nature research

Peer Review Information

Journal: Nature Genetics
Manuscript Title: Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment
Corresponding author name(s): Dr. Tyler Jacks, Dr Aviv Regey

Editorial Notes:

Transferred manuscripts - This manuscript has been previously reviewed at another journal.

Redactions – confidential patient information - Parts of this Peer Review File have been redacted as indicated to maintain patient confidentiality.

Reviewer Comments & Decisions:

Decision Letter, initial version:

Dear Will,

6th Oct 2021

Your Article, "Single-nucleus and spatial whole transcriptome profiling of frozen pancreatic cancer reveals multicellular communities and enrichment of a neuronal-like phenotype after neoadjuvant treatment" has now been seen by the original 4 referees. You will see from their comments below that while they continue to 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.

Taken as a whole, we think that the four referees all agree that this paper has improved in revision (for example, the increased sample size and the computational analysis).

Reviewer #1, while happy with the other aspects, thinks that this study's "significance is moderate at this point" and says this is due to a lack of validation. They suggest the paper could be further focused by removing some extraneous analyses.
Reviewer #2 asks for "additional validation" of the novel neuronal program, from which "a lot of the novelty of this paper" arises.
Reviewer #3 seems largely satisfied with the additional work performed, though they have a few remaining requests (e.g. the malignant cell markers used). They think that the conclusions should be toned down further, in light of the limitations of this work.
Reviewer #4 is supportive of publication, asking only for minor additional changes.

Our reading of these reports is that, in sum, they sound supportive of an eventual publication, but we think these remaining comments must be satisfactorily addressed before such. The most important concern raised (by Reviewers #2 and #4), in our mind, is the request for further validation of your

findings; this should be comprehensively addressed in a revision. We note also that the reviewers do not provide specific guidance for what form this validation should take; this means they might be satisfied with further in silico validation, although we leave it to you and your co-authors' judgement as to how this request may best be fulfilled.

To guide the scope of the revisions, the editors discuss the referee reports in detail within the team, including with the chief editor, with a view to identifying key priorities that should be addressed in revision and sometimes overruling referee requests that are deemed beyond the scope of the current study. We hope that you will find the prioritized set of referee points to be useful when revising your study. Please do not hesitate to get in touch if you would like to discuss these issues further.

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.

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

[READCTED]

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,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Referee expertise:

Referees #1, #3: pancreatic cancer.

Referee #2: single cell genomics, spatial transcriptomics.

Referee #4: single cell proteomics.

Reviewers' Comments:

Reviewer #1:
Remarks to the Author:
The authors consent 50 patients to look at 43 PDAC from resected patients.
Many are pretreated. A snRNA variant of scRNA is used. They identify epithelial, fibroblast, endothelial, and immune cells present in PDAC, across all tumors. CAN is used to verify the malignant compartment. Cellular identity and protein expression is compared to the scRNA profiles of individual cells on 7 samples profiled with MIBI.
They describe and ADM-like cell type which may move into an atypical ductal cell type expressing CFTR as well as CK19. The Latter is described as a malignant marker but normal epithelia in the lung and gut express this CK, so this comment is a bit confusing.

The comparisons of the 1,718 cells from Peng et al seems platform specific and perhaps not appropriate to include here, given the vagaries of sample prep and non-overlapping methodologies. It is noted that this is in response to a prior reviewer's concerns.

The argument that some acinar cell ◊ ADM ◊ PanIN / atypical ductal ductal ◊ PDAC over time is logical and appealing but has the logical failing of time. Presumagably, if transformation has occurred and if PDAC cells have a proliferative advantage over atypical ductal and ADM cells, then why are we recovering all of these at the time of resection? Field effect might explain this, but it'd be great to have normal pancreata profiled for these to see what's actually going on here, although this may well be beyond the scope here.

CD8 cells are up with CRTL and Treg went down. This is an underdeveloped but potentially interesting sidebar enabled by the greater patient numbers here on trial and profiled. Of note, the individual samples are unmatched.

RNA subtypes are promiscuous, overlapping in a single cell, or absent altogether. Fibroblast signatures from others were also not clearly recapitulated so the authors embark on a new cNMF-driven subtype finding exercise. They find Classical subtype cells as well as individual subtypes of mesenchymal, basal and squamoid that are distinct. They also find cells that are malignant by CNA and express an exocrine or endocrine transcriptional profile. A neuronal programme (distinct from the neuroendocrine programme) is also a unique sample type. Seven other cell state subtypes are also described.
4 or so fibroblast types are described as well. These are leveraged in interpretation of treatments applied to patients over time in a variety of neoadjuvant maneuvers (CRT, CRTL, etc.). A theme, if any can be gleaned, is a diminution of classical gene expression patterns in surviving (?regrowing) cells after treatment.

A classical and malignant cell population as well as an immunomodulatory CAF program was prognostically good when superimposed on TCGA samples. Neuronal gene expression in malignant cells was prognostically poor.

Digital special profiling (DSP) is done with Nanstring WTA to assess where in the tumor these different cells reside. Fig 3b shows that these cells are disperse into apparent treatment communities, some of which are enriched with treatment.

Finally, we get a glimpse into Ligand receptor pairs. CXCL12 and CXCR4 pair strongly in this analysis. The malignant compartment has RL pairs with both the fibroblasts as well as themselves. ERBB2 is of note here, possibly as a resistance mediator given its enrichment in CRT samples.

This is a technical tour de force. Lots of information here with some solid conclusions.
The authors are to be congratulated for collecting these samples in a mindful way, processing them to optimize data acquisition and then analyzing them with rigor and creativity.

The work scores very highly in terms of originality, the technical aspects as well as the novel findings make this very original.

Significance is moderate at this point. Without validation of the many findings in this dataset, it is not clear in which direction the field should move, but there are several hypotheses raised here which could have a large significance.

Data and methodology are innovative. One concern is overfitting of the subtypes and the opacity between the # of cell types discovered and the number of samples analysed. This could be clarified. The comparison of their snRNA success rate to others' scRNA might be best in discussion and not as a primary result and this is not a robust interrogation given the control or lack thereof.

The immune analysis is very underdeveloped and perhaps leaving it out is best.

This reader found comparisons to outside sets of scRNA to be unconvincing.

the cell communities

Fig 2b , "Patients" seems to imply that we are nowhere near consensus / saturation, and that each patient seems to dominate a cluster. Can the authors comment on what saturation of cell types in the tumor would look like? Since we discover so many unique subtypes, one worries that anther 50 samples would yield another 17 subtypes.

The technique used to map spatial heterogeneity using the Nanostring WTA is very complex and difficult to follow.

Reviewer #2:
Remarks to the Author:
In this manuscript, Hwang et al have performed snRNA-Seq and Nanostring GeoMX on a large cohort of archived PDAC samples. The authors are commended for collecting and presenting an impressive dataset, and performing sophisticated analyses on the data. The authors have also addressed several comments put forth in the previous round of reviews and this has greatly improved the manuscript. Figure 4c in particular very nicely sums up the results and provides a compelling description of the contributions of this manuscript.

We have the following comments regarding the current version of the manuscript.

Related to figure 1: There seems to be discrepancies with the number of samples for snRNA-seq between the figure and the results section. Line 142 describes 43 snRNA-Seq samples while Figure 1a describes 45 scRNA-Seq.

In Figure 1d,e, and f, the authors claim a trajectory among the "normal" epithelial cells, adm and atypical ductal cells, with the latter two being PDAC intermediate states. However, further evidence would be required to support this claim. The analysis shown is at the level of all patients but it would be useful to understand if these states exist in all patients. A UMAP in Figure 1d for each cell type labeled with a patient sample would address this question. Furthermore, does treatment have an effect on the prevalence/occurrence of these states?

Additional evidence to support the claim about the novel neuronal program: Given that a lot of the novelty of the paper stems from its delineation, additional validation would be required to claim its existence.

The Figure 3 analysis is very interesting.

Can the authors explain Figure 4c? It is only cited in the results in a short sentence.

One suggestion: Figure 4a seems a bit disorganized, and it appears that rearranging the rows and columns would better show the clusters.

Minor comment: Line 175 refers to fig 1e but should it refer to figure 1g?

Reviewer #3:
Remarks to the Author:
Hwang et al present a revised version of the previous manuscript that looks almost like a new paper. The authors should be congratulated for expanding their study to include more cases both in the snRNA-Seq analysis and in the DSP work. In addition, I believe that have made a substantial effort to respond to my previous comments and to improve the quality of the paper, most notably as it refers to the accuracy of the statements made in the Results section. Overall, therefore, this is a highly improved manuscript that will have a significant impact in the field.

General comments

Notwithstanding these considerations, some of the major issues raised in my review have not been overcome (and it might not be possible to overcome them at the present time). These issues need to be addressed, in my opinion, in a paragraph on the limitations of the present work at the end of the discussion. Again, and despite the efforts made, there is a need to be modest and prudent with the conclusions of the results presented and I highlight here some of the main general issues I have with the paper:
- a large number of comparisons are made in some analyses and even if the authors have placed emphasis on the improved statistical analysis, the possibility that the occurrence of significant p-values might be inflated due to the multiple comparisons made cannot be dismissed (e.g. several panels from Figure 2 and ED Suppl. Figs. 12,13);
- the study is based on an enlarged, but still small and very heterogeneous, set of samples which means that caution is mandatory. Furthermore, this is likely to be a biased series, even if unintentionally, as most retrospective series are. For example: while interesting, a group of 5 patients treated with CRTL, is a very tiny group....
- many bioinformatics parameters and technical limitations derived from the use of snRNA-Seq impact on the robustness of the cell types called and the new classifications. What is really important is whether the latter can be confirmed in independent studies and they really exist in the patient (not that I am questioning it....);
- this caution should be applied to the title as this is one of the aspects of the paper that will have wide visibility.

Specific comments

1. Fig. 1e-g are often referred to in the text concerning data that I do not see shown in the corresponding panel.
2. Lines 174-179: what is the possible impact of these biases? and their cause?

