nature portfolio

Peer Review File

Genomic evolution of pancreatic cancer at single-cell resolution

Corresponding Author: Dr Christine Iacobuzio-Donahue

This manuscript has been previously reviewed at another journal. This document only contains information relating to versions considered at Nature Genetics.

Version 0:

Decision Letter:

19th Feb 2025

Dear Dr Iacobuzio-Donahue,

First, please accept my sincere apologies for the delay in returning this decision to you. Thank you so much for your patience.

Your Article, "Genomic evolution of pancreatic cancer at single-cell resolution" has now been seen by 3 referees. You will see from their comments below that while they find your work of interest, some important points are raised. We are interested in the possibility of publishing your study in Nature Genetics, but would like to consider your response to these concerns in the form of a revised manuscript before we make a final decision on publication.

We therefore invite you to revise your manuscript taking into account all reviewer and editor comments. Please highlight all changes in the manuscript text file. At this stage we will need you to upload a copy of the manuscript in MS Word .docx or similar editable format.

We are committed to providing a fair and constructive peer-review process. Do not hesitate to contact us if there are specific requests from the reviewers that you believe are technically impossible or unlikely to yield a meaningful outcome.

When revising your manuscript:

*1) Include a “Response to referees” document detailing, point-by-point, how you addressed each referee comment. If no action was taken to address a point, you must provide a compelling argument. This response will be sent back to the referees along with the revised manuscript.

*2) If you have not done so already please begin to revise your manuscript so that it conforms to our Article format instructions, available
here.
Refer also to any guidelines provided in this letter.

*3) Include a revised version of any required Reporting Summary: https://www.nature.com/documents/nr-reporting-summary.pdf
It will be available to referees (and, potentially, statisticians) to aid in their evaluation if the manuscript goes back for peer review.
A revised checklist is essential for re-review of the paper.

Please be aware of our guidelines on digital image standards.

EXTENDED DATA FIGURES

When re-submitting your manuscript, please ensure that any supplementary figures and tables that are crucial to the manuscript's conclusions are converted into Extended Data figures and tables to increase visibility of these data. Extended Data figures and tables are online-only (present in the online PDF and full-text HTML versions of the paper), peer-reviewed display items that provide essential background to the article but are not included in the main article due to space constraints. A maximum of ten Extended Data display items (figures and tables) is permitted.

Please use the link below to submit your revised manuscript and related files:

Link Redacted

Note: This URL links to your confidential home page and associated information about manuscripts you may have submitted, or that you are reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage.

We hope to receive your revised manuscript within four to eight weeks. If you cannot send it within this time, please let us know.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further.

Nature Genetics is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as 'corresponding author' on published papers create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System (MTS), prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the MTS by clicking on 'Modify my Springer Nature account'. For more information please visit please visit www.springernature.com/orcid.

We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Sincerely,

Safia Danovi, PhD
Senior Editor, Nature Genetics
ORCiD: 0009-0007-7822-5479

Referee expertise:

Referee #1: somatic mutations, genomics

Referee #2: pancreatic cancer, evolution

Referee #3: pancreatic cancer, single-cell, computational

Reviewers' Comments:

Reviewer #1 (Remarks to the Author):

In a manuscript entitled "Genomic evolution of pancreatic cancer at single-cell resolution", Zhang et al. describe the analysis of ~140K single nuclei, representing 72 samples from 24 pancreatic cancer patients. The nuclei were analyzed on the commercially available Tapestri platform, using a panel of ~ 600 amplicons targeting common pancreatic cancer driver variants. The authors begin their narrative with a broad description of the genetic characteristics of the cohort, then comment on combinations of mutations found in patients with (presumed?) BRCA2 germline mutation, then explore CNV heterogeneity and report common alterations resulting in TGF-beta inactivation. Finally, a few points are made about evolution during metastasis and the paper closes with a few data points from longitudinally collected samples.

I think the data set underlying this paper is very interesting, and I appreciate the authors' deep expertise in pancreatic cancer evolution and computational biology. I also think that the manuscript is written in a clear and accessible manner and has a logical structure. My main concerns relate to the substance and novelty of the findings, as well as to the narrative style, which largely consists of a succession of patient-level anecdotes but lacks any cohort-wide quantitative analysis that would provide new insights into how pancreatic cancer evolves.

It appears that the authors' main novelty argument is that the single cell sequencing assay applied to this cohort – unsurprisingly – has improved sensitivity for picking up rare subclonal mutations, this is contrasted with general "bulk sequencing", although the authors do not specifically explain which kind of bulk sequencing they mean (e.g. error-corrected ultra-deep bulk sequencing could easily outperform the Tapestri assay for detection of rare mutations). It makes sense that sequencing more and more deeply or on the single cell level will reveal ever rarer mutations, but why is this important? The authors largely fail to make a convincing case in my opinion. They do mention that some rare subclonal variants might lead

to therapy failure, but this is not at all demonstrated in the manuscript, and perhaps it would be hard to do so with a sequencing panel that is focused on variants that are essential in PDAC (and thus will largely be truncal and cannot be under selection during therapy).

The majority of the paper consists of mini-vignettes that describe the evolution of specific mutation combinations in particular patients (e.g. “PC01’s phylogeny showed that the somatic BRCA2 nonsense mutation was truncal, followed by several other LOH events and CDKN2A homozygous deletion” or “The continuous evolution also manifests as LOH of somatic driver mutation loci in PC10. Two TGFBR1 nonsense mutations were subclonally gained in one of the two main lineages of metastatic clones (Figure 3D).”). I do not doubt that the authors were able to ascertain these facts from a technical perspective, but I am not sure what we can learn from this. Some mutations will be clonal, and others will be selected subclonally, I think the principle is well-established. Particular pathways (namely those included in the sequencing panel based on prior knowledge) appear to be under selection, which is expected. But I cannot discern where the paper accomplishes the crucial task of abstracting the many patient-level findings into any kind of general rule that could be considered novel or memorable.

This weakness stands out starkly throughout the paper, but particularly in the metastasis evolution and longitudinal sampling sections. The main summary figure of the metastasis evolution section (Fig. 5B) tells us that the authors did not observe any quantifiable differences in diversity among metastatic sites, or in divergence from the primary, and could find no correlation between metastasis genetics and spatial localization. What do we learn here? That PDAC metastasis does not follow any discernible rules or that the resolution of the sequencing assay is too limited to relevant effects? It is unclear to me as a reader.

In summary, I very much like the underlying data set, but I struggle to identify the biological insights that would make the manuscript of interest to a broad audience. I would encourage the authors to transcend the rather anecdotal patient-level narrative and aim to expose broad patterns in the data which could then hopefully be reported as genuinely novel insights into PDAC evolution.

Reviewer #2 (Remarks to the Author):

This manuscript by Dr Iacobutio-Donahue and colleagues provides an in-depth analysis of clonal change in pancreatic cancer, by utilizing single cell resolution DAN sequencing. The manuscript builds on a large tissue database and unique samples available to the investigator, such as rapid autopsy samples and longitudinal samples from the same patients pre and post-surgery. Using a custom panel of 596 genes, the authors evaluate mutations, copy number alterations across tumors, including analysis across space (primary vs metastases, and different metastatic loci) and time (over the course of treatment). The findings offer an in-depth view in the genomic landscape. Of pancreatic cancer, building on this group past track record. Of additional interest, the authors include samples from BRCA mutant pancreatic cancer, providing new information on the molecular evolution of disease in the context of BRCA loss. Overall, this manuscript provides a greatly needed resource and will be a go-to reference for future work. I only have some minor comments to note:

1) Presumably non tumor cells (components of the stroma) were captured in the sequencing samples. If so, it would be interesting to analyze whether they have any mutation in pancreatic cancer drivers, as that is a still debated concept in the field.
2) Recent work on PanIN, pancreatic cancer precursor lesions, has reported the presence of multiple different mutations in the KRAS gene within the same lesion. It would be interesting to know whether multiple mutations in KRAS were detected in any of the tumors, even if only in a small subset of clones.
3) The cartoons provided in Figures 1 and 5 are informative, but an additional summary cartoon that shows the key findings of the study to a broader readership would further enhance the manuscript’s impact.

Reviewer #2 (Remarks on code availability):

Reviewing the code goes beyond my area of expertise

Reviewer #3 (Remarks to the Author):

The manuscript describes application of snDNA-seq of a moderately-sized PDAC cohort to investigate its evolutionary genomics. The work seems like a logical follow-up to the authors' earlier paper by Zhang et al. (2023). Some of the observations recapitulate known findings, but the focus on evolution and lineage tracking reveals some interesting connections among cell groups. One concern is that the confidence of these copy number calls the authors relied on heavily for making clonal assignments and other further conclusions. Also, there are a number of comments related to methods, statements, and presentation.
1. The cancer cell fraction (CCF) is a useful metric, but I can think of several issues.
A. The denominator is set by the number of cells having KRAS mutations, which is presumed to be a founder and thus in all tumor cells in a sample.
B. Are there true tumor cells lacking KRAS mutation, and could these even possibly have the clonal mutation in question? This would also affect CCF.

