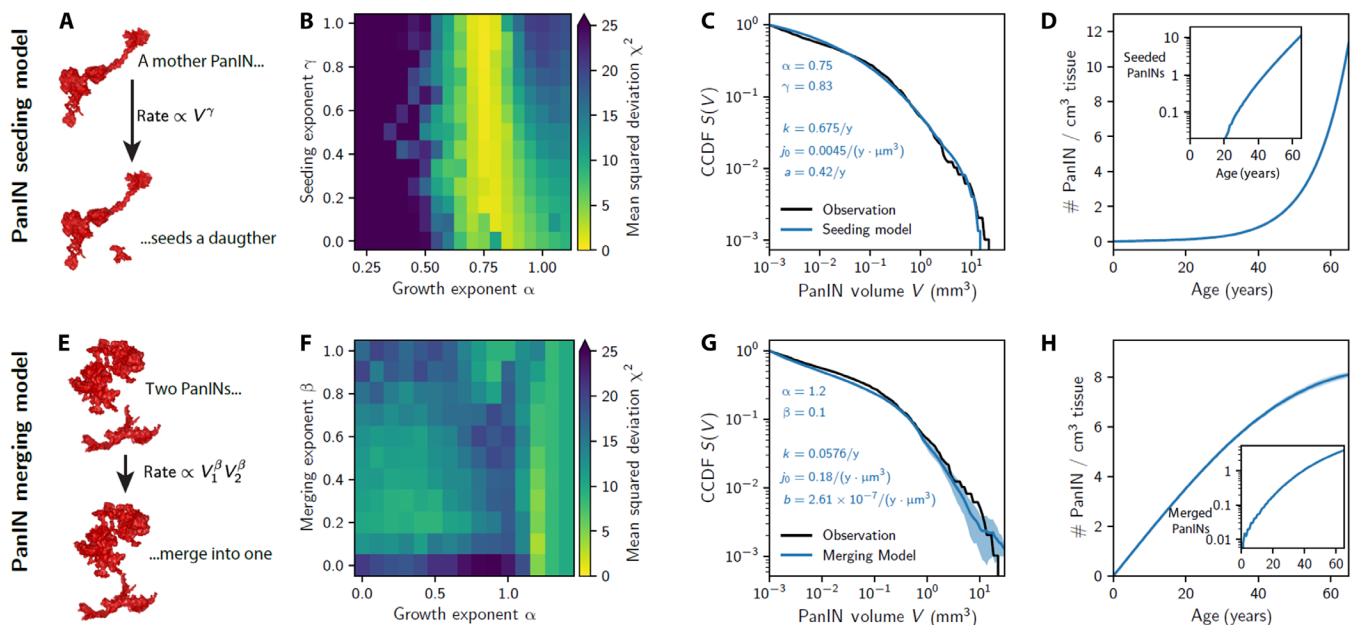


Fig. 5. Seeding and merging models can explain observed size distribution quantitatively. These models combine the simple growth described by Eq. 1 with spontaneous seeding of daughters from older PanINs (A to D) or merging of two PanINs (E to H). (B) Mean squared deviation χ^2 as a function of the growth exponent α and seeding exponent γ indicates that seeding model with $\alpha \approx 0.75$ can explain the observed data. (C) Comparison of the PanIN size distribution $S(V)$ of the seeding model (blue line; $\alpha = 0.75$, $\gamma = 0.83$) to the observed data (black line). The parameters in the inset refer to a sample of volume $V_s = 100 \text{ cm}^3$ simulated for $T = 65$ years. (D) Predicted PanIN count N as a function of age t . Inset shows the number of seeded PanINs as a function of t indicating an exponential increase. (F) χ^2 as a function of α and the merging exponent β indicates that the merging model with $\alpha \approx 1.2$ and $\beta \approx 0.1$ can explain the observed data. (G) Comparison of $S(V)$ of the merging model (blue line; $\alpha = 1.2$, $\beta = 0.1$; shaded area indicates STD for $n = 32$ repetitions) to the observed data (black line). (H) Predicted N as a function of t suggests $N \sim t$. Inset shows the number of merged PanINs as a function of age t .

seeding exponent γ influences χ^2 only weakly, whereas the growth exponent α is strongly constrained by the data. This analysis now suggests that PanINs grow sub-exponentially ($0.6 < \alpha < 0.9$) in contrast to the simpler model without seeding. In any case, the direct comparison of the theoretical prediction with experimental measurements shown in Fig. 5C indicates that seeding can account for the observed data quantitatively. In essence, seeding from existing PanINs leads to an exponentially increasing initiation rate density j (see Fig. 5D), which is consistent with Fig. 4E and accounts for the many observed small PanIN lesions.