3. Lines 206-208: the markers used here to highlight malignant cells are inadequate: KRT19 is expressed in all ductal cells (normal and tumoral) and KRT17 is expressed in a selected subset of non-neoplastic ductal cells, including the pancreas of organ donors (PMID 7686885).
4. Lines 227-230 and several other paragraphs: it is not clear that the authors have adjusted the p-values of comparisons in figure 2 - among others - for multiple testing, as would be required when a priori hypotheses are not considered.
5. Ext. Fig. 8a: it is striking that the signatures from the Chang-Seng-Yue paper are so poorly represented in this series' data. Any explanation?
6. Lines 265 and subsequent: not clear how the adequacy of the number of programs chosen was assessed, both for tumor cells and for fibroblasts.
7. Designation of programs/cell types: I think that the authors need to be cautious here again. For example, their list of the 200 top genes from ED Table 2 for the neuronal subtype, when run on GO pathways, top annotations are "central nervous system" or "brain development" with modest adjusted p-values; however, when run on KEGG pathways, the top annotations include "bile secretion" and "pancreatic secretion" with similar adjusted p-values. I have not run analyses for many other types of questions in the paper but similar issues might come up. Therefore, I would be careful because once these "statements" are made in a published paper in an excellent journal, it becomes "dogma" while the knowledge is not necessarily more solid.
8. Lines 401-403: a standard definition of response to neoadjuvant therapy should be used rather than the "poor, minimal, or moderate" grading which is not clinically acceptable.
9. Lines 412-419 and following: the use of the PanCuRx dataset to analyze signatures derived from non-tumor cells is not appropriate since this series comprises laser-microdissected samples, generally with >70% tumor cells. The survival analyses are deeply limited by treatment heterogeneity and sample size. In fact, the only clearly significant association with outcome is observed for the classical subtype (Fig. 2e), a well-established association in the literature.
10. Lines 477-495: the limited contribution of the novel tumor cell classification to our understanding of PDAC heterogeneity is highlighted by the fact that two of the tree communities proposed actually correspond largely to previously described PDAC subytpes. The third one (treatment-enriched) is defined through the subtypes that were less robustly defined. These findings call, again, for a cautious revision of the conclusions of the paper.
11. In my opinion, much of the content of the last section of the Results should be presented differently, avoiding excessive speculation which is not based on mechanistic experiments and which could be summarized in a Table, either in the main text or as supplementary material.

Minor comments

1. Lines 91-94: I don't think that the issue here is whether these "normal genes" have a role in cancer, this should be rephrased.
2. Lines 139-140: specify here what about the 7 "missing" cases.
3. The font of some of the panels in the figures should be carefully revised to make sure that the words can be read easily.

Reviewer #4:
Remarks to the Author:
Hwang et al. present a substantially revised manuscript describing their snRNA-seq methods and how they were able to derive clinically meaningful molecular signatures for PDAC – many for the first time.

The heavily revised manuscript is much more streamlined with more simplified PDAC molecular signature hazard analysis and additional, more granular molecular and spatial proteomic analysis. All of this is underpinned by an almost doubling of the data (patient samples) included in the original manuscript. This is a much-improved study compared to the original submission and should serve as standard bearer for multi-omic clinical studies and data integration of complementary features from different analyses.

One minor critique that could possibly improve the manuscript further would be to include the hazard analysis for one or two previously published PDAC molecular signature into the revised 2E. Even though it is clear the new signatures identified here have significant and complementary clinical implications it would still be good for readers to understand how this is couched within previous knowledge by having a few points of comparison.

Also minor, Figure 2 would be even clearer if the variable abbreviations from 2E were included in the plot titles of 2B.

Author Rebuttal to Initial comments

Response to Reviewers

Overview

We thank the Reviewers and editors for their thoughtful feedback and guidance. In particular, the editors have guided us that "the most important concern raised (by Reviewers #2 and #4) is the request for further validation of [our] findings."

In the revised manuscript, we have addressed the **suggestion for further validation of our findings, especially the novel neuronal-like program, in several ways** that we summarize here and describe in more detail in the point-by-point response below.

Briefly, we focused our validation of the novel neuronal-like malignant program on (**1**) its detailed features and annotation, and (**2**) its treatment-induced enrichment.

1. Detailed features and annotation. We have validated the most highly-weighted genes and gene pathways that define the neuronal-like program by several **new experiments and analyses**. First, we found a significant program enrichment for annotated brain-enhanced genes in the **Human Protein Atlas** (**new Fig. 2c**). Furthermore, the program contains **NFIB**, a transcription factor that promotes pro-metastatic neuronal gene expression programs in other cancer types [Denny et al, Cell 2016]. Next, we performed **multiplexed immunofluorescence** on an independent PDAC specimen and demonstrated heterogeneous co-expression of cytokeratins (epithelial marker) and neurexin-3 at the single-cell and gland levels (**new Fig. 2d**). We chose neurexin-3 because it is a brain-enhanced adhesion protein that is involved in synaptic plasticity and typically expressed in neurons and neuropil (**Fig. 2c**), weighted strongly in the neuronal-like program, and one of the most differentially-expressed genes in CRT relative to untreated specimens (P_{adj} = 1.25 x 10^{-7}; **Extended Data Fig. 6**). **These results further support our annotation of this program as "neuronal-like**." Because the program also includes additional features beyond the neuronal ones, especially for early tissue development and a potential progenitor state (*e.g.*, enrichment for gene sets for hepatocyte nuclear factor activity, adult tissue stem cell modules, and organ morphogenesis), we now name it "**neural-like progenitor**" throughout the manuscript.

2. Treatment-induced enrichment. To further test the impact of treatment, we **derived organoids from an untreated tumor (PDAC_U_12) and treated them with ex vivo FOLFIRINOX-like chemotherapy and radiotherapy**. We performed snRNA-seq on pre- and post-treatment organoids and found a significant enrichment in the neural-like progenitor score in the latter (**new Fig. 3d**).

Our new analyses and experiments strongly supports our conclusion that the novel neural-like progenitor malignant program is appropriately annotated and that chemotherapy and radiotherapy are capable of enriching expression of this program in residual cancer cells that survive treatment.

Response to Reviewer 1

The authors consent 50 patients to look at 43 PDAC from resected patients.
Many are pretreated. A snRNA variant of scRNA is used. They identify epithelial, fibroblast, endothelial, and immune cells present in PDAC, across all tumors. CAN is used to verify the malignant compartment. Cellular identity and protein expression is compared to the scRNA profiles of individual cells on 7 samples profiled with MIBI.

We thank the Reviewer for their comments, thoughtful suggestions, and appreciation of our work.

They describe and ADM-like cell type which may move into an atypical ductal cell type expressing CFTR as well as CK19. The Latter is described as a malignant marker but normal epithelia in the lung and gut express this CK, so this comment is a bit confusing.

We thank the Reviewer for this comment and regret the confusion. Although CK19 is indeed expressed in normal epithelia in certain tissues, prior studies have shown that it is significantly overexpressed in malignant PDAC relative to non-malignant epithelia [Yao et al, Cancer Biomark 2016]. Consistent with prior work, KRT19 has markedly higher expression and CFTR has significantly lower expression in the malignant relative to non-malignant epithelial cells (**Fig. 1b**).

To further address the Reviewer's question, we now also present the expression of other malignant-enriched markers, including KRT5 (which is more malignant-specific; **Fig. 1b**), as well as KRT17, KRT7, KRT6A, KRT14, TACSTD2, S100A11, S100A16, TFF1, and CLDN18 (**Fig. R1**, **Methods,** page 27). These show highest expression in malignant cells, and intermediate expression in ductal (atypical) and ADM cells.

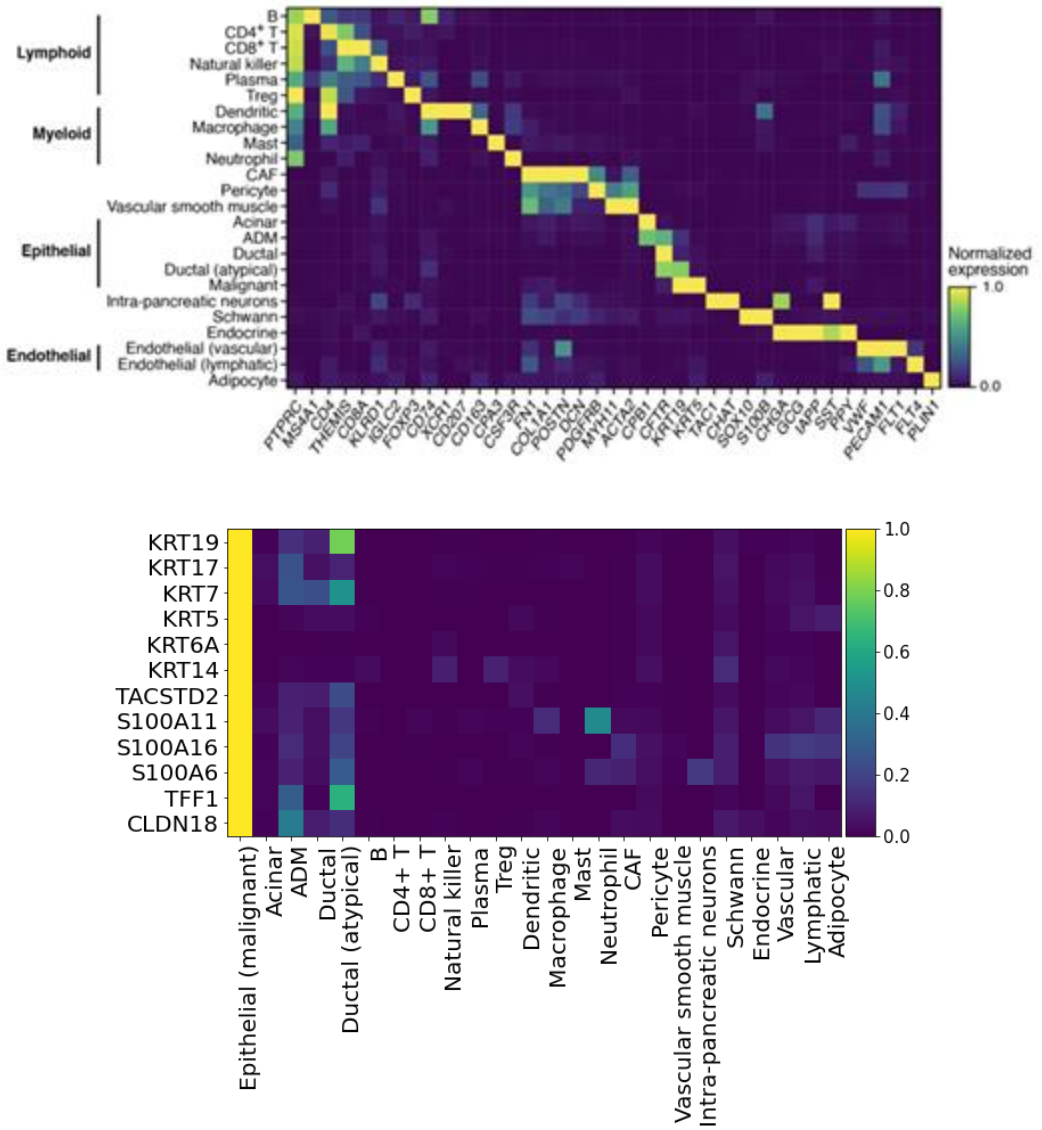

Fig. R1. Expression of key malignant cell markers in malignant, ADM and other cell subsets.