Since CCF is based on a sample (in the statistical sense) of all the cells in a tumor mass, one could even think about CCF error bars, though I think that is beyond the scope of this work. Some comments on this overall issue in the manuscript would be helpful.
2. How do you choose/justify a CCF cutoff to distinguish clonal drivers (CCF ~0.9) versus sub-clonal drivers (CCF ~0.6)? If CCF is not totally immune to the effects just mentioned, could this be a more subtle process of distinguishing than you are accounting for? Does the choice of cutoff affect your results?
3. There are other hard-coded cut-offs baked into the analysis, for example the positive mutation numbers quoted in the paper. Without having more explanation available as to how these are chosen, it leads to the question of whether any changes in these values would affect the scientific results. It would add confidence if we were to know why these are the best values, or if some of the analyses could be repeated with other values, showing that results are not overly sensitive to their choices.
4. (lines 124-126) and (lines 137-139, 148-150) higher alteration of CDKN2A and SMAD4 and attribution to higher sensitivity than TCGA bulk: You place the sensitivity of bulk at 2%, but I don't think bulk is traditionally thought of as being reliably this sensitive. Does this estimation consider issues of depth, purity, etc? The required minimum depth 4 reads for detection that you quote would suggest that sensitivity is a function of depth.
5. TCGA comparison: For comparison of ARID1A mutation frequency, it sounds like this was made to the snDNA-seq results, not the investigators' own WES results. Have the investigators compared frequencies of their snDNA-seq and their own bulk WES?
6. Authors claim no general rules of organ-specific selective pressure: Absence of evidence here may not be evidence of absence of a pattern, since the cohort size is relatively small. You may not have enough data to make this statement confidently.
7. What is the clinical ramifications of this study given that the cohort is clinical?

Other comments
1. VAF is conventionally understood now as "variant allele fraction" rather than "variant allele frequency".
2. I did not see very much on the uniformity of coverage or other QC assessments.
3. Print and symbols in many of the figures are way too small and do not likely satisfy the journal's minimum font size.

Version 1:

Decision Letter:

Our ref: NG-A66776R

11th Sep 2025

Dear Christine,

How are you? I hope you're well.

Thank you for submitting your revised manuscript "Genomic evolution of pancreatic cancer at single-cell resolution" (NG-A66776R). It has now been seen by the original referees and their comments are below. The reviewers find that the paper has improved in revision, and therefore we'll be happy in principle to publish it in Nature Genetics, pending minor revisions to satisfy the referees' final requests and to comply with our editorial and formatting guidelines. In particular, please note the requests for full data availability by Reviewer #1; acceptance will be contingent on this.

If the current version of your manuscript is in a PDF format, please email us a copy of the file in an editable format (Microsoft Word or LaTex)-- we can not proceed with PDFs at this stage.

We are now performing detailed checks on your paper and will send you a checklist detailing our editorial and formatting requirements soon. Please do not upload the final materials and make any revisions until you receive this additional information from us.

Thank you again for your interest in Nature Genetics Please do not hesitate to contact me if you have any questions.

Sincerely,

Safia Danovi, PhD
Senior Editor, Nature Genetics
ORCiD: 0009-0007-7822-5479

Reviewer #1 (Remarks to the Author):

I thank the authors for responding to my questions and for improving the presentation of novel key observations and concepts. I have no more questions with respect the manuscript, but a different type of request. To make the data from this publication widely useful to the community, could I request that the authors please provide all single cell genotypes (along with a basic set of metadata) in an unrestricted manner? I believe there is no reason to make the somatic data access-restricted in EGA, as it would provide no identifiable information whatsoever. Presumably, the full data set could even be

shared in a large supplementary file in the format of cell 1 - SNV1, SNV2, SNV3, Indel2; cell 2 - SNV4, SNV5 etc., along with the patient ID and the identity of the sample. It is possible the data is already available on the authors' Github repository but I could not easily find it there, I could only identify more "polished" versions in which cells had already been assigned to clones which may not be desirable to some. Many publications already do this but not all. I think it is paramount that this kind of simple data sharing become more common and I thank the authors in advance.

Reviewer #2 (Remarks to the Author):

The current m,anuscript provides some new insight in the clonal evolution of pancreatic cancer and constitutes a technical advancement over previous genomics work (much from the same lab) that has established a roadmap of pancreatic cancer evolution. I think the manuscript will be of interest; the limitation in number of samples for some group is inevitable given the difficulty in procuring this kind of samples. I believe the data provided will be of broad interest to the pancreatic cancer community once made public. Of particular interest is that RAS resistance signatures are likely pre-existing any treatment, which will be important for patient stratification in the future.

Reviewer #3 (Remarks to the Author):

The authors attempted to address some technical questions but have not convincingly demonstrated the novelty of this study. Subclonal KRAS mutations in PDAC are well established, and this work does not provide substantial new insights. In addition, BRCA2 bi-allelic loss is expected for a tumor suppressor, and its role in PDAC has been previously reported. Furthermore, I believe the dataset generated from this moderately sized cohort should be more thoroughly integrated with the broader PDAC omics literature.

MANUSCRIPT NG-A66776R

Response to Reviewers' Comments
Haochen Zhang, Christine A. Iacobuzio-Donahue

Reviewer #1 (Remarks to the Author):

In a manuscript entitled "Genomic evolution of pancreatic cancer at single-cell resolution", Zhang et al. describe the analysis of ~140K single nuclei, representing 72 samples from 24 pancreatic cancer patients. The nuclei were analyzed on the commercially available Tapestri platform, using a panel of ~ 600 amplicons targeting common pancreatic cancer driver variants. The authors begin their narrative with a broad description of the genetic characteristics of the cohort, then comment on combinations of mutations found in patients with (presumed?) BRCA2 germline mutation, then explore CNV heterogeneity and report common alterations resulting in TGF-beta inactivation. Finally, a few points are made about evolution during metastasis and the paper closes with a few data points from longitudinally collected samples.

I think the data set underlying this paper is very interesting, and I appreciate the authors' deep expertise in pancreatic cancer evolution and computational biology. I also think that the manuscript is written in a clear and accessible manner and has a logical structure.

Response: We appreciate your acknowledgement of our work.

My main concerns relate to the substance and novelty of the findings, as well as to the narrative style, which largely consists of a succession of patient-level anecdotes but lacks any cohort-wide quantitative analysis that would provide new insights into how pancreatic cancer evolves.

It appears that the authors' main novelty argument is that the single cell sequencing assay applied to this cohort – unsurprisingly – has improved sensitivity for picking up rare subclonal mutations, this is contrasted with general "bulk sequencing", although the authors do not specifically explain which kind of bulk sequencing they mean (e.g. error-corrected ultra-deep bulk sequencing could easily outperform the Tapestri assay for detection of rare mutations). It makes sense that sequencing more and more deeply or on the single cell level will reveal ever rarer mutations, but why is this important? The authors largely fail to make a convincing case in my opinion.

Response: Thank you for raising this valid point. We acknowledge that our argument that the Tapestri single-cell sequencing assay is more sensitive than general "bulk sequencing" is inaccurate and are aware of enhanced bulk sequencing such as Nanoseq [1] which would be more suitable than our method for the purpose of detecting rare single nucleotide mutations. We also recognize that this should not be the main novelty point of our paper as we have already published our method [2]. Therefore, we have now removed any such claims from our paper in the Results and Discussion sections.

We believe the strength of snDNA sequencing is not for its sensitivity of picking up rare short mutations (single nucleotide variants, SNVs) alone, but rather for its ability to discriminate co-occurring versus mutually exclusive SNVs, something that bulk sequencing is unable to do. The advantages of distinguishing such cell populations within a mixture of normal and tumor cell nuclei has allowed us to make novel observations pertaining to *KRAS* wildtype (WT) populations, timing of *BRCA2* biallelic loss, and convergent evolution within the TGF-β pathway (these three points will be discussed in greater detail on the following page).

We also highlight here specific examples within individual PDACs. For example, in PC05 we found a somatic *DNMT3A* p.R882H hot spot mutation assumed to be tumor-specific by bulk whole exome sequencing (WES) but ultimately found to belong to an unrelated lineage within this sample, likely reflecting clonal hematopoiesis of lymphocytes within the sample (see below **Figure R1A**). In patients PC14 and PC22, we found functionally deleterious mutations to *CTNNA2*, *RNF43* and *MTOR* respectively. Both are known low frequency PDAC driver genes [3], yet our data clearly show these mutations are in cells unrelated to the PDAC within each of these samples **(Figure R1B-C)**. We have added these result to **new Supplementary Figure 2C-E** of the manuscript.

