nature research

Corresponding author(s): Vinod Balachandran

☐ Initial submission ☒ Revised version ☐ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

Experimental design

1. Sample size

Describe how sample size was determined.

We were unable to calculate pre-specified effect sizes to optimally determine sample sizes in the two groups, given the extreme rarity of published literature examining tumor specimens in long term survivors of pancreatic cancer. The largest study (prior to ours) to perform whole exome sequencing on long term survivors had a sample size of 8 patients (PMID: 25623214). Following our initial analysis in the MSKCC cohort (n=58), we extended our findings to one of the largest cohorts of human pancreatic cancer with available next generation sequencing, and that is unselected by survival (n=166).

2. Data exclusions

Describe any data exclusions.

MSKCC PDAC Cohort: All tumor samples were surgically resected primary pancreatic ductal adenocarcinomas. Patients treated with neoadjuvant therapy were excluded. All tumors were subjected to pathologic re-review and histologic confirmation by two expert PDAC pathologists prior to analyses. Long term survivors were defined as patients with overall survival > 3 years from surgery, short term survivors as patients with survival >3 months and < 1 year from surgery to exclude perioperative mortalities. (Reported on page # 8, paragraph # 1, line # 288-292)

3. Replication

Describe whether the experimental findings were reliably reproduced.

All experimental findings reported in this manuscript were reliably reproducible.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

MSKCC PDAC Cohort: Long term survivors were defined as patients with overall survival > 3 years from surgery, short term survivors as patients with survival >3 months and < 1 year from surgery to exclude perioperative mortalities. (Reported on page # 8 paragraph # 1, line # 285-292.

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

Tissue Microarray Multiplexed Immunohistochemistry and Immunofluorescence: Automated digital quantification of tissue microarrays was performed blinded to the investigator with respect to the short and long term cohorts.

Whole Exome Sequencing and Neoantigen Prediction: Computational mutation calling and neoantigen prediction was performed blinded to the investigator with respect to the short and long term cohorts.

Transcriptome analysis: Bulk tumor transcriptome analysis was performed blinded to the investigator with respect to the short and long term cohorts.

T Cell Receptor Sequencing: TCR Sequencing was performed blinded to the investigator with respect to the short and long term cohorts.

Neoantigen fitness modeling: Neoantigen fitness modeling was performed blinded

to the investigator with respect to the short and long term cohorts.

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.	For statistical measurements in Figure 3D, the glm function in R version 3.4 was used for model fitting these models. All data analysis was performed using Prism 7.0, GraphPad Software unless otherwise indicated. Cox regression was performed using STATA 13.1. Mutations were called using Mutect (https://software.broadinstitute.org/gatk/download/mutect), Mutect Rescue (https://github.com/soccin/BIC-variants_pipeline/tree/master/rescue), and HaplotypeCaller (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php_. Neoantigens were called using NASeek (reference # 9), and pVAC_Seq (https://github.com/griffithlab/pVAC-Seq). MHC amplitude calculations were performed using NetMHC 3.4 (http://www.cbs.dtu.dk/services/NetMHC-3.4). TCR recognition probability was calculated using blastp Biopairwise 2 (https://github.com/biopython/biopython.github.io/). Tumor clonal structure was calculated using PhyloWGS. (https://github.com/morrislab/phylowgs)

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

There are no restrictions on availability of unique materials and there are no materials available for distribution by a for-profit company.

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

Immunohistochemistry:
Human specific antibodies to MUC16 (clone OCT125, dilution 1:130), WT1 (clone CAN-R9 (IHC)-56-2, dilution 1:30), and Annexin A2 (ab54771, 5 ug/ml) were purchased from Abcam (MA, USA). Antibodies to MUC1 (clone M695, dilution 1:100), and Mesothelin (clone 5B2, dilution 1:50) were purchased from Vector laboratories (CA, USA).

Multiplexed Immunohistochemistry:
Antibodies to Granzyme B (clone GrB-7, Dako), CD3 (clone 2GV6, Ventana), CD8, (clone C8/144b, Dako), FoxP3 (clone 236A/E7, Abcam), CD20 (clone L26, Dako), CD68 (clone KP1, Dako), DC-LAMP (clone 1010E1.01, Novus Biologicals), MHC class-I (HLA-ABC, clone EMR8-5, Abcam) and CK19 (clone EP1580Y, Abcam) were used.