The comparisons of the 1,718 cells from Peng et al seems platform specific and perhaps not appropriate to include here, given the vagaries of sample prep and non-overlapping methodologies. It is noted that this is in response to a prior reviewer's concerns.

We thank the Reviewer for this comment. Indeed, this analysis was included at the specific request of another Reviewer. We agree that there could be multiple reasons why we were able to achieve an average of 6,054 nuclei per sample, relative to 1,718 cells per sample described in the Peng et al. study, despite

comparable quantities of loaded nuclei/cells onto the 10x Chromium platform. However, we believe this comparison may nonetheless be important to readers, in light of the very limited number of single-cell/nucleus extraction protocols for PDAC, and the historical challenges of extracting sufficient high-quality cells from this tissue. An estimate of the expected number of high-quality cells or nuclei is helpful to other experimentalists when they select protocols or attempt to implement them, and helps determine the relative strengths and weaknesses of different approaches. We further facilitated the comparison by reannotating the Peng et al. profiles with our snRNA-seq cell type signatures. Following the Reviewer's comment, we have moved the discussion of this comparison from the **Results** to the **Methods** (page 27) and refer to the analysis (**Extended Data Fig. 19**) in the **Discussion** (page 17).

The argument that some acinar cell ADM PanIN / atypical ductal ductal PDAC over time is logical and appealing but has the logical failing of time. Presumably, if transformation has occurred and if PDAC cells have a proliferative advantage over atypical ductal and ADM cells, then why are we recovering all of these at the time of resection? Field effect might explain this, but it'd be great to have normal pancreata profiled for these to see what's actually going on here, although this may well be beyond the scope here.

We thank the Reviewer for this important question. We agree with the Reviewer that it is interesting that we observe the presence of ADM and atypical ductal populations in PDAC specimens, and concur that field effect may at least partly explain their coexistence. In addition, we note that relative to malignant cells, these putative intermediate states along the malignant transformation spectrum exist at a low prevalence within the untreated specimens. For instance, the ADM and atypical ductal populations represent only 1.2% and 2.4% of all untreated epithelial cells, respectively (*vs*. 83.3% and 16.7% for malignant and non-malignant epithelial cells, respectively). We do observe greater proportions of non-malignant epithelial cells in the treated cohort (63.0%), but this is consistent with the enhanced cytoxicity of rapidly-dividing cancer cells, as well as regenerating exocrine pancreas post-treatment. We agree with the Reviewer that further examination of normal pancreata toward this end would be outside the scope of this specific study but nonetheless worth pursuing in another context. We note this in the revised **Discussion** (page 17).

CD8 cells are up with CRTL and Treg went down. This is an underdeveloped but potentially interesting sidebar enabled by the greater patient numbers here on trial and profiled. Of note, the individual samples are unmatched.

We agree with the Reviewer that the finding is interesting, which prompted us to include the data with an eye towards stimulating further study. However, given word count limitations and the fact that quantitative immune changes with treatment is not a focus of this study, we have condensed the description of these results and merged it with the preceding section.

RNA subtypes are promiscuous, overlapping in a single cell, or absent altogether. Fibroblast signatures from others were also not clearly recapitulated so the authors embark on a new cNMF-driven subtype finding exercise. They find Classical subtype cells as well as individual subtypes of mesenchymal, basal and squamoid that are distinct. They also find cells that are malignant by CNA and express an exocrine or endocrine transcriptional profile. A neuronal programme (distinct from the neuroendocrine programme) is also a unique sample type. Seven other cell state subtypes are also described.
4 or so fibroblast types are described as well. These are leveraged in interpretation of treatments applied to patients over time in a variety of neoadjuvant maneuvers (CRT, CRTL, etc.). A theme, if any can be gleaned, is a diminution of classical gene expression patterns in surviving (?regrowing) cells after treatment.

A classical and malignant cell population as well as an immunomodulatory CAF program was prognostically good when superimposed on TCGA samples. Neuronal gene expression in malignant cells was prognostically poor.

Digital special profiling (DSP) is done with Nanostring WTA to assess where in the tumor these different cells reside. Fig 3b shows that these cells are disperse into apparent treatment communities, some of which are enriched with treatment.

Finally, we get a glimpse into Ligand receptor pairs. CXCL12 and CXCR4 pair strongly in this analysis. The malignant compartment has RL pairs with both the fibroblasts as well as themselves. ERBB2 is of note here, possibly as a resistance mediator given its enrichment in CRT samples.

This is a technical tour de force. Lots of information here with some solid conclusions.
The authors are to be congratulated for collecting these samples in a mindful way, processing them to optimize data acquisition and then analyzing them with rigor and creativity.

The work scores very highly in terms of originality, the technical aspects as well as the novel findings make this very original.

We are grateful for the Reviewer's appreciation of the technical aspects, conclusions, innovation, and creativity in our work.

Significance is moderate at this point. Without validation of the many findings in this dataset, it is not clear in which direction the field should move, but there are several hypotheses raised here which could have a large significance.

We agree with the Reviewer that additional validation is warranted in this study as well as follow-up investigations. Following the Reviewer's feedback, we focused additional validation efforts on our discovery of the **neuronal-like malignant program**. In the revised manuscript, we provide **new experiments and analyses** to validate the novel neuronal-like malignant program.

1. Detailed features and annotation. We validated the neuronal-like program in several ways. First, we show that the program is enriched for annotated brain-enhanced genes in the **Human Protein Atlas** (**new Fig. 2c**). Next, **multiplexed immunofluorescence** on an independent PDAC specimen demonstrated heterogeneous co-expression of cytokeratins (epithelial marker) and neurexin-3 at the single-cell and gland levels (**new Fig. 2d**). Neurexin-3 is a brain-enhanced adhesion protein that is involved in synaptic plasticity and typically expressed in neurons and neuropil (**Fig. 2c**), is weighted strongly in the neuronal-like program, and is one of the most differentially-expressed genes in CRT relative to untreated specimens ($P_{adj} = 1.25 \times 10^{-7}$; **Extended Data Fig. 6**). These data support our annotation of this program as "neuronal-like." Notably, because other program genes are enriched for features of early tissue development and progenitor state (e.g., hepatocyte nuclear factor activity, adult tissue stem cell modules, and organ

morphogenesis), we now name the program "neural-like progenitor" throughout the manuscript to highlight these additional nuances.

2. Impact of treatment. To further test the impact of treatment and address the lack of matched pre- and post-treatment specimens in our cohort, we **derived organoids from an untreated tumor (PDAC_U_12) and treated them with *ex vivo* FOLFIRINOX-like chemotherapy and radiotherapy**. We performed snRNA-seq on pre- and post-treatment organoids and found a significant enrichment in the neural-like progenitor score in the latter (**new Fig. 3d**).

Taken together, our validation experiments and analyses support our conclusion of a novel neural-like progenitor malignant program that is enriched in residual cancer cells that survive chemotherapy and radiotherapy.

Data and methodology are innovative. One concern is overfitting of the subtypes and the opacity between the # of cell types discovered and the number of samples analysed. This could be clarified.

We thank the Reviewer for these comments, which we address in turn.

With regard to **malignant cell subtypes**, we performed consensus NMF and selected a number of components/programs that **optimized for the stability** of the computed solution. For the malignant and fibroblast compartments, this led to a total of 49 and 4 components/programs, respectively (**Fig. R2** and **new Extended Data Fig. 9a**). As noted by the Reviewer, while many of these programs were patient-specific, we focused on those that were biologically distinct and distributed across multiple patients (**Fig. R3**). While prior studies had identified fewer malignant subtypes, they relied on bulk profiles that allow detecting majority cell states distinct between subsets of patients, while our single cell resolution can not only find such majority states, but also identify *minority* cell states shared across multiple patients, which are otherwise obscured at the bulk level. While it is theoretically possible that additional programs could be discovered if more samples were included, they would likely exist at very low prevalence. Indeed, when we performed cNMF with subsampling of the samples (patients) in our cohort, we recovered all 14 malignant programs at 80% subsampling and all 4 fibroblast programs at 50% subsampling, suggesting that we are at or near saturation with the current cohort size and that further increases to the cohort size

are unlikely to yield additional core programs in either cell type (**Fig. R4** and **new Extended Data Fig. 9b**).

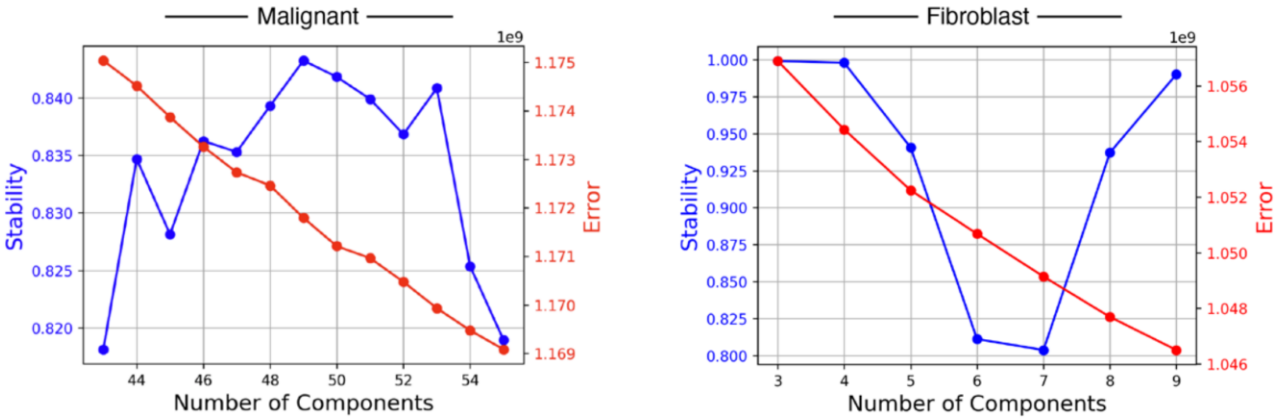

Fig. R2. Impact of number of programs (components, x axis) on the stability (blue, left y axis) and error (red, right y axis) of the solution for malignant cells (left) and fibroblasts (right).