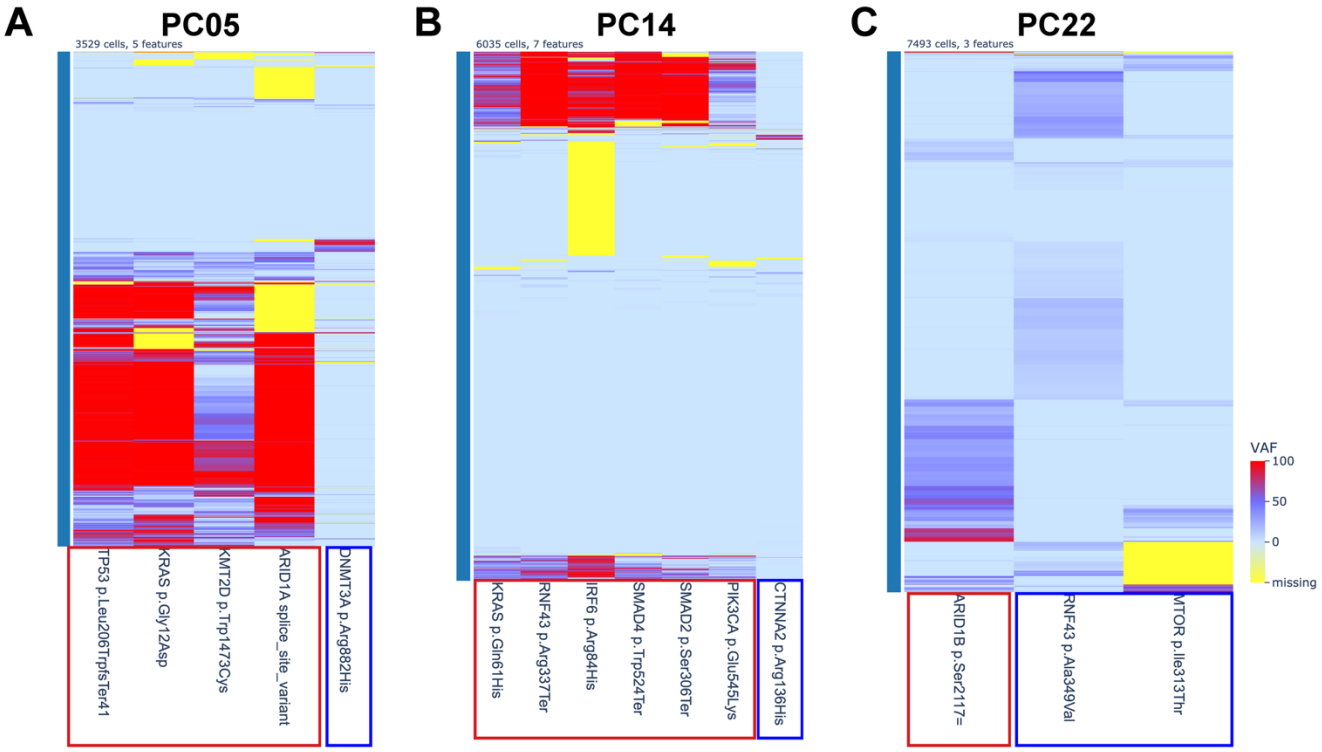


Figure R1: colocalization pattern select SNVs detected by Tapestri in cases PC05, PC14 and PC22
Single-cell (Y-axis) by SNV (X-axis) heatmaps are shown for PC05 (**A**), PC14 **(B)** and PC22 **(C),** where the heat is variant allele frequency (VAF) for each SNV in each single-cell on a scale from 0 to 100 (missing indicates that the locus had zero depth). For each case, SNVs in the red square were identified as belonging to the tumor based on colocalization of multiple PDAC-relevant SNVs and CNVs (not shown). SNVs in the blue square were those that were picked up by Tapestri but did not colocalize with the tumor-SNVs. They are likely carried by cells from the stroma that were sampled along with the tumor tissue.

This technical advantage also has utility for identification of small (200-300 base pairs) or subclonal deletions that are more challenging to detect than SNVs in bulk DNA sequencing. For example, we found subclonal homozygous deletions in *CDKN2A and SMAD4* at higher frequencies than previously described by bulk-sequencing, as we describe in the main text (**page 7, paragraph 2**):

"Subclonal and subgene (affecting only a part of the gene) monoallelic/biallelic losses may be difficult to detect with bulk sequencing methods, particularly for those PDACs where tumor purity may be limited. Our targeted snDNA-seq method evaluated each mutant loci's variant allele fraction (VAF) and single amplicons' (200-300 base pairs) read count while accounting for amplicon dropout rates, to detect such events in small groups of cells"

Finally, this orthogonal view of PDAC mutations has formed the basis of our new single-cell phylogenetic inference algorithm (fast-ConDoR), which leverages not only mutations but also loss of heterozygosity (VAF~0.5 to VAF~0 / 1) events for reconstructing single cell phylogenies.

In sum, we agree that our prior claim was misleading in emphasizing the sensitivity aspect of our method on rare SNVs; we have deleted the relevant claims and shifted our focus on the three novelty points described in page 4 of this document.

They do mention that some rare subclonal variants might lead to therapy failure, but this is not at all demonstrated in the manuscript, and perhaps it would be hard to do so with a sequencing panel that is focused on variants that are essential in PDAC (and thus will largely be truncal and cannot be under selection during therapy).

Response: Thank you very much for raising this point. We would like to point out that our panel was not only designed based on the known genomic features of our patient cohort but was also designed to detect mutations to known PDAC driver genes based on the cBioportal database of >6000 PDACs sequenced at our institution. Drivers affected by both SNV or CNV were considered, as well as those that occur at both high frequency and low frequency based on the collective literature of PDAC. We have included **new Supplementary Tables 4 and 5** with the full list of genes and amplicons on our panel.

In our recently published manuscript in *Cancer Discovery* by Mullen et al [4], we found many functionally deleterious *subclonal* mutations (~28%) occur in PDAC that are potentially under selection in the appropriate treatment context. This is confirmed by our current data using snDNA sequencing. We have identified cases with subclonal *PIK3CA* or *MAP2K4* mutations (for example PC14, PC11 and PC03), *CDKN2A* and *SMAD4* deletions (PC02, PC20), and amplifications of *KRAS* and *MYC* (PC10, PC11). *PIK3CA, MAP3K4, KRAS* amplification and MYC amplification are known resistance mechanisms to *KRAS* targeted therapy [5]. We have added relevant results to the ***Heterogeneity of KRAS dependence Section*** (**page 11, paragraph 1**) to better emphasize such findings.

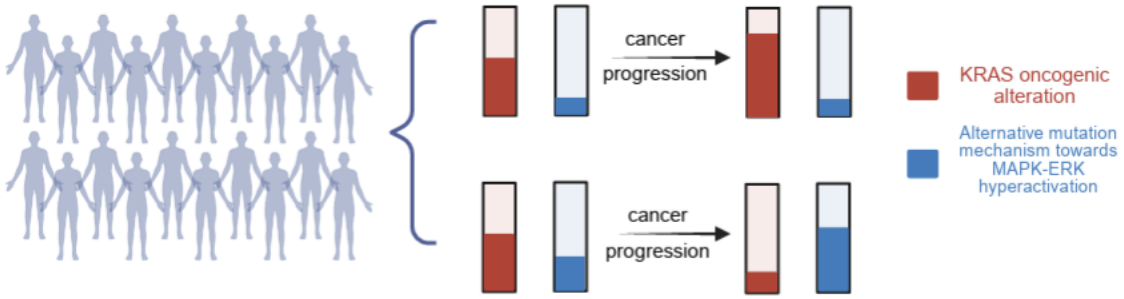
The majority of the paper consists of mini-vignettes that describe the evolution of specific mutation combinations in particular patients (e.g. "PC01's phylogeny showed that the somatic BRCA2 nonsense mutation was truncal, followed by several other LOH events

and CDKN2A homozygous deletion" or "The continuous evolution also manifests as LOH of somatic driver mutation loci in PC10. Two TGFBR1 nonsense mutations were subclonally gained in one of the two main lineages of metastatic clones (Figure 3D)."). I do not doubt that the authors were able to ascertain these facts from a technical perspective, but I am not sure what we can learn from this. Some mutations will be clonal, and others will be selected subclonally, I think the principle is well-established. Particular pathways (namely those included in the sequencing panel based on prior knowledge) appear to be under selection, which is expected. But I cannot discern where the paper accomplishes the crucial task of abstracting the many patient-level findings into any kind of general rule that could be considered novel or memorable.

Response: We thank the reviewer for raising this valid criticism, and for encouraging us to dig deeper into our dataset. This has been a worthy endeavor as we believe we are now able to more clearly describe three key novelties to understanding of the pancreatic cancer genome. These three observations are highlighted in a new Figure 6 that we also include below for reference.

A summary of our novel observations is as follow: First, regarding oncogenic *KRAS*, we have identified subclones with likely **loss** of the mutant *KRAS* allele in 42.8% of PDACs (**new Figure 6A, below**). While these cell populations were not the predominant clone in each PDAC, the fact that they exist raises the possibility that *loss of KRAS mutant alleles should be anticipated as a resistance mechanism to allele specific KRAS inhibitors.* Clinically, we provide an example from our institution of this phenomenon in a patient with a G12C mutant colon cancer treated with sotorasib (manuscript in review): progressive disease was noted to have loss of the mutant allele. We also noted that there is heterogeneity across (and even within) PDACs with respect to dependence on *KRAS* oncogenic signaling. Moreover. as none of the patients in our cohort were treated with KRAS inhibitors, it indicates that many of the mechanisms of resistance to KRAS-targeting therapies that have been described in ongoing trials are pre-existent.