PanIN merging could explain frequent large PanINs

Merging events, where two PanINs grow so large that they touch and merge with each other within the effected duct, are a second alternative for a process that affects the size distribution (see Fig. 5E). Experimental evidence supports the existence of polyclonal PanIN lesions, as DNA sequencing has shown that large, highly branched PanIN lesions can contain multiple distinct localized somatic mutations in the same initiating driver gene (14). Instead of describing the intricate details of spatial PanIN growth, we also capture this behavior by extending the power-law growth model given by Eqs. 1 and 2. The main idea is that the probability that two PanINs meet and merge is roughly inversely proportional to sample volume V_S and might also depend on their individual volumes V_1 and V_2 . We thus merge two PanINs stochastically with rate $K(V_1, V_2)$, which we model as a power law

$$K(V_1, V_2) = \frac{b}{V_S} \left(\frac{V_1 V_2}{V_{\min}^2} \right)^\beta \quad (4)$$

where b determines the merging rate, whereas β encodes the size dependence: For $\beta = 0$, the merging rate is independent of PanIN size, whereas, for instance, $\beta = 2/3$ implies a rate that scales with the surface area of both PanINs. This merging model is similar to Smoluchowski's coagulation model, which describes merging clusters like liquid droplets (20, 21). For simplicity, we consider a constant initiation rate density j_0 of de novo formation of PanINs. The model is inherently stochastic, so we simulate multiple samples and collect all PanIN volumes at the final time to compare their distribution to the experimentally measured one. Because we replace two merging PanINs by a single one with the total volume $V_1 + V_2$, this model leads to fewer but larger PanINs over time, which could explain the higher-than-expected portion of large PanINs that we observe.

The PanIN merging model has five parameters (k , α , j_0 , q , and β), where α and β distinguish qualitatively different growth scenarios, whereas k , j_0 , and q set quantitative rates. We thus again fit the rates by minimizing χ^2 as a function of the parameter pair (α , β). Figure 5F indicates that there is an optimal region for these two parameters, although it is less sharply defined than in the seeding model. The best fit occurs for super-exponential growth ($\alpha \approx 1.2$) and a merging rate that is roughly constant ($\beta \approx 0.1$), although larger merging exponents are also plausible. Figure 5G shows that the best fit can indeed explain the observed size distribution, but there is appreciable uncertainty, particularly for the larger PanINs with worse statistics. In any case, merging of PanINs happens predominantly for larger volumes, leading to even larger PanINs, implying that PanIN count decreases with time (see Fig. 5H) and the size distribution becomes skewed toward larger sizes.

Seeding and merging model predict different PanIN counts over time

The seeding and the merging model can both explain the experimentally observed PanIN size distribution. However, the reasons are fundamentally different: The seeding model exhibits a strongly increasing initiation rate, resulting in more small PanINs than the simple power-law growth model predicts. Conversely, the merging model leads to an excess of large PanINs even for a constant initiation rate. Crucially, both models account for the deviation between the power-law growth model and the observed data that we identified in Fig. 4E. Clearly, the combination of both models could also explain the observed experimental data of PanIN sizes. However, both models make distinct predictions for the number N of PanINs as a function of time: The seeding model yields exponentially increasing N (Fig. 5D), due to the exponential increase in the initiation rate, whereas the merging model predicts even fewer PanINs than in the basic growth model due to merger events (Fig. 5H). This difference also explains why the seeding model predicts a lower growth exponent ($\alpha \approx 0.75$) than the merger model ($\alpha \approx 1.2$), which is consistent with our observations in Fig. 4F that a smaller α coincides with strongly increasing initiation rates. Together, the two models could thus be distinguished, and their relative contribution could be quantified, if PanINs were identified in much younger samples.