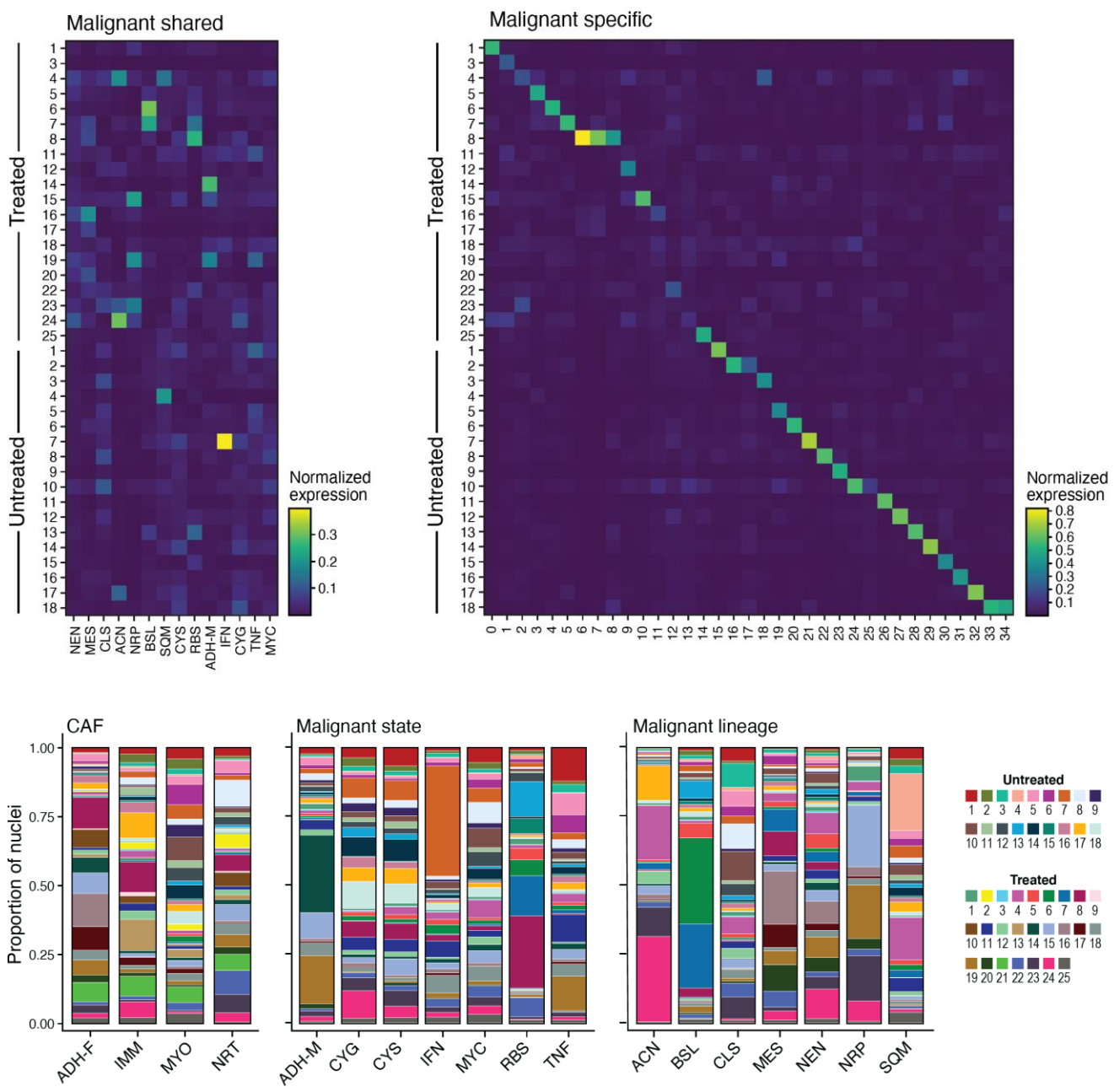

Fig. R3. (Top) Normalized mean expression (color scale) of each distributed program (left columns with names) and patient-specific program (right columns with numbers) for all patient tumors (rows) with at least 10 malignant cells (left) or fibroblasts (right). (Bottom) Proportion of nuclei (y axis) contributed by each patient tumor (color legend) to each CAF (left, x axis), malignant state (middle, x axis), and malignant lineage (right, x-axis) program.

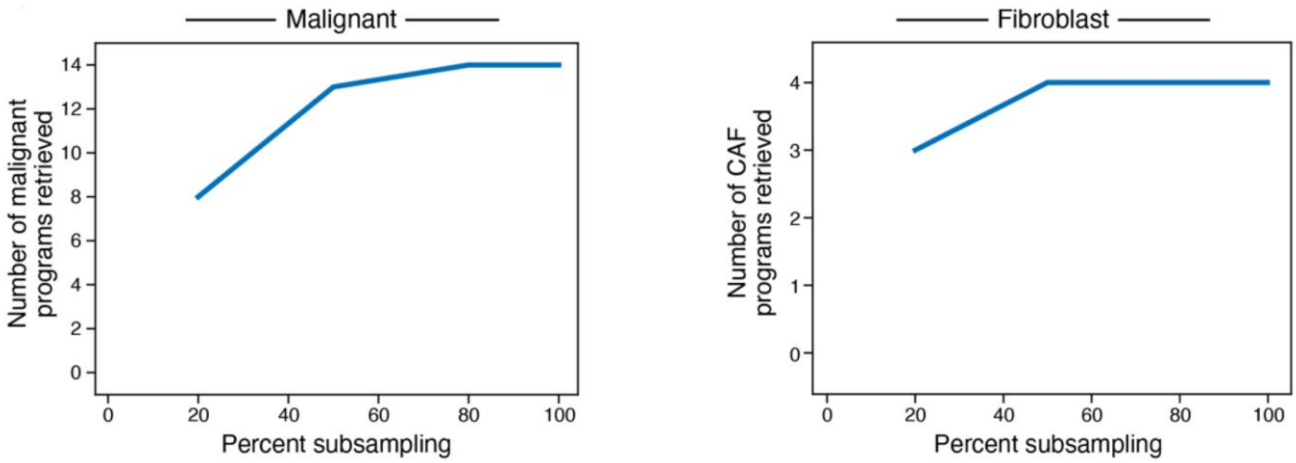

Fig. R4. Number of malignant (out of 14; left) and CAF (out of 4; right) programs recovered in the cNMF solution learned with different degrees of subsampling of patients from our cohort.

The comparison of their snRNA success rate to others' scRNA might be best in discussion and not as a primary result and this is not a robust interrogation given the control or lack thereof.

We thank the Reviewer for this suggestion. As noted above, this analysis was added at the request of another Reviewer and our goal was to help users understand the different features of various protocols. We facilitated the comparison by reannotating the Peng et al. data. Following the Reviewer's comment, we have moved this comparison from the **Results** to the **Methods** section and refer to it in the **Discussion** (**Extended Data Fig. 19**, page 17).

The immune analysis is very underdeveloped and perhaps leaving it out is best.

We agree with the Reviewer and we have condensed the description of quantitative immune changes with treatment considerably, merging it with the prior section rather than presenting it in a separate section as before. We did not remove it altogether so that interested readers will be able to follow on this direction with our data.

This reader found comparisons to outside sets of scRNA to be unconvincing.

As noted above, these analyses were at the request of another Reviewer. We have moved these from the **Results** to the **Methods** section and refer to them from the **Discussion** (**Extended Data Fig. 19**, page 17).

the cell communities

Fig 2b , "Patients" seems to imply that we are nowhere near consensus / saturation, and that each patient seems to dominate a cluster. Can the authors comment on what saturation of cell types in the tumor would look like? Since we discover so many unique subtypes, one worries that another 50 samples would yield another 17 subtypes.

We thank the Reviewer for this question.

First, for context, multiple scRNA-seq studies across many tumor types have shown that in patients, malignant cells first separate by patient (e.g., from our own work: Patel et al *Science* 2014 in GBM; Tirosh et al *Science* 2016 in melanoma; Puram et al *Cell* 2017 in HNSCC, Izar et al *Nature Medicine* 2020 in ovarian cancer, Jerby et al *Nature Medicine* 2021 in synovial sarcoma, Pelka et al *Cell* 2021 in CRC; but also studies from many other labs and tumor types). This is not a technical artifact as in the same studies, non-cancer cells in the TME group by type, with clusters of cells from multiple patients. This is due to several factors, including some impact of SNVs on gene expression, as we have shown in the controlled setting of animal models (Marjanovic et al., *Cancer Cell* 2019), as well as other factors. However, it was also shown in multiple tumor types (including all the above cited studies), that analyses that focus on the gene program level (rather than cell clustering) can readily recover the shared expression features across patients, and that the number of such shared programs is much smaller.

This is what we observe in PDAC as well. In **Fig. 2b**, cells from each patient generally partition into separate clusters, and yet consensus NMF identified both gene programs that were unique to patients and programs that were shared among cells from multiple patients (**Fig. R3**). We focus on the latter: 14 of 49 malignant programs, and 4 of 4 fibroblast programs. We further selected the number of programs/components based on maximizing the stability of the resultant cNMF solution (**Fig. R2** and **new Extended Data Fig. 9a**). While there could in principle be other programs that we did not capture, when we performed cNMF with subsampling of our patient samples in our cohort, we recovered all 14 malignant

programs at 80% subsampling and all 4 fibroblast programs at 50% subsampling, suggesting that we are at or near saturation with the current cohort size and that further increases to the cohort size are unlikely to yield additional core programs in either cell type (**Fig. R4** and **new Extended Data Fig. 9b**).

The technique used to map spatial heterogeneity using the Nanostring WTA is very complex and difficult to follow.

We regret the complexity and lack of clarity in our description and have revised the relevant **Results** section (page 14-15, **Fig. 4b**). Specifically, our goal is to ask if the variation we observe between patients is higher than the spatial variation we observe within a patient tumor. To answer this, we compared the dispersion (interquartile range) in program scores within individual tumors (intra-patient, when we compare different spatial niches (ROIs)) to the dispersion across different tumors (inter-patient). Overall, we found greater inter-patient heterogeneity in the expression of most snRNA-seq programs across tumors than intra-patient heterogeneity spatially across ROIs within individual tumors. The only exceptions were the mesenchymal, immunomodulatory, and myofibroblastic progenitor programs. We have also added clarifying details to the **Methods** section (page 38).