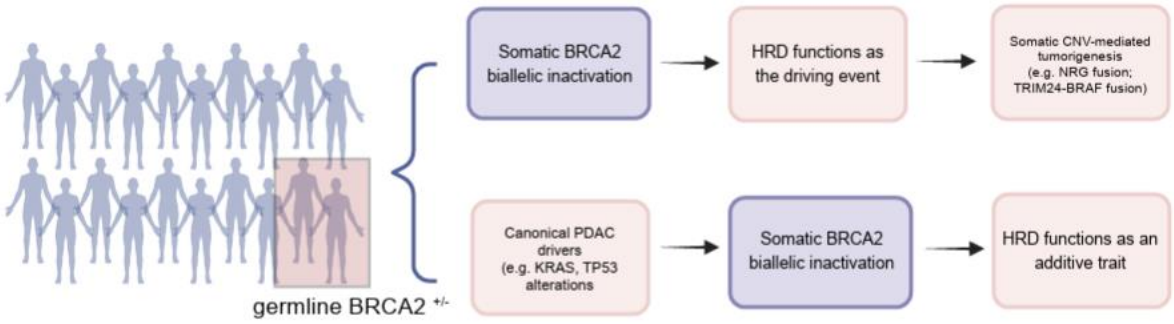
A. PDACs may display heterogeneous dependence on oncogenic KRAS alteration



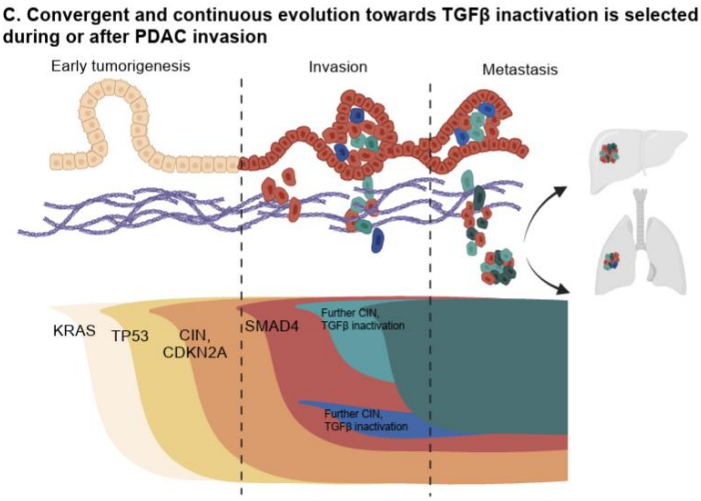
Additionally, contrary to reports in PDAC precursors, we did not observe any secondary *KRAS* mutations in PDAC. There was one secondary *KRAS* in PC11, but with snDNA sequencing we were able to discern that this mutation is in an unrelated lineage within the pancreas (**new Supplementary Figure 5B**). This is clinically important, suggesting that pre-existent clones carrying secondary oncogenic *KRAS* alterations are uncommon, even in *KRAS* WT PDACs.

The second novel finding is that we identified heterogeneity with respect to the timing of *BRCA2* biallelic inactivation in patients with germline heterozygous *BRCA2* (*gBRCA2*) mutations (**new Figure 6B, below**). For example, some patients' loss of the wild type allele was the earliest somatic event and the most recent common ancestor (MRCA) of these PDACs *is defined by biallelic BRCA2 loss and onset of the homologous recombination deficiency (HRD)phenotype*. In these patients' non-canonical forms of ERK activation due to gene rearrangements were identified (*BRAF* fusion, *NRG* fusion) by clinical grade sequencing, providing the basis for why *KRAS* WT genotypes are more common in PDACs that arise in *gBRCA2* carriers. By contrast, in other *gBRCA2* patients' loss of the wild type *BRCA2* allele occurred subsequent to acquisition of *KRAS* and *TP53* mutations. We believe these PDACs represent those with more conventional biology for which HRD may be an additive trait. Ultimately, the broader significance of this finding remains to be determined given the small number of patients analyzed. Nonetheless, while reversion mutations are a known mechanism of resistance and mixed clinical responses, we posit that the timing of biallelic inactivation of *BRCA2* may provide an additional variable to be considered in anticipation of mixed clinical responses among patients with similar *BRCA2* germline variants; this is important when making clinical management decisions regarding use of KRAS inhibitors versus platinum-salts/PARPi.

B. Heterogeneous timing of BRCA2 biallelic inactivation in PDAC patients with germline BRCA2 alteration



The third novel finding is that unlike common driver genes that are selected for early (truncal/clonal) in the evolutionary life history of PDAC, inactivating mutations of the TGF-β pathway occur in parallel lineages over time and space (**new Figure 6C, following page**). In some PDACs inactivation of different members of the pathway are inactivated in the same lineage. Both patterns suggesting ongoing selection for phenotypes with greater extents of non-responsiveness to TGF-β signals from the microenvironment. Consistent with this interpretation we find these events to take place late in PDAC progression, often coinciding with invasion and metastasis. This pattern is contrary to the genetic alterations in other drivers of PDAC such as *KRAS* and *TP53* that are selected for during tumorigenesis.



This weakness stands out starkly throughout the paper, but particularly in the metastasis evolution and longitudinal sampling sections. The main summary figure of the metastasis evolution section (Fig. 5B) tells us that the authors did not observe any quantifiable differences in diversity among metastatic sites, or in divergence from the primary, and could find no correlation between metastasis genetics and spatial localization. What do we learn here? That PDAC metastasis does not follow any discernible rules or that the resolution of the sequencing assay is too limited to relevant effects? It is unclear to me as a reader.

Response: Thank you for raising this concern. We have removed and/or significantly edited down our findings pertaining to metastasis given they are negative data.

We agree that the previous Figure 5B can be misleading by concluding that PDAC metastasis does not follow any discernible rules; it should be corrected to specify that coverage of the sequencing assay (DNA-level, targeted on 280 genes that are enriched for mutation in PDAC) failed to identify any discernible rules. We have added such a sentence in the new section ***Major patterns of evolutionary trajectory:***

(Page 9, paragraph 2*): As our panel targets driver mutations which are previously known and hereby confirmed to exhibit homogeneity across metastatic sites (two example cases shown in* ***Supplementary Figure 3C-D)****, our capacity of assessing genetic heterogeneity and in turn, metastatic seeding patterns was limited in most metastatic cases we sampled.*

We still favor that the pattern observed in PC10 is worth mentioning, as it provides direct evidence to the single-stage evolution model metastasis (founder cells of the metastatic lesions were apparently randomly sampled from the available primary tumor lineages), which was indirectly inferred by bulk sequencing [6], [7], [8].

We also agree with the abundance of literature claiming that the evolution of PDAC during progression to secondary sites largely manifests at the transcriptomic level. This may be due to epigenetic changes (not studied in this manuscript) or by copy number variation across the genome. Our data illustrating this latter point using the longitudinal sample PC20 is now moved to section ***Major patterns of evolutionary trajectory*** (page 9, paragraph 1) to emphasize this

point. Notably, we have orthogonally conducted single-nucleus RNA sequencing thoroughly across multiregional samples of one of the metastatic cases in this cohort (PC13), and have written a separate manuscript that robustly underscores this point (Jimenez-Sanchez et al. Cancer Research 2025 *in press* and bioRxiv [9]).

In summary, I very much like the underlying data set, but I struggle to identify the biological insights that would make the manuscript of interest to a broad audience. I would encourage the authors to transcend the rather anecdotal patient-level narrative and aim to expose broad patterns in the data which could then hopefully be reported as genuinely novel insights into PDAC evolution.

Response: Thank you for your critique which prompted us to restructure our paper. It has greatly improved as a result. We hope you agree.

Reviewer #2 (Remarks to the Author):

This manuscript by Dr Iacobuzio-Donahue and colleagues provides an in-depth analysis of clonal change in pancreatic cancer, by utilizing single cell resolution DAN sequencing. The manuscript builds on a large tissue database and unique samples available to the investigator, such as rapid autopsy samples and longitudinal samples from the same patients pre and post-surgery. Using a custom panel of 596 genes, the authors evaluate mutations, copy number alterations across tumors, including analysis across space (primary vs metastases, and different metastatic loci) and time (over the course of treatment). The findings offer an in-depth view in the genomic landscape. Of pancreatic cancer, building on this group's past track record. Of additional interest, the authors include samples from BRCA mutant pancreatic cancer, providing new information on the molecular evolution of disease in the context of BRCA loss. Overall, this manuscript provides a greatly needed resource and will be a go-to reference for future work.

Response: Thank you for your recognition of our work.

I only have some minor comments to note:

1) Presumably non tumor cells (components of the stroma) were captured in the sequencing samples. If so, it would be interesting to analyze whether they have any mutation in pancreatic cancer drivers, as that is a still debated concept in the field.

Response: Thank you very much for bringing up this intriguing question. Our samples indeed included nuclei from non-neoplastic stromal cells, although we are unable to discern if these nuclei were fibroblastic, neural, vascular or immune in origin. In PC05 a *DNMT3A* p.R882H mutation was found that was mutually exclusive with the neoplastic cells (response to reviewer 1 section **Figure R1A**). This mutation is a common hotspot in clonal hematopoiesis. PC14 had a *CTNNA2* p.R136H mutation that was also mutually exclusive with the neoplastic cells (**Figure R1B**). PC22 had a *RNF43* p.A349V mutation and a MTOR p.I313T mutation, both mutually exclusive with the neoplastic cells and each other (**Figure R1C**). We found only one case (PC11) with a secondary oncogenic *KRAS* mutation that did not belong to the tumor lineage (further

elaborated below). All these results are added to the new manuscript (**new Supplementary Figure 2C-E**; **Supplementary 5A-B**).

2) Recent work on PanIN, pancreatic cancer precursor lesions, has reported the presence of multiple different mutations in the *KRAS* gene within the same lesion. It would be interesting to know whether multiple mutations in *KRAS* were detected in any of the tumors, even if only in a small subset of clones.

Response: Thank you for raising this interesting point. We did observe a second *KRAS* mutation in a single patient (PC11) that was mutually exclusive to the main tumor clone carrying a *KRAS* G12V mutation and *SMAD4* mutations (**new Supplementary Figure 5A-B)**. This minor clone carried a *KRAS* G12D mutation without any other detectable somatic driver mutation. Interestingly the minor clone was detected in all three samples of the primary tumor taken from different spatial location, corresponding to either one clone colonizing the pancreatic ductal system [10] or different clones with the same *KRAS* hotspot mutation.

We have added relevant results to **Page 10 Paragraph 2** in the revised manuscript.