DISCUSSION

Here, we show that simple growth models can describe experimentally observed size distributions in human pancreatic precancer incidence and volume, whereas the age, sex, and location seem to have a weak influence. We demonstrate that there are two general models of lesion growth that can lead to the experimentally measured size distribution: (i) sub-exponential lesion growth with exponentially increasing initiation rate, e.g., due to discontinuous intraductal spread, and (ii) exponential lesion growth with significant merging of larger lesions, e.g., fused polyclonal PanIN lesions. Both models fit experimentally collected genomic data—likely, a combination of the two models is true [this is studied in related fields as coagulation-fragmentation processes (22)].

Although both mechanisms lead to the same measured PanIN size distribution at their endpoints, the dynamics of the two are very different. This is apparent in the predicted de novo initiation rate densities j_0 , which differ by more than two orders of magnitude (Fig. 5, C and G), the number of lesions as a function of time (Fig. 5, D and H), and in the lesion size distribution as a function of time (fig. S1). The PanIN seeding model exhibits sub-exponential growth of individual PanIN lesions, but the number of PanINs grow exponentially because more PanINs can, in turn, seed more PanINs. Conversely, the PanIN merging model requires super-exponential growth of individual PanIN lesions, but the number of PanIN lesions actually decreases over time as multiple PanIN combine into one. Because we do not observe significant differences in PanIN counts between two age groups (Fig. 3D), the merging model might explain real PanIN growth more accurately. However, reality might be best described by a combination of seeding, merging, and a time-dependent de novo initiation rate density. More detailed data, particularly from samples from younger individuals, are needed to quantify the relative contributions of these different processes.

We note several limitations of our study. As we analyze all PanINs from all 3D samples reconstructed, our PanIN volumetric data were biased by non-fully contained lesions (PanIN that were cut at the boundaries and should thus be larger than we measure). If we were to analyze only the fully contained PanIN, we would lose all the largest lesions, shifting our distribution. In the future, larger sample volumes could circumvent this problem. Because of these challenges, the numbers obtained from the model should be interpreted carefully. However, the general relations between initiation, merging, and the growth exponent would still hold. In addition, as the volumetric PanIN data generated by CODA was limited by the resolution of the histological staining schema thickness [4- μm -thick serial sections with every third section stained with hematoxylin and eosin (H&E)] (14), the resolution of the experimental data analyzed here was 12 μm by 12 μm by 12 μm , which may have limited our ability to accurately measure the smallest PanIN lesions. Last, as the pancreas samples analyzed here were collected during surgical resections for pancreatic abnormalities, the incidence and size of lesions reported here may not fully represent the general population (as most of our samples came from older individuals, and there is an association between age, pancreatic cancer, and PanIN incidence). Future work modeling the growth properties of PanIN as measured from pancreata lacking any abnormality and samples from younger individuals is important for correcting this bias. Last, in this study, we assume that all PanINs grow at the same rate. As knowledge of precursor lesions is constantly evolving, it is highly possible that more nuanced information will reveal different growth rates for PanIN from different populations [such as PanINs in individuals with mutations that put them at high risk for development of pancreatic cancer (BRCA) or high-grade PanIN] grow at different rates. However, we believe that it was a reasonable assumption that all low-grade PanINs grow at similar rates given the genetic similarities between all low-grade PanINs, which are primarily driven by KRAS hotspot mutations (23–25).

Our model gives a general overview for how precancerous lesions could evolve. More detailed experimental data, such as high-resolution spatial genetic information for both very small PanIN lesions and very large PanIN lesions, would be valuable to measure seeding and merging rates directly. Similarly, more data on PanIN sizes and shapes from samples of various ages could be used to directly test different growth models of individual PanIN, e.g., whether they grow in 2D sheets along the pancreatic ducts or have some 3D leeway to expand their volume in all directions (for example, into the lumen of the ducts or into the acinar lobules), which will likely also depend on PanIN size. If such data become available, our model can serve as a basis for developing more detailed models that describe PanIN in the actual physical space provided by the pancreatic ducts. Moreover, our generic approach to describing lesion growth is likely transferable to other lesion types, including other common cancer precursors in the fallopian tubes or esophagus. Differences and similarities between different precancerous lesions could then unveil universal principles of how cancers originate.