Response to Reviewer 2

In this manuscript, Hwang et al have performed snRNA-Seq and Nanostring GeoMX on a large cohort of archived PDAC samples. The authors are commended for collecting and presenting an impressive dataset, and performing sophisticated analyses on the data. The authors have also addressed several comments put forth in the previous round of reviews and this has greatly improved the manuscript. Figure 4c in particular very nicely sums up the results and provides a compelling description of the contributions of this manuscript.

We thank the Reviewer for their appreciation of our revised work, and are grateful for their help in improving our study and manuscript.

We have the following comments regarding the current version of the manuscript.

Related to figure 1: There seems to be discrepancies with the number of samples for snRNA-seq between the figure and the results section. Line 142 describes 43 snRNA-Seq samples while Figure 1a describes 45 scRNA-Seq.

We apologize for any confusion regarding the number of snRNA-seq samples in **Fig. 1a** and the text. There were indeed 45 snRNA-seq samples total, but, as specified in **Fig. 1a**, two of those samples were non-malignant pancreas tissue. Therefore, on line 142 and elsewhere, we refer to the 43 malignant snRNA-seq samples that were the subject of downstream analyses. We have clarified this in the text on page 4.

Unrelated to this comment, however, we did find a typographical error in **Fig. 1a** that we have now corrected. The total number of specimens included in the updated manuscript is **48** rather than 50. The two additional patient specimens (one CRT, one other treatment) were analyzed by DSP using only the more limited cancer transcriptome assay (~1,800 plex) rather than the whole transcriptome assay and are now omitted from the final analyses/manuscript. We have corrected **Fig. 1a**.

In Figure 1d, e, and f, the authors claim a trajectory among the "normal" epithelial cells, adm and atypical ductal cells, with the latter two being PDAC intermediate states. However, further evidence would

be required to support this claim. The analysis shown is at the level of all patients but it would be useful to understand if these states exist in all patients. A UMAP in Figure 1d for each cell type labeled with a patient sample would address this question. Furthermore, does treatment have an effect on the prevalence/occurrence of these states?

We thank the Reviewer for this great suggestion. Following the Reviewer's comment, we have added **new Extended Data Fig. 7** (also presented as **Fig. R5** below) with **(1)** UMAPs pseudo-colored by patient ID for each non-malignant epithelial cell type (acinar, acinar-REG+, ADM, ductal, atypical ductal); **(2)** stacked proportion bar plots showing the proportion of cells in each tumor for each epithelial cell subset and the proportion of epithelial cell subsets for each tumor; and **(3)** stacked proportion bar plots showing the distribution of epithelial subpopulations (denominator = total epithelial cells) as a function of treatment status (untreated, CRT, CRTL).

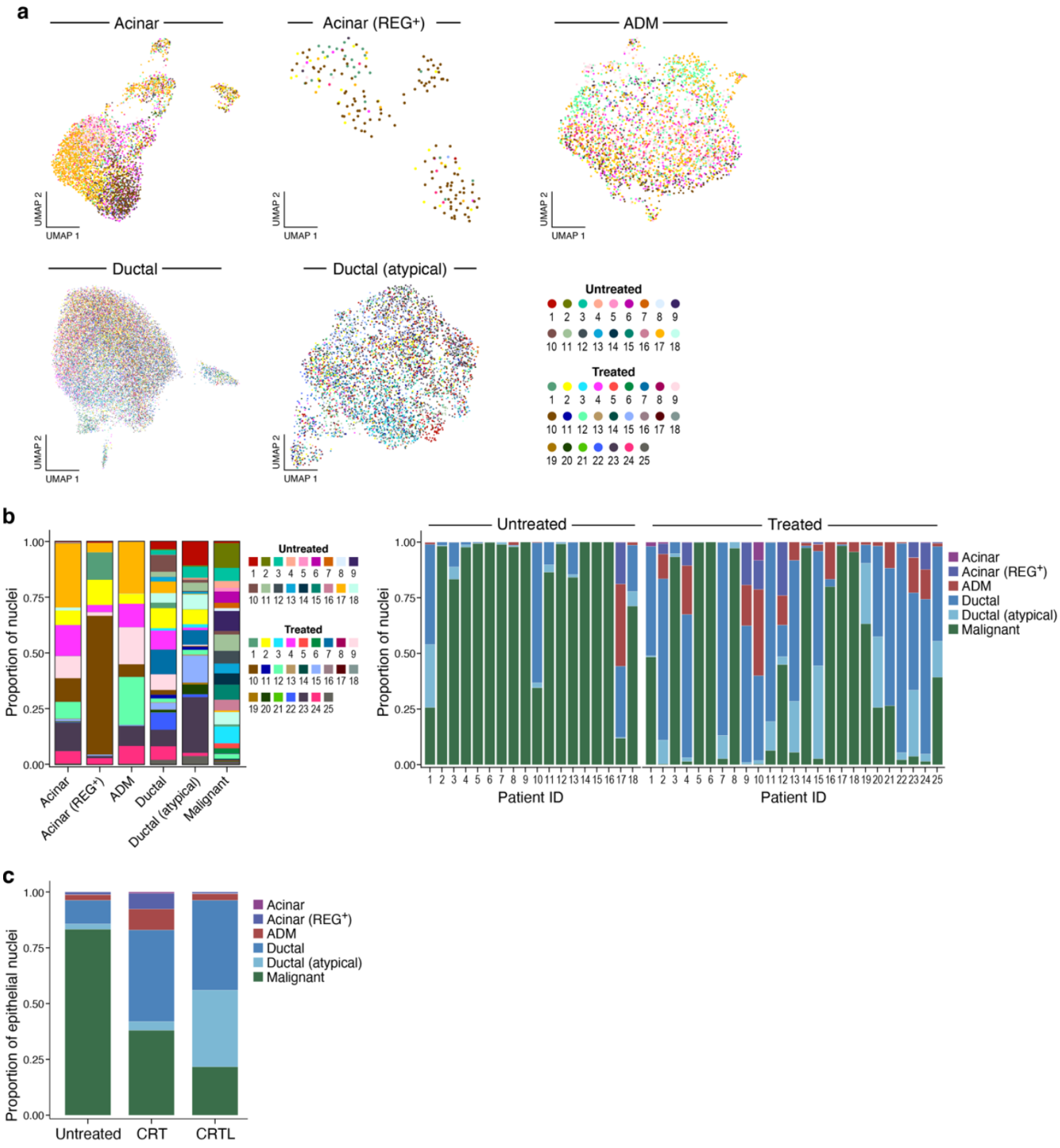

Fig. R5. Epithelial cell type composition across PDAC tumors. a, UMAP embeddings of single nucleus profiles (dots) for different epithelial cell subsets (panels) colored by patient ID (color legend). **b**, Left: proportions (y axis) of cells in each tumor (color legend) for each epithelial cell subset (x axis); Right: proportions (y axis) of epithelial cell subsets (color legend) for each tumor (x axis). **c**, Proportions (y axis) of epithelial cell subsets (color legend) summed across all tumors for each treatment category (x axis).

Additional evidence to support the claim about the novel neuronal program: Given that a lot of the novelty of the paper stems from its delineation, additional validation would be required to claim its existence.

We thank the Reviewer for their appreciation of the discovery of the **neuronal-like malignant program.** Following the Reviewer's suggestion, in the revised manuscript we validate the program with **new experiments and analyses**.

1. Validation of the program features and annotation. First, we show that the program is enriched for annotated brain-enhanced genes in the **Human Protein Atlas** (**new Fig. 2c**). Next, **multiplexed immunofluorescence** on an independent PDAC specimen demonstrated heterogeneous co-expression of cytokeratins (epithelial marker) and neurexin-3 at the single- cell and gland levels (**new Fig. 2d**). Neurexin-3 is a brain-enhanced adhesion protein that is involved in synaptic plasticity and typically expressed in neurons and neuropil (**Fig. 2c**), is weighted strongly in the neuronal-like program, and is one of the most differentially-expressed genes in CRT relative to untreated specimens (P_{adj} = 1.25 x 10^{-7}; **Extended Data Fig. 6**). These data support our annotation of this program as "neuronal-like." Notably, because other program genes are enriched for features of early tissue development and progenitor state (e.g., hepatocyte nuclear factor activity, adult tissue stem cell modules, and organ morphogenesis), we now named the program "neural-like progenitor" throughout the manuscript to reflect these additional nuances.

2. Validation of impact of treatment. To further test the impact of treatment, and address the lack of matched pre- and post-treatment specimens in our cohort, we **derived organoids from an untreated tumor (PDAC_U_12) and treated them with *ex vivo* FOLFIRINOX-like chemotherapy and radiotherapy**. We performed snRNA-seq on pre- and post-treatment organoids and found a significant enrichment in the neural-like progenitor score in the latter (**new Fig. 3d**).

Overall, our validation experiments and analyses support our conclusion of a novel neural-like progenitor malignant program that is enriched in residual cancer cells that survive chemotherapy and radiotherapy.

The Figure 3 analysis is very interesting.

We thank the Reviewer for their appreciation of these results.

Can the authors explain Figure 4c? It is only cited in the results in a short sentence.

We appreciate the Reviewer's question. This panel (now **Fig. 5c**) is a summary of the various associations we found for each of the malignant lineage programs. As there is no new data presented here, we now cite this summary figure in the **Discussion** (page 17). Briefly, this includes a summary of the detailed classification of each malignant lineage program and its associated state programs, CAF phenotypes, immune milieu, and clinical outcomes.

One suggestion: Figure 4a seems a bit disorganized, and it appears that rearranging the rows and columns would better show the clusters.

We thank the Reviewer for this helpful suggestion. We have reorganized this figure (**new Fig. 5a**) so that the immune cell type associations with expression programs are more visually apparent and use the same row order on the two panels, which emphasizes the similarities (and some distinctions) in organization.

Minor comment: Line 175 refers to fig 1e but should it refer to figure 1g?
We thank the Reviewer for pointing out this typographical error, which has now been corrected.

Response to Reviewer 3

Hwang et al present a revised version of the previous manuscript that looks almost like a new paper. The authors should be congratulated for expanding their study to include more cases both in the snRNA-Seq analysis and in the DSP work. In addition, I believe that have made a substantial effort to respond to my previous comments and to improve the quality of the paper, most notably as it refers to the accuracy of the statements made in the Results section. Overall, therefore, this is a highly improved manuscript that will have a significant impact in the field.

We thank the Reviewer for their appreciation of our work and appreciate their comments that helped us improve our study.

General comments

Notwithstanding these considerations, some of the major issues raised in my review have not been overcome (and it might not be possible to overcome them at the present time). These issues need to be addressed, in my opinion, in a paragraph on the limitations of the present work at the end of the discussion.