3) The cartoons provided in Figures 1 and 5 are informative, but an additional summary cartoon that shows the key findings of the study to a broader readership would further enhance the manuscript's impact.

Response: Thank you very much for your suggestion to our paper. We agree that cartoons demonstrating the novel patterns we found in PDAC evolution would enhance the manuscript's impact to a broader readership. We have now added **new Figure 6** for this purpose.

Remarks on code availability: Reviewing the code goes beyond my area of expertise

Response: All code has been made available on Github, as described in our manuscript:

The custom computational analysis pipeline consists of three modules:
(i) A bioinformatics pipeline for single-cell variant calling, available at https://github.com/haochenz96/TapVarCallSmk.
(ii) Copy number clone calling pipeline, Tapestri-CN, available at https://github.com/haochenz96/tap_cn_calling.
(iii) The fast-ConDoR pipeline for phylogenetic analysis, available at https://github.com/Bolladeen/full-ConDoR.

Reviewer #3 (Remarks to the Author):

The manuscript describes application of snDNA-seq of a moderately-sized PDAC cohort to investigate its evolutionary genomics. The work seems like a logical follow-up to the authors' earlier paper by Zhang et al. (2023). Some of the observations recapitulate known findings, but the focus on evolution and lineage tracking reveals some interesting connections among cell groups.

Response: Thank you very much for your recognition of our work.

One concern is that the confidence of these copy number calls the authors relied on heavily for making clonal assignments and other further conclusions. Also, there are a number of comments related to methods, statements, and presentation.

1. The cancer cell fraction (CCF) is a useful metric, but I can think of several issues.
A. The denominator is set by the number of cells having KRAS mutations, which is presumed to be a founder and thus in all tumor cells in a sample.
B. Are there true tumor cells lacking KRAS mutation, and could these even possibly have the clonal mutation in question? This would also affect CCF.
Since CCF is based on a sample (in the statistical sense) of all the cells in a tumor mass, one could even think about CCF error bars, though I think that is beyond the scope of this work. Some comments on this overall issue in the manuscript would be helpful.

Response: Thank you very much for raising this important point. When writing the original manuscript, we considered three ways of defining the denominator for CCF calculation:

1. The number of cells carrying *KRAS* mutations, except for those that do not have a KRAS hotspot mutation (PC01, PC22 and PC23).
2. The number of cells carrying the most prevalent somatic driver mutation in each tumor.
3. The number of cells inferred by the copy number (CN) clone clustering method to have any non-diploid state across the targeted genomic region.

We chose Method 1 in our original subission as sufficient evidence from prior literature have showed that KRAS hotspot mutations are the earliest oncogenic driver in most PDACs [11], [12]. However, we acknowledge its limitation in that it relies on *a priori* knowledge and some tumor cells may artificially (due to technical allelic dropout, ADO) or factually not carry *KRAS* mutations.

Method 2 might reduce the bias of relying on a priori knowledge compared to Method 1, but would be subject to the same technical noise. As a hypothetical example, suppose a sample has a heterozygous *KRAS* mutation present in 100% of the tumor cells and a *TP53* mutation in 95% of those cells, but all TP53 mutant cells lost their wildtype allele and gained one extra copy of the mutant TP53 allele (copy neutral LOH, CNLOH). Because we have a ~20% ADO rate as estimated by our prior publication [2], in a diploid cell there will be a ~10% chance of not measuring either one of the two alleles of any gene- If p is the probability of an allele dropping out in a cell, then (1-p)^2 is the probability of no dropout, which would be 1-ADO. Thus, given ADO=0.2 we can calculate p = 1-sqrt(1-ADO) = 1 - sqrt(0.8) = 0.1056. Therefore for the example case, we should observe 10% tumor cells not carrying KRAS and 10% cells carrying homozygous KRAS mutation. But for TP53 because of the CNLOH, the chance of not observing a TP53 mutation at all would be 10%*10% = 1%, so we should only see 95% * 1% = 0.95% tumor cells without that mutation and 94.05% tumor cells with that mutation. The result would be that TP53 would be more prevalent than KRAS and be erroneously selected as the clonal driver.

Method 3 might be subject to technical noise with our CN clustering method, but we have decided to switch to this approach in this updated manuscript because:
(a) it does not make any assumption based on *a priori* knowledge.

(b) it relies on CNV signals to define tumor clones similar to prevalent methods in single-cell RNA sequencing analysis of PDAC [13].
(c) The primary goal for CCF calculation was to quantify the relative clonality among SNVs at the population-level. Using CNV as an index avoids using any particular SNV as reference, thereby providing an objective view.

We have edited the results, in particular the oncoprint in **new Figure 1B** and relevant descriptive text in both the main text and methods to reflect this change. Importantly, the conclusions stay the same, with *TP53, BRCA2, KRAS*, CDKN2A remaining as clonal drivers while SMAD4, ARID1A as subclonal drivers across the cohort.

Nonetheless, we caution that a population-mean CCF can be misleading. For instance, there are many instances where *CDKN2A* was inactivated subclonally, exemplified by PC21 (**new Figure 3F)**; or *SMAD4* was inactivated clonally, such as in PC11 (**new Figure 5A**). Therefore, we added the sentence:

Page 7, paragraph 1: *"...yet heterogeneity of clonality existed for almost every gene"*

We admit that sampling error that the reviewer mentioned, along with the technical noise mentioned above, would affect CCF calculation. Therefore we added the following sentence to remind readers to focus on the relative values of CCF among SNVs at cohort-level:

Page 7, paragraph 1*: "... It should be cautioned that this scDNA-seq CCF calculation was subject to noise from technical dropout and sampling errors, and the values should only be used for quantifying relative clonality among SNVs".*

2. How do you choose/justify a CCF cutoff to distinguish clonal drivers (CCF ~0.9) versus sub-clonal drivers (CCF ~0.6)? If CCF is not totally immune to the effects just mentioned, could this be a more subtle process of distinguishing than you are accounting for? Does the choice of cutoff affect your results?

Response: The choice of a CCF cutoff to distinguish clonal versus subclonal drivers was based on both prior literature and empirical observations on technical noises. Prior literature from Landau et al. 2013 and Dentro et al. 2021 [14], [15] chose cutoffs for clonal vs subclonal drivers of CCF=0.95 and 0.90, respectively. In our scDNA-seq dataset, to more leniently account for factors such as copy number variation (CNV), and technical errors (such as ADO mentioned above) that may introduce variability to our CCF estimates, we lowered the cutoff to 0.75.

As mentioned above, the calculation of CCF was not immune to technical noise and sampling error, and the values were only used for quantifying *relative* clonality among SNVs at cohort-level. The choice of cutoff certainly would affect our conclusion - for instance, if we chose a cutoff of 0.8, KRAS with a mean CCF of 0.76 would be considered subclonal, which would be against much of the prior literature on PDAC genetics. We have added the following sentence to **Page 24, paragraph 3**:

The CCF value of 0.75 to distinguish between clonal and subclonal drivers was chosen empirically (similar to prior literatures using bulk sequencing [Dentro et al. 2021]) as we observed a clear gap dividing two clusters (cluster1: TP53, CDKN2A, KRAS; cluster2: SMAD4, ARID1A etc.) at that value.

To assess relative clonality of genetic events in each individual tumors, we did not rely solely on CCF but ran several other analyses such as fast-ConDoR phylogeny reconstruction, statistical testing of multiple events (for instance, in **page 12, paragraph 3**, to validate the *BRCA2* homdel clone in PC21, we looked for signals of *CDKN2A* homdel as well as *TP53* heterozygosity), inspection of raw signals of variant allele frequency etc. For these analyses, the choice of CCF cutoff did not affect the conclusions on clonality at all.

3. There are other hard-coded cut-offs baked into the analysis, for example the positive mutation numbers quoted in the paper. Without having more explanation available as to how these are chosen, it leads to the question of whether any changes in these values would affect the scientific results. It would add confidence if we were to know why these are the best values, or if some of the analyses could be repeated with other values, showing that results are not overly sensitive to their choices.

Response: We would assume that the reviewer was referring to the below sentence in our Methods description:

Page 21, subsection *"**Mutation calling**", paragraph 1: "We concatenated mutations called in each single cell, and with the belief that the false positive rate of a mutation decreases exponentially with the number of barcodes it is detected in, we only considered mutations present in >= 3 single cells."*

The choice of 3 single cells was arbitrary and came from empirical testing - where 3 single cells is a lenient cutoff to filter out a good amount of false positive mutations (because the false positive rate of a mutation decreases exponentially with the number of barcodes it is detected in) while leaving enough mutations for further filtering, described in the following sections in Methods:

Page 22, subsection ***"Filtering by relative prevalence and mutational signature" -*** this is a much stronger filter than the hard cutoff of 3 cells (because most of our samples are >1000 cells so the 0.5% prevalence filter would filter out any mutations present in less than 5 cells) and the rationale has been explained in **new Supplementary Figure 1A & B** and relevant texts in Methods:

*"In both neoplastic and normal samples taken from different organ sites, we noticed a significantly similar mutational signature, signified by a large proportion of T>C variants, enriched in mutations <0.5% single-cell prevalence detected by this snDNA-seq technology (**Supplementary Figure 1B**). As different types of human tissues have been shown to display distinct mutational signatures and the signature did not match any of them [16], we deemed it likely artifactual and decided to filter out all mutations below 0.5% single-cell prevalence. T>C variants in the 0.5% to 1% were also filtered."*

"Assembly of a panel of normal" - this type of filtering is common in bulk sequencing [17] and is also a much stronger filter than the hard cutoff of 3 cells.