MATERIALS AND METHODS

Experiments

Generation of a 3D human pancreas tissue cohort

The 3D pancreas maps used here were previously described in a work mapping the prevalence and spatially resolved genomic properties of

pancreatic cancer precursor lesions (14). Briefly, thick slabs of grossly normal human pancreas tissue were collected from 48 individuals who underwent surgical resection at the Johns Hopkins Hospital for pancreatic abnormalities including PDAC, well-differentiated pancreatic neuroendocrine tumors, metastatic disease of non-pancreatic origin, and non-malignant pathologies. Tissue was formalin-fixed, paraffin-embedded (FFPE), and serially sectioned at a thickness of 4 μm . Every third section was stained with H&E and digitized at 20 $\times$ magnification, for a lateral (xy) resolution of 0.5 μm per pixel and an axial (z) resolution of 12 μm . CODA (Fig. 2), a recently developed tool for 3D reconstruction of serially sectioned tissues (15), was used to register the serial images and segment nine pancreatic microanatomical structures on the serial H&E images at a reduced resolution of 2 μm : PanIN, normal pancreatic ducts, pancreatic acini, islets of Langerhans, vasculature, nerves, fat, lymph nodes, and stroma to an accuracy of 96.6% (14). Resulting models were fully visualizable and quantifiable. Spatially distinct PanINs identified using CODA were validated through inspection of corresponding histology, and parameters including number of PanIN lesions per cubic centimeter of pancreas tissue, lesion size, cellularity, and aspect ratio were obtained.

Power-law growth model

The power-law growth model given by Eq. 1 in Results with initial condition $V(t = 0) = V_{\min}$ results in the growth curve

$$V(t) = V_{\min} [1 + kt(1 - \alpha)]^{\frac{1}{1-\alpha}}$$

for $\alpha \neq 1$, and $V(t) = V_{\min} e^{kt}$ for $\alpha = 1$, where we use $V_{\min} = 0.001 \text{ mm}^3$. The maximal PanIN size in a sample of age T is thus $V_{\max} = V(T)$, which is achieved when the PanIN is initiated at $t = 0$. Assuming a constant PanIN initiation rate density j , the distribution of PanIN sizes can be expressed by the CCDF after duration T , which reads

$$S(V) = \frac{1 - \left(\frac{V_{\max}}{V}\right)^{\alpha-1}}{1 - \left(\frac{V_{\max}}{V_{\min}}\right)^{\alpha-1}}$$

for $\alpha \neq 1$, and $S(V) = \ln(V_{\max}/V) / \ln(V_{\max}/V_{\min})$ for $\alpha = 1$. These expressions do not depend on j because the distribution is normalized, $S(V_{\min}) = 1$. To compare the power-law growth model to observed data, we predict the distribution $S(V)$ for each individual sample, taking its age T into account. The overall distribution, determined by averaging weighted by sample volume, then only depends on the parameters k and α . We determine these parameters by minimizing the squared difference between the predicted and measured $S(V)$ using the Levenberg-Marquardt algorithm (in Fig. 4E).

Inverting the growth curve $V(t)$ allows us to determine the duration it takes for a PanIN to grow to the observed volume V , so we can predict when a PanIN must have been initiated for a given k and α . Choosing the minimal plausible value for k (such that PanINs must have originated after $t = 0$) as well as pooling observed data by years and weighing them with the inverse sample volume, we predict the number N_i of PanINs that were initiated in year i per unit volume of the sample. These discretized data are smoothed with a Gaussian filter of width 5 years to generate Fig. 4F.

Simulating extended models

We simulate the seeding model by explicitly propagating forward in time a collection of PanIN sizes $\{V_i\}$. For each step, we first use Eq. 3 (see Results) to determine the average waiting time Δt until a new PanIN is initiated and then grow all PanINs for this duration according to the power-law growth curve given above. We quantify the resulting distribution $S^{\text{pred}}(V)$ from the final volumes after time T .

In contrast, we simulate the merging model using a fixed time step $\Delta t = 0.01 \gamma$. During each step, we first grow all PanINs according to the power-law growth given by Eq. 1, then randomly choose M of the $N_{\text{pairs}} = N(N - 1)/2$ possible PanIN pairs for an attempted merge, and lastly initiate new PanIN with the constant initiation rate j_0 . The merge is performed stochastically, i.e., when $P_{ij} > \xi$, where $P_{ij} = \Delta t \cdot N_{\text{pairs}} K(V_i, V_j)/M$ with $K(V_1, V_2)$ given by Eq. 4, and ξ is a random number chosen uniformly between 0 and 1. Here, M is a control parameter, which is chosen minimally while still obeying $P_{ij} < 1$. While initiation is still implemented deterministically, merging is done stochastically; thus, we obtain the respective distribution $S^{\text{pred}}(V)$ from an average of eight independent runs. In all cases, we run simulations for $T = 65 \gamma$, the median age of the patients analyzed. The volume V_S of the model sample does not affect results and we chose $V_S = 100 \text{ cm}^3$ to get adequate statistics.