We thank the Reviewer for this feedback and have now included more details on the limitations of the study throughout the **Discussion**. In particular, we highlight several limitations. First, we note that snRNA-seq may not be optimized to capture particular immune subsets based on our comparison of immune cell type proportions in snRNA-seq and MIBI (**new Fig. 1e**; **new Extended Data Fig. 3**). Second, we emphasize that the lack of matched pre- and post-treatment specimens limits the strength of conclusions that can be drawn from treatment group comparisons. Third, we note limitations in sample size for certain treatment groups and treatment heterogeneity.

In addition, in this revision, we have included matched **pre- and post-treatment data from a patient-derived organoid line** demonstrating that *ex vivo* chemoradiotherapy can enrich the neural-like progenitor program (**new Fig. 3d**).

Again, and despite the efforts made, there is a need to be modest and prudent with the conclusions of the results presented and I highlight here some of the main general issues I have with the paper:
- a large number of comparisons are made in some analyses and even if the authors have placed emphasis on the improved statistical analysis, the possibility that the occurrence of significant p-values might be inflated due to the multiple comparisons made cannot be dismissed (e.g. several panels from Figure 2 and ED Suppl. Figs. 12,13);

We agree with the Reviewer that there is the possibility of obtaining false positives from multiple statistical tests. We have addressed this with Bonferroni corrections for multiple hypothesis testing (shown as adjusted p-values) in all analyses where we are comparing multiple treatment groups (**new Fig. 1f, 3b**; **new Extended Data Fig. 4, 12**) or multiple treatment response grades (**new Fig. 3c**).

- the study is based on an enlarged, but still small and very heterogeneous, set of samples which means that caution is mandatory. Furthermore, this is likely to be a biased series, even if unintentionally, as most retrospective series are. For example: while interesting, a group of 5 patients treated with CRTL, is a very tiny group....

We agree with the Reviewer that the CRTL cohort of 5 patients is not large in most contexts. However, our overall cohort size is large for a PDAC single-cell/nucleus transcriptomic study (and substantially larger than any previously published study), especially given the challenges of obtaining high-quality frozen primary resection specimens and the resources required to successfully process them. We nevertheless, note this limitation in the revised **Discussion**, as noted above.

Furthermore, our statistical approaches directly account for (**1**) patient sample size in each treatment group, and/or (**2**) patient-of-origin for each individual cell using a mixed effects model with patient ID as a random effect (that is, we do not treat cells from the same patient as independent observations). Hence, the CRTL-related findings in our study are significant even though there are only 5 patients, which suggests that the patterns and associations observed are rather pronounced.

- many bioinformatics parameters and technical limitations derived from the use of snRNA-Seq impact on the robustness of the cell types called and the new classifications. What is really important is whether the

latter can be confirmed in independent studies and they really exist in the patient (not that I am questioning it....);

We agree with the Reviewer that there are limitations to annotations and analyses performed on dissociated single cells or extracted single nuclei, and note this in the revised **Discussion** on page 18. Moreover, we conducted **additional experiments** to **validate the existence and potential clinical relevance of the neuronal-like program** (now termed neural-like progenitor program; please see our Overview and response to the Reviewer's Question #7). Specifically, we performed **multiplexed immunofluorescence** to demonstrate that neurexin-3, a protein expressed primarily in neuronal and glial cells in the brain, is heterogeneously expressed in a subset of malignant epithelial cells as determined by pan-cytokeratin co-staining. Furthermore, we had previously **derived an organoid line from one of our untreated study cohort patients** (PDAC_U_12), and now **treated** this organoid line with *ex vivo* FOLFIRINOX-like chemotherapy and radiotherapy followed by snRNA-seq on pre- and post-treatment samples. Our analyses revealed that **the neural-like progenitor program was indeed upregulated in cells from treated vs. untreated organoids** (**new Fig. 3d**). Taken together, these additional findings validate that malignant PDAC cells express neural-like genes and that the neural-like progenitor program is enriched post-treatment in a matched and controlled *ex vivo* setting.

- *this caution should be applied to the title as this is one of the aspects of the paper that will have wide visibility.*

We agree with the Reviewer on this important point. With our additional validation experiments for the neural-like progenitor program, we believe the title now accurately represents the content of the manuscript.

Specific comments

1. Fig. 1e-g are often referred to in the text concerning data that I do not see shown in the corresponding panel.

We thank the Reviewer for pointing this out. In the revision, we have rearranged the figure panels, such that prior Fig. 1e is **new Fig. 1g**; prior Fig 1f is **new Fig. 1h**; and prior Fig. 1g is **new Fig. 1e**. We have also carefully checked the revised text citations to these figure panels.

2. Lines 174-179: what is the possible impact of these biases? and their cause?

We thank the Reviewer for these important questions and now address them in revised **Discussion** (page 17). Specifically, differential capture of various immune cell subsets may limit immune cell enumeration analyses based on snRNA-seq. This can be due to differences in isolation efficiency of nuclei from different immune cell types, differences in cell size (smaller cells have smaller nuclei and less RNA, which can lead to quantities below QC thresholds; Habib *et al*. *Nature Methods* 2017, and some immune cells, like naïve T cells, are especially small; Wallrap *et al*., *Nature* 2017), or differences in structural integrity/viability of nuclei from different cell types. Focused examination of particular immune subsets may therefore benefit from application-specific optimization and complementary spatial (*in situ*) approaches. We note however, that our study is one of very few that even tested consistency with *in situ* measurements, and our consistency with *in situ* data is high overall, and better than that of scRNA-seq (which often enriches for immune cells because of poor capture of epithelial and/or stromal cells).

3. Lines 206-208: the markers used here to highlight malignant cells are inadequate: KRT19 is expressed in all ductal cells (normal and tumoral) and KRT17 is expressed in a selected subset of non-neoplastic ductal cells, including the pancreas of organ donors (PMID 7686885).

We thank the Reviewer for this helpful feedback. On lines 206-208 of the prior manuscript, we had listed a few example marker genes but did not provide a complete list. Following the Reviewer's comment, we have now added several other markers that were used in annotating the malignant cell population in the (**Methods,** page 27): KRT6A, KRT7, KRT14, KRT17, KRT19, TACSTD2, S100A11, S100A16, TFF1, and CLDN18 (**Fig. R1**, bottom). Many of these markers were also previously reported as enriched in malignant relative to normal epithelia [Lu et al, Onco Targets Ther 2021]. In our data, they are very highly expressed in malignant cells (**Fig. R1**), a few are moderately or lowly expressed in ductal (atypical) and ADM cells and are not expressed or extremely lowly expressed in normal cells. It is highly unlikely that

non-malignant cells express all these markers very highly. In addition, we further used the absence of expression of specific ductal markers (*e.g.*, CFTR) to support the malignant cell annotation.

4. Lines 227-230 and several other paragraphs: it is not clear that the authors have adjusted the p-values of comparisons in figure 2 - among others - for multiple testing, as would be required when a priori hypotheses are not considered.

We regret our lack of clarity. As described above, we have addressed this with Bonferroni corrections (adjusted p-values) in all analyses where we are comparing multiple treatment groups (**new Fig. 1f, 3b**; **new Extended Data Fig. 4, 12**) or multiple treatment response grades (**new Fig. 3c**).

5. Ext. Fig. 8a: it is striking that the signatures from the Chang-Seng-Yue paper are so poorly represented in this series' data. Any explanation?

We thank the Reviewer for pointing this out and have sought to address this question. Examining our analysis following the Reviewer's question, we realized that the visualization of the expression of the Chang-Seng-Yue signatures across malignant cells in our prior UMAPs was obscured in part by the presence of outlier cells that expressed the program very strongly (and dominated the saturation of the color scale). We now provide new UMAPs with a Vmax threshold of 0.99 such that all cells at the 99th percentile of expression and above are pseudocolored the same. The updated UMAPs are included in **new Extended Data Fig. 8**. For reference we show below both the previous and new visualization; the only distinction is the saturation of the color (**Fig. R6**).

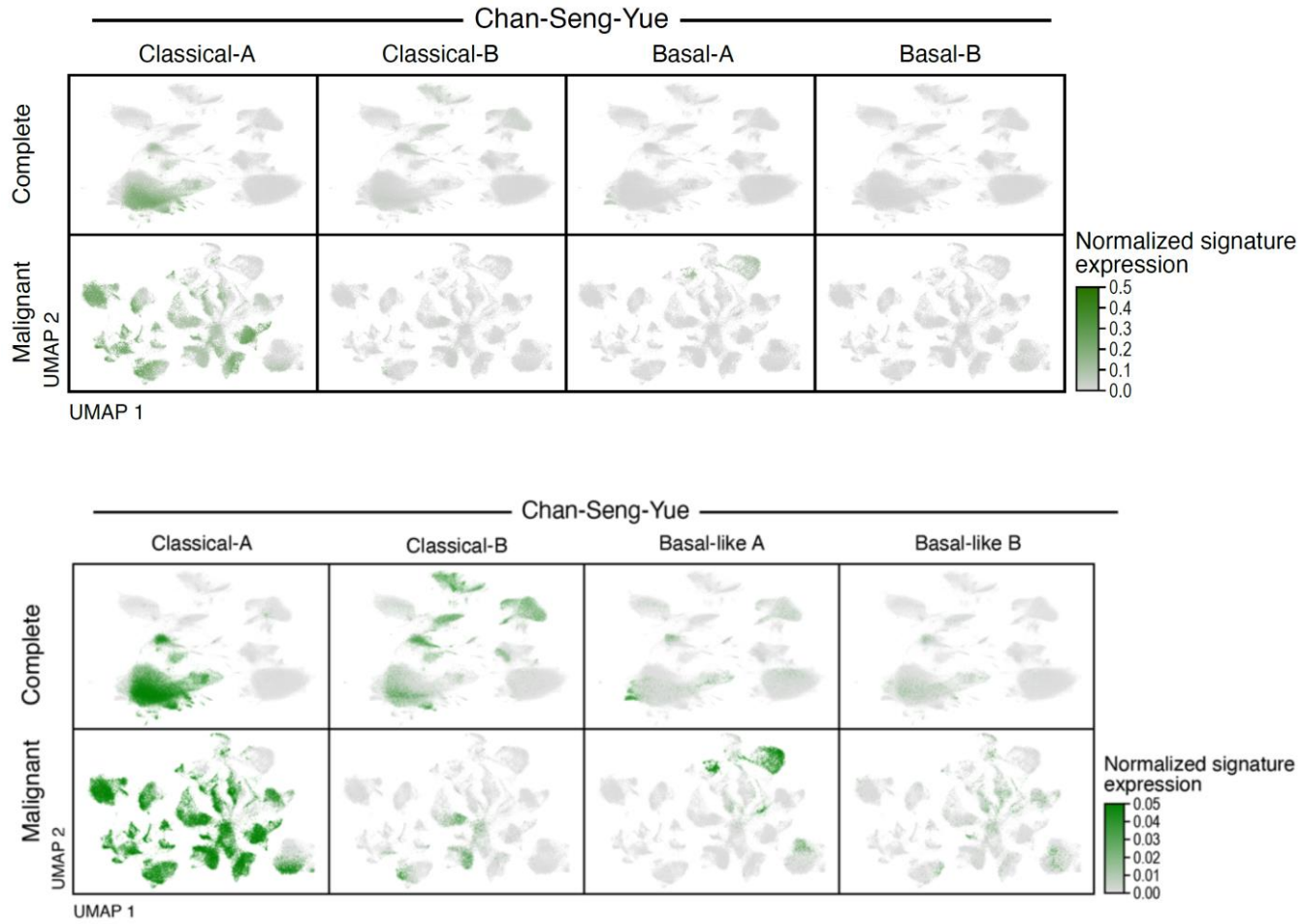

Fig. R6. UMAP visualization of the Chan-Seng-Yue et al signature mapped onto our snRNA-seq data at with two different saturation levels.