Given the many more stringent filtering downstream of the hard cutoff of 3-cell prevalence filter, small shifts to changes to this filter did not affect scientific results. We added the following sentence in **Page 21**, subsection *"**Mutation calling**", paragraph 1* to clarify this point.

"It should be noted that the choice of 3 single cells was arbitrary and came from empirical testing, where a prevalence filter of 3 single cells filtered out good amounts of false positive mutations while leaving enough mutations for downstream, more stringent filters to be described below."

4. (lines 124-126) and (lines 137-139, 148-150) higher alteration of CDKN2A and SMAD4 and attribution to higher sensitivity than TCGA bulk: You place the sensitivity of bulk at 2%, but I don't think bulk is traditionally thought of as being reliably this sensitive. Does this estimation consider issues of depth, purity, etc? The required minimum depth 4 reads for detection that you quote would suggest that sensitivity is a function of depth.

Response: Thank you for pinpointing the flaw in our claim. We want to clarify to the reviewer that "minimum depth of 4 reads for detection" (mentioned in **Methods** subsection "***Mutation calling"*** paragraph 1*)* was not what we quoted for estimation of *bulk sequencing*'s sensitivity; it was a parameter we used for filtering mutation calls in single-cells (Tapestri), which, as mentioned in the response above, was one of the many steps of filtering that we used in Tapestri's bioinformatics and often did not directly affect the final filtering due to the more stringent filtering steps downstream.

The 2% sensitivity of bulk was really a naive estimation that we made assuming 100% tumor purity, sufficient depth (must be greater than threshold on depth; must also satisfy variant allele count = 2% * depth > threshold on variant allele count). As stated in the previous **Methods** subsection **"*Estimation of the highest resolution of bulk sequencing and Tapestri in calling somatic mutations*"**, we quoted Frankell et al [18] "for a state-of-the-art bulk sequencing bioinformatics pipeline.

> *'An SNV was considered a true positive if the variant allele frequency (VAF) was greater than 2% and the mutation was called by both VarScan2, with a somatic P ≤ 0.01, and MuTect. Alternatively, a frequency of 5% was required if only called in VarScan2, again with a somatic P ≤ 0.01. Additionally, the sequencing depth in each region was required to be ≥30, and ≥10 sequence reads had to support the variant call.'*

"From here, we took the most relaxed threshold 2% VAF, and assumed that the driver mutations of interest have loss of heterozygosity in cancer cells. We arrived at 2% single-cell prevalence as bulk sequencing's highest resolution of somatic mutation calling."

We admit the resulting 2% is an overestimation, and we used it in the previous manuscript mainly for contrasting against Tapestri's higher sensitivity 0.5%. Realizing that this might raise

confusion and detecting rare mutations is not the most significant capability of single-cell DNA sequencing, we have removed relevant sentences from our revised manuscript.

5. TCGA comparison: For comparison of ARID1A mutation frequency, it sounds like this was made to the snDNA-seq results, not the investigators' own WES results. Have the investigators compared frequencies of their snDNA-seq and their own bulk WES?

Response: Thank you for this suggestion. We did compared the snDNA-seq results with paired bulk sequencing results on our N=24 cohort and *ARID1A* was detected in all but one case (PC17) which presented a very rare subclone carrying the mutation. However this was not a fair comparison as our Tapestri panel was designed to cover PDAC-relevant mutations detected by multiregional bulk sequencing on the same samples, so the high congruence was expected. Regardless, we recognize that the higher mutation frequency of ARID1A observed here was not solely due to increased sensitivity of Tapestri but likely also (1) increase sensitivity of multiregional sampling (2) random enrichment of ARID1A mutations in this cohort. The latter may have contributed more as our multiregional bulk sequencing data on a larger cohort (N=91) suggested only 5% prevalence [4].

As for the previous comment, we have removed relevant sentences on comparing scDNA-seq's sensitivity with bulk sequencing.

6. Authors claim no general rules of organ-specific selective pressure: Absence of evidence here may not be evidence of absence of a pattern, since the cohort size is relatively small. You may not have enough data to make this statement confidently.

Response: Thank you for pointing out this issue that was also raised by Reviewer #1. We have rewritten the previous results section on metastasis to remove any such claims and added such a sentence in the new Section **"Major patterns of evolutionary trajectory"** to acknowledge that with our data we are unable to make claims regarding organ-specific selective pressure**.**

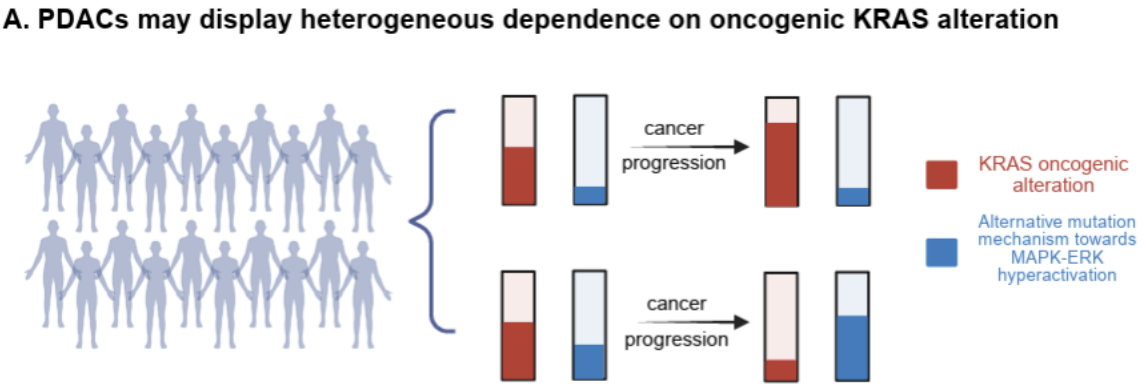
*Page 8, paragraph 3: "As our panel targets driver mutations which are previously known and hereby confirmed to exhibit homogeneity across metastatic sites (two example cases shown in **Supplementary Figure 3C-D**), our capacity of assessing genetic heterogeneity and in turn, metastatic seeding patterns was limited in most metastatic cases we sampled."*

7. What are the clinical ramifications of this study given that the cohort is clinical?

Response: Thank you for the insightful question. Reviewer #1 raised a similar question, which prompted us to rewrite our results to better emphasize the following three main points, which we summarize in a new Figure 6. We provide below the same response below as for Reviewer #1's question.

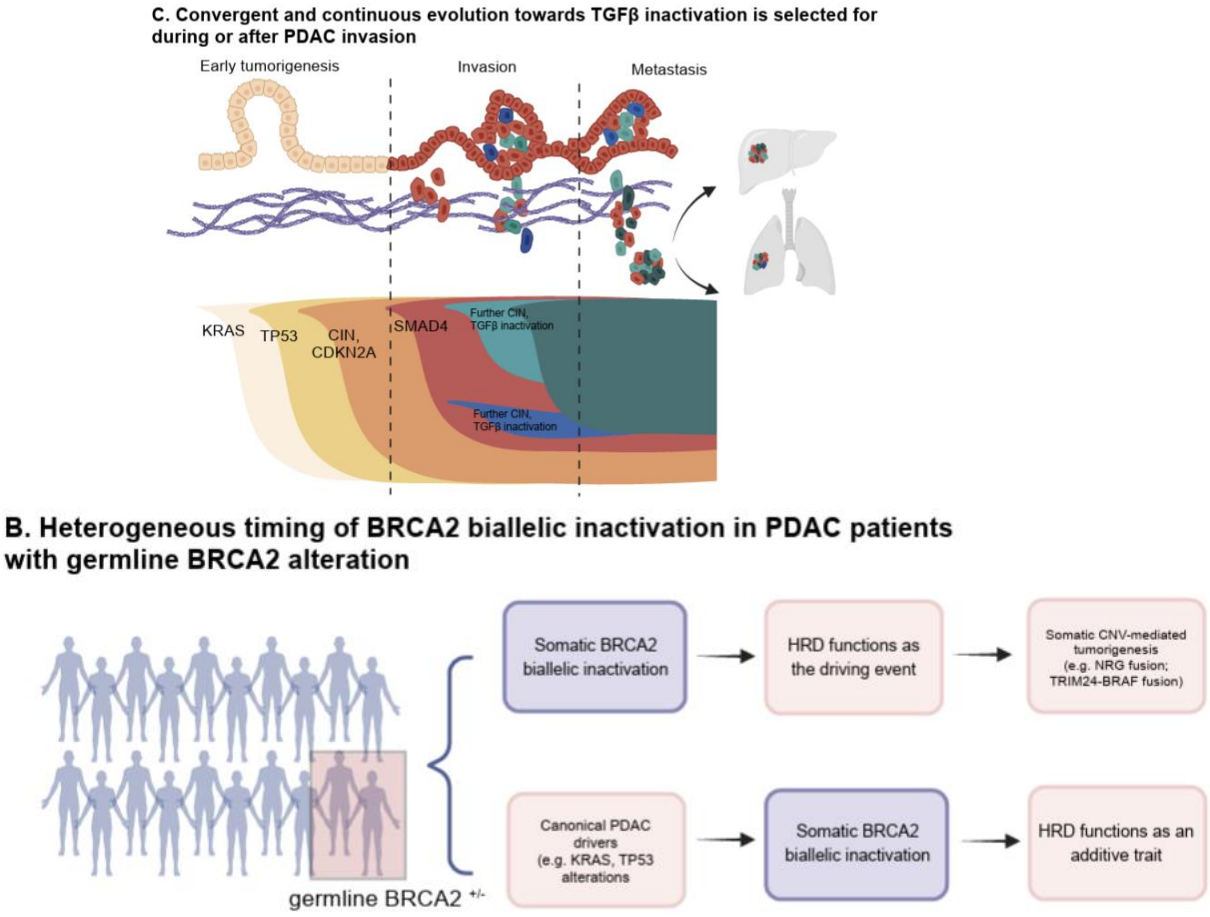
A summary of our novel observations is as follow: First, regarding oncogenic *KRAS*, we have identified subclones with likely **loss** of the mutant *KRAS* allele in 42.8% of PDACs (**new Figure 6A, below**). While these cell populations were not the predominant clone in each PDAC, the fact that they exist raises the possibility that *loss of KRAS mutant alleles should be anticipated as a resistance mechanism to allele specific KRAS inhibitors.* Clinically, we provide an example