Fitting of simulated models

To fit the predictions of the models described above to the observed data $S^{\text{obs}}(V)$, we minimize the logarithmically scaled mean squared deviation

$$\chi^2 = \sum_{i=1}^{64} \left[\ln \frac{S_i^{\text{obs}}}{S_i^{\text{pred}}} \right]^2$$

where the CCDFs $S(V)$ are evaluated in 64 logarithmically distributed intervals between $V_{\min}$ and the maximally observed volume; thus, $S_i = P(V > V_i)$ gives the fraction of PanIN with a volume above V_i . We minimize χ^2 by adapting the model parameters using the differential evolution algorithm (26) over eight independent repetitions, each with 2048 steps (in Fig. 5).

Supplementary Materials

This PDF file includes:

Sections S1 to S3

Fig. S1

References

REFERENCES AND NOTES

- R. L. Siegel, K. D. Miller, N. S. Wagle, A. Jemal, Cancer statistics, 2023. *CA Cancer J. Clin.* **73**, 17–48 (2023).
- L. Rahib, B. D. Smith, R. Aizenberg, A. B. Rosenzweig, J. M. Fleshman, L. M. Matrisian, Projecting cancer incidence and deaths to 2030: The unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74**, 2913–2921 (2014).
- L. Rahib, M. R. Wehner, L. M. Matrisian, K. T. Neale, Estimated projection of US cancer incidence and death to 2040. *JAMA Netw. Open* **4**, e214708 (2021).
- B. L. Mazer, J. W. Lee, N. J. Roberts, L. C. Chu, A. M. Lennon, A. P. Klein, J. R. Eshleman, E. K. Fishman, M. I. Canto, M. G. Goggins, R. H. Hruban, Screening for pancreatic cancer has the potential to save lives, but is it practical? *Expert Rev. Gastroent.* **17**, 555–574 (2023).
- O. Basturk, S. M. Hong, L. D. Wood, N. V. Adsay, J. Albores-Saavedra, A. V. Biankin, L. A. A. Brosens, N. Fukushima, M. Goggins, R. H. Hruban, Y. Kato, D. S. Klimstra, G. Klöppel, A. Krasinskas, D. S. Longnecker, H. Matthaei, G. J. A. Offerhaus, M. Shimizu, K. Takaori, B. Terris, S. Yachida, I. Esposito, T. Furukawa, Baltimore Consensus Meeting, A revised classification system and recommendations from the baltimore consensus meeting for neoplastic precursor lesions in the pancreas. *Am. J. Surg. Pathol.* **39**, 1730–1741 (2015).
- A. L. Kiemen, L. Dequiedt, Y. Shen, Y. Zhu, V. Matos-Romero, A. Forjaz, K. Campbell, W. Dhana, T. Cornish, A. M. Braxton, P. H. Wu, E. K. Fishman, L. D. Wood, D. Wirtz, R. H. Hruban, PanIN or IPMN? Redefining lesion size in 3 dimensions. *Am. J. Surg. Pathol.* **48**, 839–845 (2024).
- J. R. Lin, S. Wang, S. Coy, Y. A. Chen, C. Yapp, M. Tyler, M. K. Nariya, C. N. Heiser, K. S. Lau, S. Santagata, P. K. Sorger, Multiplexed 3D atlas of state transitions and immune interaction in colorectal cancer. *Cell* **186**, 363–381.e19 (2023).
- A. L. Kiemen, A. I. Damanakis, A. M. Braxton, J. He, D. Laheru, E. K. Fishman, P. Chames, C. Almagro-Perez, P. Wu, D. Wirtz, L. D. Wood, R. H. Hruban, Tissue clearing and 3D reconstruction of digitized, serially sectioned slides provide novel insights into pancreatic cancer. *Med.* **4**, 75–91 (2023).
- J. T. C. Liu, A. K. Glaser, K. Bera, L. D. True, N. P. Reder, K. W. Eliceiri, A. Madabhushi, Harnessing non-destructive 3D pathology. *Nat. Biomed. Eng.* **5**, 203–218 (2021).
- D. S. Richardson, J. W. Lichtman, Clarifying tissue clearing. *Cell* **162**, 246–257 (2015).
- A. L. Kiemen, Y. Choi, A. M. Braxton, C. Almagro Perez, S. Graham, M. P. Grahm, N. Nanda, N. Roberts, L. Wood, P. Wu, R. H. Hruban, D. Wirtz, Intraparenchymal metastases as a cause for local recurrence of pancreatic cancer. *Histopathology* **82**, 504–506 (2023).
- A. L. Kiemen, M. Dbouk, E. A. Diwan, A. Forjaz, L. Dequiedt, A. Baghdadi, S. P. Madani, M. P. Grahm, C. Jones, S. Vedula, P. Wu, D. Wirtz, S. Kern, M. Goggins, R. H. Hruban, I. R. Kamel, M. I. Canto, Magnetic resonance imaging-based assessment of pancreatic fat strongly correlates with histology-based assessment of pancreas composition. *Pancreas* **53**, e180–e186 (2024).