6. Lines 265 and subsequent: not clear how the adequacy of the number of programs chosen was assessed, both for tumor cells and for fibroblasts.

We thank the Reviewer for this important question. We performed consensus NMF and selected a number of components/programs that **optimized for the stability of the computed solution** (**Fig. R2** and **new Extended Data Fig. 9a**), leading to 49 components/programs for malignant cells, and 4 for fibroblasts. Among malignant cell programs, many, as expected, are patient-specific, and we therefore focused on the 14 that were biologically distinct and shared across cells from multiple patients (**Fig. R3**). While prior studies had identified fewer malignant subtypes, they had relied on bulk profiles that enable detecting majority cell states distinct among subsets of patients, while our single cell resolution can not only find such majority states, but also identify *minority* cell states shared across multiple patients, which are

otherwise obscured at the bulk level. The fewer number of programs/components for fibroblasts suggests that there is intrinsically less diversity within this cell type relative to malignant cells, which is not unexpected.

7. Designation of programs/cell types: I think that the authors need to be cautious here again. For example, their list of the 200 top genes from ED Table 2 for the neuronal subtype, when run on GO pathways, top annotations are "central nervous system" or "brain development" with modest adjusted p-values; however, when run on KEGG pathways, the top annotations include "bile secretion" and "pancreatic secretion" with similar adjusted p-values. I have not run analyses for many other types of questions in the paper but similar issues might come up. Therefore, I would be careful because once these "statements" are made in a published paper in an excellent journal, it becomes "dogma" while the knowledge is not necessarily more solid.

We agree with the Reviewer that it is important to be prudent about the nomenclature of our gene expression programs. We therefore performed **new analyses and experiments** in this revision to validate and hone the annotation of the neuronal-like program.

First, in our prior analyses, our top 200 cNMF-weighted genes corresponding to the neuronal-like program overlapped with multiple distinct Gene Ontology (GO) gene sets from biological processes for neuronal development and differentiation, many of which with FDR q-values below $1x10^{-15}$. In our **revised manuscript**, we also describe a disproportionate overlap ($p < 0.001$) between the top 200 genes of the neuronal-like program and a list of genes defined by the **Human Protein Atlas as being 'brain-enhanced'** (*i.e.*, having at least 4-fold greater expression in the brain relative to the average of all other tissues). Furthermore, we have now added a **new multiplexed immunofluorescence experiment** to demonstrate that neurexin-3, a protein expressed primarily in neuronal and glial cells in the brain, indeed colocalizes with a subset of epithelial cells as determined by pan-cytokeratin.

In addition, following the Reviewer's comment, we examined closely <u>other</u> program genes not captured by the neuronal enrichment but present in the same program. From our analyses, we also observed overlap with an adult tissue stem cell module, a gene ontology set for organ morphogenesis, and multiple gene

sets corresponding to hepatocyte nuclear factor (HNF) activity. As such, we have renamed our neuronal-like program to **'neural-like progenitor'** in the revised manuscript to incorporate these nuances.

We thank the Reviewer for this important question which helped us strengthen and refine a key conclusion.

8. Lines 401-403: a standard definition of response to neoadjuvant therapy should be used rather than the "poor, minimal, or moderate" grading which is not clinically acceptable.

The neoadjuvant therapy response grading used in this study is based on **our institutional clinical standard** for reporting surgical pathology in neoadjuvant-treated PDAC. A de-identified example is included below (**Fig. R7**). These terms were developed by board-certified pathologists and are reported in a standardized format in patients' surgical pathology reports. A score of 1 is "moderate response"; 2 is "minimal response"; and 3 is "poor response." We elected to use the descriptive terms rather than the numerical scores in the manuscript because the former are more informative for a reader unfamiliar with this grading system. We clarify this in the revised **Methods** on page 21.

Fig. R7. [REDACTED]

9. Lines 412-419 and following: the use of the PanCuRx dataset to analyze signatures derived from non-tumor cells is not appropriate since this series comprises laser-microdissected samples, generally with >70% tumor cells. The survival analyses are deeply limited by treatment heterogeneity and sample size. In fact, the only clearly significant association with outcome is observed for the classical subtype (Fig. 2e), a well-established association in the literature.

We thank the Reviewer for this question, but respectfully disagree on this point. We do not think the PanCuRx dataset is inappropriate for survival analyses of our programs. Our signatures were derived from "pure" malignant and fibroblast populations based on snRNA-seq (and thus, are not subject to issues that plagued bulk RNA-seq from large specimens). Applying these signatures to another "pure" cell type population would have been optimal, but cohorts with this level of resolution simply do not exist at a scale necessary to perform robust survival analyses. The fact that bulk studies had performed laser microdissection to >70% malignant purity actually helps with the relevance of our malignant signatures over bulk RNA-seq of large, non-microdissected specimens. On the other hand, we understand why one could reason that the fibroblast programs are less applicable here, but with up to 30% stroma as the Reviewer pointed out, we feel that it is justifiable to keep the fibroblast programs in the multivariable Cox model.

Furthermore, it is unclear to us what the Reviewer meant by 'the only clearly significant association with outcome is observed for the classical subtype', as **multiple other programs are significantly associated with time to progression (TTP)** using a standard p-value threshold of 0.05, including the neural-like progenitor program, which is concordant with similar prognostic associations observed with the neuronal subtype in muscle-invasive bladder cancer [Robertson et al., Cell 2017]. While some of the program associations observed for TTP do not reach the significance threshold for overall survival, they nonetheless trend in the same direction. Furthermore, TTP is a more specific measure of clinical outcomes attributable to disease characteristics (*e.g.*, gene expression phenotypes) relative to overall survival, which is affected by non-malignant causes of death. Indeed, in the combined TCGA and PanCuRx cohorts, there were 154 progression events and 167 deaths recorded, indicating that at least 13 deaths were not disease-related.

Lastly, the statistics applied are inherently designed to account for sample size. With a larger sample size, *additional* significant associations that we are currently underpowered to detect may be revealed, but the current associations are already significant at the current sample size.

10. Lines 477-495: the limited contribution of the novel tumor cell classification to our understanding of PDAC heterogeneity is highlighted by the fact that two of the tree communities proposed actually correspond largely to previously described PDAC subytpes. The third one (treatment-enriched) is defined through the subtypes that were less robustly defined. These findings call, again, for a cautious revision of the conclusions of the paper.

We thank the Reviewer for this feedback. In the spatial community analysis (page 15), we simply report on **spatial co-localizations** among the malignant programs, CAF programs, and immune cell types without speculating on biological claims not supported by our snRNA-seq data and established literature. Interestingly, when we performed a similar analysis at the level of patient tumors (rather than ROIs) using the snRNA-seq data, we saw a similar correlation among the neural-like progenitor malignant program, neurotropic CAF program, and $CD8^+$ T cells (**new Extended Data Fig. 17**). We have further revised the text to clarify this point.

11. In my opinion, much of the content of the last section of the Results should be presented differently, avoiding excessive speculation which is not based on mechanistic experiments and which could be summarized in a Table, either in the main text or as supplementary material.

We thank the Reviewer for this helpful suggestion. We agree that summarizing the RL analysis in a table is an efficient way of presenting the data. In the revised manuscript, we have removed the separate section on RL analysis, and added **new Table 2** to highlight select interactions that are differentially correlated between treatment groups and have been explored more extensively in prior mechanistic work.

Minor comments

1. Lines 91-94: I don't think that the issue here is whether these "normal genes" have a role in cancer, this should be rephrased.

We thank the Reviewer for this feedback. We have removed this from the revised manuscript.

2. Lines 139-140: specify here what about the 7 "missing" cases.

We appreciate the Reviewer bringing to our attention to this potential source of confusion. As laid out in **Fig. 1a**, some patient tumors were only analyzed by DSP and not snRNA-seq, and two additional samples in the study were non-malignant pancreatic tissue and therefore not included in this description of snRNA-seq on flash frozen, histologically-confirmed primary PDAC specimens (lines 130-142). We clarify this in the revised legend.

We did find a separate typographical error that we have corrected in **new Fig. 1a**. The total number of specimens in the revised manuscript is 48 rather than 50. Two additional patients samples (one CRT, one other treatment) were analyzed by DSP only with the more limited cancer transcriptome assay (~1,800 plex) and were omitted from the final analyses/manuscript.

3. The font of some of the panels in the figures should be carefully revised to make sure that the words can be read easily.

We thank the Reviewer for this feedback and have gone through the panels in the figures to ensure that they are more legible.

Response to Reviewer 4

Hwang et al. present a substantially revised manuscript describing their snRNA-seq methods and how they were able to derive clinically meaningful molecular signatures for PDAC – many for the first time. The heavily revised manuscript is much more streamlined with more simplified PDAC molecular signature hazard analysis and additional, more granular molecular and spatial proteomic analysis. All of this is underpinned by an almost doubling of the data (patient samples) included in the original manuscript. This is a much-improved study compared to the original submission and should serve as standard bearer for multi-omic clinical studies and data integration of complementary features from different analyses.

We appreciate the Reviewer's kind feedback on our revised manuscript and their comments that helped us improve our study and manuscript.