from our institution of this phenomenon in a patient with a G12C mutant colon cancer treated with sotorasib (manuscript in review): progressive disease was noted to have loss of the mutant allele. We also noted that there is heterogeneity across (and even within) PDACs with respect to dependence on *KRAS* oncogenic signaling. This heterogeneity is most pronounced in later phases of the evolutionary life history of PDACs. Moreover. as none of the patients in our cohort were treated with KRAS inhibitors, it indicates that many of the mechanisms of resistance to KRAS-targeting therapies that have been described in ongoing trials are pre-existent.



Additionally, contrary to reports in PDAC precursors, we did not observe any secondary *KRAS* mutations in PDAC. There was one secondary *KRAS* in PC11, but with snDNA sequencing we were able to discern that this mutation is in an unrelated lineage within the pancreas (**new Supplementary Figure 5B)**. This is clinically important, suggesting that pre-existent clones carrying secondary oncogenic *KRAS* alterations are uncommon, even in *KRAS* WT PDACs.

The second novel finding is that we identified heterogeneity with respect to the timing of *BRCA2* biallelic inactivation in patients with germline heterozygous *BRCA2* (*gBRCA2*) mutations (**new Figure 6B, below**). For example, some patients' loss of the wild type allele was the earliest somatic event and the MRCA of these PDACs *is defined by biallelic BRCA2 loss and onset of the HRD phenotype*. In these patients' non-canonical forms of ERK activation due to gene rearrangements were identified (*BRAF* fusion, *NRG* fusion) by clinical grade sequencing, providing the basis for why *KRAS* WT genotypes are more common in PDACs that arise in *gBRCA2* carriers. By contrast, in other *gBRCA2* patients' loss of the wild type *BRCA2* allele occurred subsequent to acquisition of *KRAS* and *TP53* mutations. We believe these PDACs represent those with more conventional biology for which HRD may be an additive trait. Ultimately, the broader significance of this finding remains to be determined given the small number of patients analyzed. Nonetheless, while reversion mutations are a known mechanism of resistance and mixed clinical responses, we posit that the timing of biallelic inactivation of *BRCA2* may provide an additional variable to be considered in anticipation of mixed clinical responses among patients with similar *BRCA2* germline variants; this is important when making clinical management decisions regarding use of KRAS inhibitors versus platinum-salts/PARPi.



The third novel finding is that unlike common driver genes that are selected for early (truncal/clonal) in the evolutionary life history of PDAC, inactivating mutations of the TGF-β pathway occur in parallel lineages over time and space (**new Figure 6C, above**). In some PDACs inactivation of different members of the pathway are inactivated in the same lineage. Both patterns suggesting ongoing selection for phenotypes with greater extents of non-responsiveness to TGF-β signals from the microenvironment. Consistent with this interpretation we find these events to take place late in PDAC progression, often coinciding with invasion and metastasis. This pattern is contrary to the genetic alterations in other drivers of PDAC such as *KRAS* and *TP53* that are selected for during tumorigenesis.

Other comments

1. VAF is conventionally understood now as "variant allele fraction" rather than "variant allele frequency".

Response: Thank you for pointing out this error. We have updated the manuscript accordingly to correct "variant allele frequency" to "variant allele fraction".

2. I did not see very much on the uniformity of coverage or other QC assessments.

Response: Thank you for raising this concern. In our prior work, we did a series of analysis on how sequencing depth and resulting library nuclei yield affected the insights we drew from Tapestri scDNA-seq, and as we mentioned in our previous paper [2]:

"…relative proportions of major clones and associated biological insights could be robust to varying technical parameters such as read depth and cell throughput"

In that study, we also found that the discovery power of Tapestri plateaued as the depth (for scDNA-seq, that is *mean reads per cell per amplicon*) neared or exceeded 100. Therefore, the major quality control (QC) assessment we made for each sample in this study was an optimal sequencing depth of 100, and less importantly, an optimal number of nuclei of above 2000. However, because most patients in our cohort had multiple samples studied, even if one sample's depth fell below 100, we could still leverage insights from other samples of the same case. The cellularity of some of the pancreas biopsies were so low that we had to exhaust the whole piece of archived tissue to get <100 cells (e.g. PC19), but we showed that we were still able to locate the tumor-specific SNVs and observe their colocalization pattern in such a sample (**new Supplementary Figure 2A**).

In sum, we did not use our QC assessment on depth and throughput to filter out any sample; we believe our method was robust to the varying depth and cell throughput in this study. We have included the depth for each sample along with the library single nuclei yield in **new Supplementary Table 3** for reference.

3. Print and symbols in many of the figures are way too small and do not likely satisfy the journal's minimum font size.

Response: Thank you for raising this concern. We have now revised many of our figures and enlarged font sizes. In the event of acceptance by the journal, we will do one more round of aesthetic review with the editing team to ensure we satisfy the journal's guidelines.

References (response letter):

[1] F. Abascal *et al.*, "Somatic mutation landscapes at single-molecule resolution | Nature," *Nature*, vol. 593, no. 7859, pp. 405–410, May 2021, doi: 10.1038/s41586-021-03477-4.

[2] H. Zhang *et al.*, "Application of high-throughput single-nucleus DNA sequencing in pancreatic cancer," *Nat. Commun.*, vol. 14, no. 1, p. 749, Feb. 2023, doi: 10.1038/s41467-023-36344-z.

[3] A. Hayashi, J. Hong, and C. A. Iacobuzio-Donahue, "The pancreatic cancer genome revisited," *Nat Rev Gastroenterol Hepatol*, vol. 18, 2021, doi: 10.1038/s41575-021-00463-z.

[4] K. M. Mullen *et al.*, "The Evolutionary Forest of Pancreatic Cancer," *Cancer Discov.*, vol. 15, no. 2, pp. 329–345, Feb. 2025, doi: 10.1158/2159-8290.CD-23-1541.

[5] J. Dilly *et al.*, "Mechanisms of Resistance to Oncogenic KRAS Inhibition in Pancreatic Cancer," *Cancer Discov.*, vol. 14, no. 11, pp. 2135–2161, Nov. 2024, doi: 10.1158/2159-8290.CD-24-0177.

[6] J. G. Reiter, "Minimal functional driver gene heterogeneity among untreated metastases," *Science*, vol. 361, 2018.

[7] A. P. Makohon-Moore *et al.*, "Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer," *Nat. Genet.*, vol. 49, no. 3, pp. 358–366, Mar. 2017, doi: 10.1038/ng.3764.

[8] K. Naxerova, "Evolutionary paths towards metastasis," *Nat. Rev. Cancer*, vol. 25, no. 7, pp. 545–560, Jul. 2025, doi: 10.1038/s41568-025-00814-x.

[9] A. Jiménez-Sánchez *et al.*, "Transcriptomic plasticity is a hallmark of metastatic pancreatic cancer," Mar. 06, 2025, *bioRxiv*. doi: 10.1101/2025.02.28.640922.

[10] A. P. Makohon-Moore *et al.*, "Precancerous neoplastic cells can move through the pancreatic ductal system," *Nature*, vol. 561, no. 7722, pp. 201–205, Sep. 2018, doi: 10.1038/s41586-018-0481-8.

[11] A. M. Braxton *et al.*, "3D genomic mapping reveals multifocality of human pancreatic precancers," *Nature*, vol. 629, no. 8012, pp. 679–687, May 2024, doi: 10.1038/s41586-024-07359-3.

[12] S. R. Hingorani *et al.*, "Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice," *Cancer Cell*, vol. 7, no. 5, pp. 469–483, May 2005, doi: 10.1016/j.ccr.2005.04.023.

[13] W. L. Hwang *et al.*, "Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment," *Nat. Genet.*, vol. 54, no. 8, pp. 1178–1191, Aug. 2022, doi: 10.1038/s41588-022-01134-8.

[14] D. A. Landau *et al.*, "Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia," *Cell*, vol. 152, no. 4, pp. 714–726, Feb. 2013, doi: 10.1016/j.cell.2013.01.019.

[15] S. C. Dentro *et al.*, "Characterizing genetic intra-tumor heterogeneity across 2,658 human cancer genomes," *Cell*, vol. 184, no. 8, pp. 2239-2254.e39, Apr. 2021, doi: 10.1016/j.cell.2021.03.009.

[16] L. B. Alexandrov *et al.*, "The repertoire of mutational signatures in human cancer," *Nature*, vol. 578, no. 7793, pp. 94–101, Feb. 2020, doi: 10.1038/s41586-020-1943-3.

[17] D. Benjamin, T. Sato, K. Cibulskis, G. Getz, C. Stewart, and L. Lichtenstein, "Calling Somatic SNVs and Indels with Mutect2," Dec. 02, 2019, *bioRxiv*. doi: 10.1101/861054.