- A. Forjaz, E. Vaz, V. M. Romero, S. Joshi, A. M. Braxton, A. C. Jiang, K. Fujikura, T. Cornish, S.-M. Hong, R. H. Hruban, P. H. Wu, L. D. Wood, A. L. Kiemen, D. Wirtz, Three-dimensional assessments are necessary to determine the true spatial tissue composition of diseased tissues. *bioRxiv* 569986 [Preprint] (2022). <https://doi.org/10.1101/2023.12.04.569986>.
- A. M. Braxton, A. L. Kiemen, M. P. Grahm, A. Forjaz, J. Parksong, J. Mahesh Babu, J. Lai, L. Zheng, N. Niknafs, L. Jiang, H. Cheng, Q. Song, R. Reichel, S. Graham, A. I. Damanakis, C. G. Fischer, S. Mou, C. Metz, J. Granger, X. D. Liu, N. Bachmann, Y. Zhu, Y. Liu, C. Almagro-Perez, A. C. Jiang, J. Yoo, B. Kim, S. Du, E. Foster, J. Y. Hsu, P. A. Rivera, L. C. Chu, F. Liu, E. K. Fishman, A. Yuille, N. J. Roberts, E. D. Thompson, R. B. Scharpf, T. C. Cornish, Y. Jiao, R. Karchin, R. H. Hruban, P. H. Wu, D. Wirtz, L. D. Wood, 3D genomic mapping reveals multifocality of human pancreatic precancers. *Nature* **629**, 679–687 (2024).
- A. L. Kiemen, A. M. Braxton, M. P. Grahm, K. S. Han, J. M. Babu, R. Reichel, A. C. Jiang, B. Kim, J. Hsu, F. Amoa, S. Reddy, S. M. Hong, T. C. Cornish, E. D. Thompson, P. Huang, L. D. Wood, R. H. Hruban, D. Wirtz, P. H. Wu, CODA: Quantitative 3D reconstruction of large tissues at cellular resolution. *Nat. Methods* **19**, 1490–1499 (2022).
- E. S. Carpenter, A. M. Elhossiny, P. Kadiyala, J. Li, J. McGue, B. D. Griffith, Y. Zhang, J. Edwards, S. Nelson, F. Lima, K. L. Donahue, W. Du, A. C. Bischoff, D. Alomari, H. R. Watkoske, M. Mattea, S. The, C. E. Espinoza, M. Barrett, C. J. Sonnenday, N. Olden, C. T. Chen, N. Peterson, V. Gunchick, V. Sahai, A. Rao, F. Bednar, J. Shi, T. L. Frankel, M. Pasca di Magliano, Analysis of donor pancreata defines the transcriptomic signature and microenvironment of early neoplastic lesions. *Cancer Discov.* **13**, 1324–1345 (2023).
- B. Milholland, A. Auton, Y. Suh, J. Vijg, Age-related somatic mutations in the cancer genome. *Oncotarget* **6**, 24627–24635 (2015).
- R. A. Risques, S. R. Kennedy, Aging and the rise of somatic cancer-associated mutations in normal tissues. *PLOS Genet.* **14**, e1007108 (2018).
- A. M. Makohon-Moore, K. Matsukuma, M. Zhang, J. G. Reiter, J. M. Gerold, Y. Jiao, L. Sikkema, M. A. Attiyeh, S. Yachida, C. Sandone, R. H. Hruban, D. S. Klimstra, N. Papadopoulos, M. A. Nowak, K. W. Kinzler, B. Vogelstein, C. A. Iacobuzio-Donahue, Precancerous neoplastic cells can move through the pancreatic ductal system. *Nature* **561**, 201–205 (2018).
- S. Cueille, C. Sire, Droplet nucleation and Smoluchowski's equation with growth and injection of particles. *Phys. Rev. E* **57**, 881–900 (1998).
- M. von Smoluchowski, Three presentations on diffusion, molecular movement according to Brown and coagulation of colloid particles. *Phys. Z* **17**, 557–571 (1916).
- C. M. Sorensen, H. X. Zhang, T. W. Taylor, Cluster-size evolution in a coagulation-fragmentation system. *Phys. Rev. Lett.* **59**, 363–366 (1987).
- A. D. Singhi, L. D. Wood, Early detection of pancreatic cancer using DNA-based molecular approaches. *Nat. Rev. Gastro. Hepat.* **18**, 457–468 (2021).
- M. Kanda, H. Matthaei, J. Wu, S. M. Hong, J. Yu, M. Borges, R. H. Hruban, A. Maitra, K. Kinzler, B. Vogelstein, M. Goggins, Presence of somatic mutations in most early-stage pancreatic intraepithelial neoplasia. *Gastroenterology* **142**, 730–733.e9 (2012).
- S. M. Hong, A. Vincent, M. Kanda, J. Leclerc, N. Omura, M. Borges, A. P. Klein, M. I. Canto, R. H. Hruban, M. Goggins, Genome-wide somatic copy number alterations in low-grade PanINs and IPMNs from individuals with a family history of pancreatic cancer. *Clin. Cancer Res.* **18**, 4303–4312 (2012).
- R. Storn, K. Price, Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **11**, 341–359 (1997).
- S. Cueille, C. Sire, Smoluchowski's equation for cluster exogenous growth. *Europhys. Lett.* **40**, 239–244 (1997).