One minor critique that could possibly improve the manuscript further would be to include the hazard analysis for one or two previously published PDAC molecular signature into the revised 2E. Even though it is clear the new signatures identified here have significant and complementary clinical implications it would still be good for readers to understand how this is couched within previous knowledge by having a few points of comparison.

We thank the Reviewer for this suggestion. We had originally chosen not to include prior molecular signatures to avoid overfitting this multivariable analysis as with 17 covariates we were already at risk of overfitting the time to progression (TTP) and overall survival (OS) endpoints, which had 154 and 167 events, respectively.

Nevertheless, to address the Reviewer's point, we re-performed the analysis with the Collisson quasi-mesenchymal and classical signatures included and removed our malignant state categories (secretory, cycling) to keep the total number of covariates at 17 (**Fig. R8**). With this set of programs, **the neural-like progenitor (NRP) program (HR = 1.57, p = 0.028) and squamoid (SQM) program (HR = 1.35, p = 0.025) remained significantly associated with shorter TTP.** Consistent with prior data, the quasi-mesenchymal subtype was also associated with shorter TTP (HR = 1.52, p = 0.004). Interestingly, the inclusion of the largely redundant Collisson classical program may have abrogated the significant

prognostic association previously observed with our classical malignant program, as neither were significantly associated with longer TTP. These results reinforce the notion that the NRP program is prognostically relevant.

_t	Haz. Ratio	Std. Err.	z	P>|z|	[95% Conf.	Interval]
age	1.011021	.0083931	1.32	0.187	.9947042	1.027606
sex	.8342879	.1401195	-1.08	0.281	.6002827	1.159514
stage	1.178104	.2251607	0.86	0.391	.8100292	1.71343
grade	1.122607	.1654864	0.78	0.433	.840911	1.498669
NEN_z	.9272562	.1281434	-0.55	0.585	.70724	1.215718
MES_z	1.09102	.301824	0.31	0.753	.6343852	1.876344
CLS_z	.8224194	.2046459	-0.79	0.432	.5049931	1.339372
ACN_z	1.046254	.1273853	0.37	0.710	.8241386	1.328232
NRP_z	1.56728	.3208647	2.19	0.028	1.049259	2.34105
BSL_z	.7829331	.104037	-1.84	0.066	.603415	1.015859
SQM_z	1.346724	.1789988	2.24	0.025	1.037867	1.747493
COLC_z	.9470971	.1968023	-0.26	0.794	.6302587	1.423214
COLQ_z	1.52298	.2242989	2.86	0.004	1.141123	2.03262
IMM_z	.5690949	.1194695	-2.69	0.007	.3771315	.8587693
NRT_z	1.00182	.2677616	0.01	0.995	.5933138	1.691588
ADH_F_z	1.584907	.4446494	1.64	0.101	.9145285	2.746695
MYO_z	.8143194	.2643915	-0.63	0.527	.4309516	1.538725

Fig. R8. Multi-variable Cox regression analysis for TTP with the Collisson quasi-mesenchymal and classical signatures included and secretory and cycling malignant states removed.

Also minor, Figure 2 would be even clearer if the variable abbreviations from 2E were included in the plot titles of 2B.

We thank the Reviewer for this helpful suggestion and now include the three-letter program abbreviations used in **new Fig. 3e** in **new Fig. 2a** as well.

Decision Letter, first revision:

Our ref: NG-A58304R

12th Jan 2022

Dear Will,

Thank you for submitting your revised manuscript "Single-nucleus and spatial whole transcriptome profiling of pancreatic cancer reveals multicellular communities and enrichment of a neural-like progenitor phenotype after neoadjuvant treatment" (NG-A58304R). It has now been seen by 3 of 4 original referees and their comments are below. The reviewers find that the paper has improved in revision - Reviewer #4 has not submitted a report, but their comments at the last round were minor and we think they have been satisfactorily addressed - 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.

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,

Michael Fletcher, PhD
Associate Editor, Nature Genetics

ORCID: 0000-0003-1589-7087

Reviewer #1 (Remarks to the Author):

The authors submit a revised manuscript about single cell RNA seq in pancreatic cancer.
The major findings of the paper are multiple subtypes of cancer cells and a move neuronal like subtype. Additional experiments have been added to show Neurexin-3 containing in PDAC cancer cells. They have also addressed clarity issues I encountered with the prior presentation of the data.

Reviewer #2 (Remarks to the Author):

In this revision the authors have included important validations of the findings. In particular, the novel neuronal-like program is better evidenced according to its features and annotation, together with its treatment-induced enrichment. The authors have also clarified the text and figures. I believe that the manuscript is now ready for publication.

Reviewer #3 (Remarks to the Author):

The authors have improved the manuscript and added new data that are valuable even if preliminary (e.g. the organoids). Some of the critical issues remain (e.g. small sample size).

Author Rebuttal, first revision:

Reviewer #1:
Remarks to the Author:
The authors submit a revised manuscript about single cell RNA seq in pancreatic cancer. The major findings of the paper are multiple subtypes of cancer cells and a move neuronal like subtype. Additional experiments have been added to show Neurexin-3 containing in PDAC cancer cells. They have also addressed clarity issues I encountered with the prior presentation of the data.

We thank the Reviewer for this kind feedback.

Reviewer #2:
Remarks to the Author:
In this revision the authors have included important validations of the findings. In particular, the novel neuronal-like program is better evidenced according to its features and annotation, together with its treatment-induced enrichment. The authors have also clarified the text and figures. I believe that the manuscript is now ready for publication.

We thank the Reviewer for this kind feedback.

Reviewer #3:
Remarks to the Author:
The authors have improved the manuscript and added new data that are valuable even if preliminary (e.g. the organoids). Some of the critical issues remain (e.g. small sample size).

We agree with the Reviewer that larger cohorts and further validation are warranted in future work.

Reviewer #4:
None

We thank the Reviewer for their time.

Final Decision Letter:

In reply please quote: NG-A58304R1 Regev

16th Jun 2022

Dear Dr. Regev,

I am delighted to say that your manuscript "Single-nucleus and spatial transcriptome profiling of pancreatic cancer identifies multicellular dynamics associated with neoadjuvant treatment" has been accepted for publication in an upcoming issue of Nature Genetics.

Over the next few weeks, your paper will be copyedited to ensure that it conforms to Nature Genetics style. Once your paper is typeset, you will receive an email with a link to choose the appropriate publishing options for your paper and our Author Services team will be in touch regarding any additional information that may be required.

After the grant of rights is completed, you will receive a link to your electronic proof via email with a request to make any corrections within 48 hours. If, when you receive your proof, you cannot meet this deadline, please inform us at rjsproduction@springernature.com immediately.

You will not receive your proofs until the publishing agreement has been received through our system.

Due to the importance of these deadlines, we ask that you please let us know now whether you will be difficult to contact over the next month. If this is the case, we ask you provide us with the contact information (email, phone and fax) of someone who will be able to check the proofs on your behalf, and who will be available to address any last-minute problems.

Your paper will be published online after we receive your corrections and will appear in print in the next available issue. You can find out your date of online publication by contacting the Nature Press Office (press@nature.com) after sending your e-proof corrections. Now is the time to inform your Public Relations or Press Office about your paper, as they might be interested in promoting its publication. This will allow them time to prepare an accurate and satisfactory press release. Include your manuscript tracking number (NG-A58304R1) and the name of the journal, which they will need when they contact our Press Office.

Before your paper is published online, we shall be distributing a press release to news organizations worldwide, which may very well include details of your work. We are happy for your institution or funding agency to prepare its own press release, but it must mention the embargo date and Nature Genetics. Our Press Office may contact you closer to the time of publication, but if you or your Press Office have any enquiries in the meantime, please contact press@nature.com.

Acceptance is conditional on the data in the manuscript not being published elsewhere, or announced in the print or electronic media, until the embargo/publication date. These restrictions are not intended to deter you from presenting your data at academic meetings and conferences, but any enquiries from the media about papers not yet scheduled for publication should be referred to us.

Please note that Nature Genetics is a Transformative Journal (TJ). Authors may publish their research with us through the traditional subscription access route or make their paper immediately open access through payment of an article-processing charge (APC). Authors will not be required to make a final decision about access to their article until it has been accepted. Find out more about Transformative Journals

Authors may need to take specific actions to achieve <a href="https://www.springernature.com/gp/open-research/funding/policy-compliance-

faqs"> compliance with funder and institutional open access mandates. If your research is supported by a funder that requires immediate open access (e.g. according to Plan S principles) then you should select the gold OA route, and we will direct you to the compliant route where possible. For authors selecting the subscription publication route, the journal's standard licensing terms will need to be accepted, including <a href="https://www.nature.com/nature-portfolio/editorial-policies/self-archiving-and-license-to-publish. Those licensing terms will supersede any other terms that the author or any third party may assert apply to any version of the manuscript.

Please note that Nature Portfolio offers an immediate open access option only for papers that were first submitted after 1 January, 2021.

If you have any questions about our publishing options, costs, Open Access requirements, or our legal forms, please contact ASJournals@springernature.com

If you have posted a preprint on any preprint server, please ensure that the preprint details are updated with a publication reference, including the DOI and a URL to the published version of the article on the journal website.

To assist our authors in disseminating their research to the broader community, our SharedIt initiative provides you with a unique shareable link that will allow anyone (with or without a subscription) to read the published article. Recipients of the link with a subscription will also be able to download and print the PDF.

As soon as your article is published, you will receive an automated email with your shareable link.

You can now use a single sign-on for all your accounts, view the status of all your manuscript submissions and reviews, access usage statistics for your published articles and download a record of your refereeing activity for the Nature journals.

An online order form for reprints of your paper is available at https://www.nature.com/reprints/author-reprints.html. Please let your coauthors and your institutions' public affairs office know that they are also welcome to order reprints by this method.

If you have not already done so, we invite you to upload the step-by-step protocols used in this manuscript to the Protocols Exchange, part of our on-line web resource, natureprotocols.com. If you complete the upload by the time you receive your manuscript proofs, we can insert links in your article that lead directly to the protocol details. Your protocol will be made freely available upon publication of your paper. By participating in natureprotocols.com, you are enabling researchers to more readily reproduce or adapt the methodology you use. Natureprotocols.com is fully searchable, providing your protocols and paper with increased utility and visibility. Please submit your protocol to https://protocolexchange.researchsquare.com/. After entering your nature.com username and password you will need to enter your manuscript number (NG-A58304R1). Further information can be found at https://www.nature.com/nature-portfolio/editorial-policies/reporting-standards#protocols

Sincerely,

Michael Fletcher, PhD
Senior Editor, Nature Genetics

ORCiD: 0000-0003-1589-7087