[18] A. M. Frankell *et al.*, "The evolution of lung cancer and impact of subclonal selection in TRACERx," *Nature*, vol. 616, no. 7957, pp. 525–533, Apr. 2023, doi: 10.1038/s41586-023-05783-5.

MANUSCRIPT NG-A66776R

Response to Reviewers' Comments

Haochen Zhang, Christine A. Iacobuzio-Donahue

Reviewer #1 (Remarks to the Author):

I thank the authors for responding to my questions and for improving the presentation of novel key observations and concepts. I have no more questions with respect the manuscript, but a different type of request. To make the data from this publication widely useful to the community, could I request that the authors please provide all single cell genotypes (along with a basic set of metadata) in an unrestricted manner? I believe there is no reason to make the somatic data access-restricted in EGA, as it would provide no identifiable information whatsoever. Presumably, the full data set could even be shared in a large supplementary file in the format of cell 1 - SNV1, SNV2, SNV3, Indel2; cell 2 - SNV4, SNV5 etc., along with the patient ID and the identity of the sample. It is possible the data is already available on the authors' Github repository but I could not easily find it there, I could only identify more "polished" versions in which cells had already been assigned to clones which may not be desirable to some. Many publications already do this but not all. I think it is paramount that this kind of simple data sharing become more common and I thank the authors in advance.

Response: We appreciate your approval of our revision.

We appreciate your request of unrestricted publication of our dataset, and recognize its importance for the research community.

Upon internal discussion we have decided to disclose the Tapestri H5 files through **Zenodo (https://zenodo.org/records/17155593)** which are the processed data used by downstream analysis scripts (published on the Github repository: https://github.com/haochenz96/Tapestri_main_manuscript_analysis) directly reproducing our figures.

The H5 files include multilayered CSVs that record:
For mutation analysis, for each cell, each locus:

- Reference allele read counts
- Alternate allele read counts
- Total read counts

For copy number variation (CNV) analysis, for each cell, each amplicon:

- total number of reads mapped

Notably, the files included germline mutations since loss of heterozygosity (LOH) events on them are as essential as somatic mutations for defining clonal relationships. This assumed that the targeted sequencing window of Tapestri (accounting for <1e-4 of human genome) produced such limited numbers of germline mutations that they would not enable patient identification.

We have uploaded the raw BAMs to European Genome-phenome Archive (study ID: EGAS50000001351) for interested researchers that prefer starting their analyses from more upstream.

Reviewer #2 (Remarks to the Author):

The current manuscript provides some new insight in the clonal evolution of pancreatic cancer and constitutes a technical advancement over previous genomics work (much from the same lab) that has established a roadmap of pancreatic cancer evolution. I think the manuscript will be of interest; the limitation in number of samples for some group is inevitable given the difficulty in procuring this kind of samples. I believe the data provided will be of broad interest to the pancreatic cancer community once made public. Of particular interest is that RAS resistance signatures are likely pre-existing any treatment, which will be important for patient stratification in the future.

Response: Thank you for your acknowledgement of our work. We hope the work can be valuable to the pancreatic cancer research community.

Reviewer #3 (Remarks to the Author):

The authors attempted to address some technical questions but have not convincingly demonstrated the novelty of this study.

Thank you for bringing up these concerns regarding the novelties of the study.

Subclonal KRAS mutations in PDAC are well established, and this work does not provide substantial new insights.

We acknowledge that although rare, it has been well reported that oncogenic *KRAS* mutations may appear subclonally [1]. The study reported 4 cases with multiple known oncogenic *KRAS* hotspot mutations, 3 of which had one clonal *KRAS* mutation concurrent with a subclonal *KRAS* mutation at a much lower CCF, and the tumors were *inferred* to have acquired *KRAS* sequentially resulting in a small subclonal harboring two *KRAS* mutations on each of the two alleles.

But in our cohort N=24 sequenced with snDNA-seq, we only observed one patient (PC11) where aside from a clonal *KRAS* mutation, a minor *KRAS*-mutated subclone was identified across three spatially distinct samples of the primary tumor (single-cell prevalence = 0.0083) yet *mutually exclusive* with the clonal *KRAS* mutation. The case could have resulted in similar conclusion as the abovementioned TCGA study if not studied with snDNA-seq, which did not rely on any inference based on CCF but directly measured the clonal colocalization / mutual exclusivity between mutations.

On the other hand, we are reporting a different observation: in 42.8% of PDACs where *KRAS* likely was acquired clonally, subclones with likely **loss** of the mutant *KRAS* allele were identified. We did not find prior literatures reporting such a high rate of *KRAS* mutation loss later in PDAC development, and found it clinically relevant as it may suggest a general pattern of decreased dependence on KRAS oncogenic signaling that may constitute a pre-existing resistance mechanism to KRAS inhibiting therapeutics.

In addition, BRCA2 bi-allelic loss is expected for a tumor suppressor, and its role in PDAC has been previously reported.

We acknowledge that it is expected for a tumor suppressor and its role has been well reported in PDAC. However, there was limited contemporary literatures on its inactivation timing. There was one paper in 2001 [2] that studied 14 pancreatic intraepithelial neoplasia (PanIN) from 3 PDAC patients with germline *BRCA2* mutations and concluded the late timing of *BRCA2* bi-allelic inactivation relative to the oncogenic *KRAS* mutation. Our study (for instance, case PC21) confirmed such an observation using snDNA-seq on PDAC tissues at a more advanced stage. We additionally presented with more granularity that in PC21 the BRCA2 biallelic inactivation preceded other genomic rearrangements such as *TP53* biallelic inactivation and *CDKN2A* homozygous deletion.

We did not find much literature reporting on the inactivation timing of BRCA2 in PDACs with germline BRCA2 mutation and without other canonical driver mutations such as *KRAS, TP53,* possibly due to the rarity of such cases. We presented two such cases – PC01 and PC23 – and showed that *BRCA2* biallelic inactivation was likely the clonal driver that resulted in PDAC in a non-canonical mutational route.

Furthermore, I believe the dataset generated from this moderately sized cohort should be more thoroughly integrated with the broader PDAC omics literature.

We believe in our present manuscript we have compared our results against many relevant PDAC omics literatures. Examples include:

(L139-142) Across the cohort we detected more alterations (including both short mutation and monoallelic/biallelic loss) to CDKN2A and SMAD4 (71% and 71%, respectively), many subclonal, than reported by TCGA (30%, 32%) and ICGC (30%, 27%. Specifically, we detected homdel of CDKN2A in 67% cases and SMAD4 in 25% cases, which are comparable to the numbers (~60% CDKN2A; 40% SMAD4) found in a mixed cohort of primaries/metastases using microdissection to enrich for tumor content [3].

(L151-153) We interpret these findings to reflect the increase in sensitivity by our assay rather than more extensive sampling, as these rates are also higher than reported in a recent study of 91 multiregion sampled PDAC research autopsies [4].

(L179-181) This pattern of early fixation of drivers followed by generation of intratumoral heterogeneity by CNVs is consistent with multiregional bulk sequencing of larger PDAC cohorts [5].

(L203-206) The polyclonal seeding and bifurcating migration correspond to patterns predicted by a single-stage evolution model, where founder cells of the metastatic lesions were apparently randomly sampled from the available lineages within the primary or secondary sites [6].

Additionally, our group had two other studies conducted contemporaneously as the current study – one N=91 studied with multiregionally bulk whole-exome DNA sequencing and another N=1 thoroughly sampled across primary and metastatic sites and studied with single-nucleus RNA-sequencing. These three studies had large overlap in samples but complementing methodologies. We hope that together these three studies could present a comprehensive view into PDAC genomics and transcriptomics across spatial and temporal evolution.

References

[1] Cancer Genome Atlas Research Network. Electronic address: andrew_aguirre@dfci.harvard.edu and Cancer Genome Atlas Research Network, “Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma,” *Cancer Cell*, vol. 32, no. 2, pp. 185-203.e13, Aug. 2017, doi: 10.1016/j.ccell.2017.07.007.

[2] M. Goggins, R. H. Hruban, and S. E. Kern, “BRCA2 Is Inactivated Late in the Development of Pancreatic Intraepithelial Neoplasia,” *Am. J. Pathol.*, vol. 156, no. 5, pp. 1767–1771, May 2000, doi: 10.1016/S0002-9440(10)65047-X.

[3] A. A. Connor *et al.*, “Integration of Genomic and Transcriptional Features in Pancreatic Cancer Reveals Increased Cell Cycle Progression in Metastases,” *Cancer Cell*, vol. 35, no. 2, pp. 267-282.e7, Feb. 2019, doi: 10.1016/j.ccell.2018.12.010.

[4] K. M. Mullen *et al.*, “The Evolutionary Forest of Pancreatic Cancer,” *Cancer Discov.*, vol. 15, no. 2, pp. 329–345, Feb. 2025, doi: 10.1158/2159-8290.CD-23-1541.

[5] A. P. Makohon-Moore *et al.*, “Limited heterogeneity of known driver gene mutations among the metastases of individual patients with pancreatic cancer,” *Nat. Genet.*, vol. 49, no. 3, pp. 358–366, Mar. 2017, doi: 10.1038/ng.3764.

[6] K. Naxerova, “Evolutionary paths towards metastasis,” *Nat. Rev. Cancer*, vol. 25, no. 7, pp. 545–560, Jul. 2025, doi: 10.1038/s41568-025-00814-x.