Acknowledgments

Funding: We acknowledge the following sources of support: U54CA268083 (D.W., L.D.W., R.H.H., A.L.K., and P.-H.W.); U54AR081774 (D.W., A.L.K., and P.-H.W.); UG3CA275681 (P.-H.W., D.W., A.L.K., and L.D.W.); Rolfe Pancreatic Cancer Foundation (A.L.K.); Lustgarten Foundation-AACR Career development award for pancreatic cancer research in honor of Ruth Bader Ginsburg (A.L.K.); The Carl and Carol Nale Fund for Pancreatic Cancer Research (A.L.K.); The Sol Goldman Pancreatic Cancer Research Center (A.L.K.); The Joseph C. Monastra Foundation for Pancreatic Cancer Research; The Rolfe Foundation (A.L.K.); Susan Wojcicki and Denis Troper (A.L.K.); and Max Planck Society (D.Z.). **Author contributions:** A.L.K., D.W., and D.Z. conceived the project. A.M.B., T.C.C., and R.H.H. prepared the histological specimens. A.L.K. conducted the histological image analysis. D.Z. conducted the growth simulations. L.D.W., P.-H.W., D.W., and R.H.H. oversaw biological aspects of the project. A.L.K. and D.Z. wrote the first draft of the manuscript, which all authors edited and approved. **Competing interests:** A pending patent application “Computational Techniques for

Three-Dimensional Reconstruction and Multi-labeling of Serially Sectioned Tissue” was filed on 24 June 2022 by A.L.K., R.H.H., P.-H.W., D.W., and L.D.W. The other authors declare that they have no competing interests. **Data and materials statement:** All data needed to evaluate the conclusions in the paper are present in the paper and in the following repositories. The software used to create the PanIN growth-law models is available at the following URL: <https://zenodo.org/doi/10.5281/zenodo.11045634>. The raw data used to create these models is available at the following URL: <https://zenodo.org/records/11087929>. As an additional resource, the software to create the 3D CODA maps is available at the following URL: <https://github.com/ashleylk/CODA>.

Submitted 5 February 2024

Accepted 25 June 2024

Published 26 July 2024

10.1126/sciadv.ado5103