Immunofluorescence:
Antibodies to CD4 (Ventana, cat#790-4423, 0.5ug/ml), FoxP3 (Abcam, cat#ab20034, 5 ug/ml), mouse IgG (Vector Labs, cat#MKB-22258), CK19 (Abcam, cat#ab52625, 1ug/ml), and DAPI (Sigma Aldrich, cat#D9542, 5 ug/ml) were used.

Flow Cytometry: Human-specific antibodies used in all flow cytometric phenotyping included CD45 (clone HI30, BioLegend), CD3 (clone OKT3, BioLegend), CD4 (clone SK3, BD Biosciences), CD8 (clone SK1, BioLegend), CD56 (clone B159, BD Biosciences), CD69 (clone FN50, BD Biosciences), CD19 (clone SJ25C1, BD Biosciences), PD1 (clone MIH4, BD Biosciences), CD45RA (clone HI100, BD Biosciences), CD45 RO (clone UCHL1, BD Biosciences), CD56 (clone B159, BD Biosciences) and CD107a (clone H4A3, BD Biosciences).

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

HEK293T cells were obtained from the ATCC Human Primary Cell collection (https://www.atcc.org).

b. Describe the method of cell line authentication used.

Cell line authentication was performed using Short Tandem Repeat (STR) analysis.

c. Report whether the cell lines were tested for mycoplasma contamination.

HEK293T cells were tested for mycoplasma contamination prior to use.

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

Not applicable

Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

Not applicable.

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

All clinicopathologic and population characteristics of all human research participants are provided in detail in Extended Data Tables 1-6. Briefly, all patients had surgically resected primary pancreatic ductal adenocarcinomas. Patients treated with neoadjuvant therapy were excluded. All tumors were subjected to pathologic re-review and histologic confirmation by two expert PDAC pathologists prior to analyses. Long term survivors were defined as patients with overall survival > 3 years from surgery, short term survivors as patients with survival >3m and < 1 year from surgery to exclude perioperative mortalities. Patients were not excluded based on age or gender. For patients included in the Rapid Autopsy cohort, primary and metastatic tumor samples were collected posthumously from four patients as part of the Gastrointestinal Cancer Rapid Medical Donation program. Informed consent was obtained from all subjects.

nature research

Corresponding author(s): Vinod Balachandran

☐ Initial submission ☒ Revised version ☐ Final submission

Flow Cytometry Reporting Summary

Form fields will expand as needed. Please do not leave fields blank.

Data presentation

For all flow cytometry data, confirm that:

☒ 1. The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ 2. The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ 3. All plots are contour plots with outliers or pseudocolor plots.

☒ 4. A numerical value for number of cells or percentage (with statistics) is provided.

Methodological details

5. Describe the sample preparation.

Fresh blood and tumor samples from patients undergoing elective surgery at Memorial Hospital were collected. Peripheral-blood mononuclear cells were isolated by density centrifugation of 25 ml of whole blood over 25 ml of Ficoll-Paque Plus (GE Healthcare).

Tumors were mechanically dissociated by mincing with scalpels, scissors, and forceps. Minced tumors were then enzymatically dissociated in a 2.5 ml solution containing collagenase IV (Sigma SKU# C5138, 5mg/ml), DNase I (Roche SKU# 10104159001, 1ul/ml), and a protease inhibitor cocktail mini tablet (Roche SKU# 11836170001) in HBSS, for 30 minutes in a shaker incubator at 37C. Enzymatic dissociation was quenched with 5 ml Fetal Bovine Serum (USA Scientific Ref# # 9874-5244, now discontinued). The cell suspension was then passed through a 100-micron strainer (Falcon cell strainer Ref# 352360) using the flat portion of a plunger from a 3-ml syringe, and filtered again through a 40-micron strainer (Falcon cell strainer Ref# 352340) prior to flow cytometric analysis.

Tumor draining lymph nodes were mechanically dissociated using the flat portion of a plunger from a 3-ml syringe through a 100-micron strainer (Falcon cell strainer Ref# 352360), and filtered again through a 40-micron strainer (Falcon cell strainer Ref# 352340) prior to flow cytometric analysis.

6. Identify the instrument used for data collection.

Flow cytometry was performed on an LSRFortessa (BD Biosciences; Catalog # 647177; Serial # H64717700135).

7. Describe the software used to collect and analyze the flow cytometry data.

Data were analyzed using FlowJo Software (Tree Star).

8. Describe the abundance of the relevant cell populations within post-sort fractions.

No cell sorting was performed in this report.

9. Describe the gating strategy used.

Relevant gating strategies used are indicated in Figures 3c, 4i, Extended Data Figure 2b, and Extended Data Figure 10.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information. ☒