Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer

Vinod P. Balachandran[1,2,3], Marta Łuksza[4], Julia N. Zhao[1,2,3], Vladimir Makarov[5,6], John Alec Moral[1,2,3], Romain Remark[7], Brian Herbst[2], Gokce Askan[2,8], Umesh Bhanot[8], Yasin Senbabaoglu[9], Daniel K. Wells[10], Charles Ian Ormsby Cary[10], Olivera Grbovic-Huezo[2], Marc Attiyeh[1,2], Benjamin Medina[1], Jennifer Zhang[1], Jennifer Loo[1], Joseph Saglimbeni[2], Mohsen Abu-Akeel[9], Roberta Zappasodi[9], Nadeem Riaz[6,11], Martin Smoragiewicz[12], Z. Larkin Kelley[13,14], Olca Basturk[8], Australian Pancreatic Cancer Genome Initiative*, Mithat Gönen[15], Arnold J. Levine[4], Peter J. Allen[1,2], Douglas T. Fearon[13,14], Miriam Merad[7], Sacha Gnjatic[7], Christine A. Iacobuzio-Donahue[2,5,8], Jedd D. Wolchok[3,9,16,17,18], Ronald P. DeMatteo[1,2], Timothy A. Chan[3,5,6,11], Benjamin D. Greenbaum[19], Taha Merghoub[3,9,18]§ & Steven D. Leach[1,2,5,20]§

Pancreatic ductal adenocarcinoma is a lethal cancer with fewer than 7% of patients surviving past 5 years. T-cell immunity has been linked to the exceptional outcome of the few long-term survivors[1,2], yet the relevant antigens remain unknown. Here we use genetic, immunohistochemical and transcriptional immunoprofiling, computational biophysics, and functional assays to identify T-cell antigens in long-term survivors of pancreatic cancer. Using whole-exome sequencing and *in silico* neoantigen prediction, we found that tumours with both the highest neoantigen number and the most abundant CD8$^+$ T-cell infiltrates, but neither alone, stratified patients with the longest survival. Investigating the specific neoantigen qualities promoting T-cell activation in long-term survivors, we discovered that these individuals were enriched in neoantigen qualities defined by a fitness model, and neoantigens in the tumour antigen MUC16 (also known as CA125). A neoantigen quality fitness model conferring greater immunogenicity to neoantigens with differential presentation and homology to infectious disease-derived peptides identified long-term survivors in two independent datasets, whereas a neoantigen quantity model ascribing greater immunogenicity to increasing neoantigen number alone did not. We detected intratumoural and lasting circulating T-cell reactivity to both high-quality and MUC16 neoantigens in long-term survivors of pancreatic cancer, including clones with specificity to both high-quality neoantigens and predicted cross-reactive microbial epitopes, consistent with neoantigen molecular mimicry. Notably, we observed selective loss of high-quality and *MUC16* neoantigenic clones on metastatic progression, suggesting neoantigen immunoediting. Our results identify neoantigens with unique qualities as T-cell targets in pancreatic ductal adenocarcinoma. More broadly, we identify neoantigen quality as a biomarker for immunogenic tumours that may guide the application of immunotherapies.

To define the importance of neoantigens in pancreatic ductal adenocarcinoma (PDAC), we compared stage-matched cohorts of treatment-naive, surgically resected, rare long-term survivors (LTSs, median survival 6 years, $n = 82$) to short-term survivors with a more typical poor outcome (median survival 0.8 years, $n = 68$; Memorial Sloan Kettering Cancer Center (MSKCC) cohort; Fig. 1a, Extended Data Fig. 1a–e). Using nine-parameter multiplexed immunohistochemistry[3] and immunofluorescence in tissue microarrays, we found greater densities of CD8$^+$ T cells (3-fold), cytolytic CD8$^+$ cells (12-fold), mature dendritic cells, regulatory T cells, macrophages, and decreased numbers of CD4$^+$ T cells in tumours of LTSs, yet no differences in

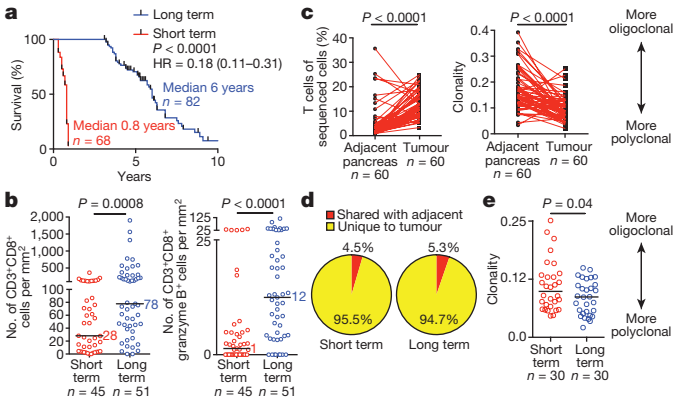


Figure 1 | Pancreatic cancer survivors display enhanced T-cell immunity. a, Survival of MSKCC cohort patients. HR, hazard ratio. **b**, Quantification of CD8$^+$ T cells using multiplexed immunohistochemistry. **c**, **d**, T-cell frequency and repertoire clonality (**c**) and clonal overlap (**d**) in tumour and matched adjacent normal pancreatic tissues, determined by TCR Vβ sequencing. **e**, Intratumoural T-cell repertoire clonality in tumours of short- and long-term survivors of PDAC. Horizontal bars indicate median values. *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test (**a**), a two-tailed Mann–Whitney *U*-test (**b**), a two-tailed paired (**c**) and a two-tailed unpaired Student's *t*-test (**e**).

[1]Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [2]David M. Rubenstein Center for Pancreatic Cancer Research, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [3]Parker Institute for Cancer Immunotherapy, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [4]The Simons Center for Systems Biology, Institute for Advanced Study, Princeton, New Jersey, USA. [5]Human Oncology and Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [6]Immunogenomics and Precision Oncology Platform, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [7]Tisch Cancer Institute, Immunology Institute, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [8]Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [9]Swim Across America/Ludwig Collaborative Laboratory, New York, New York, USA. [10]Parker Institute for Cancer Immunotherapy, San Francisco, California, USA. [11]Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [12]Cancer Research UK Cambridge Institute, University of Cambridge, Li Ka Shing Centre, Cambridge, UK. [13]Cold Spring Harbor Laboratory, New York, New York, USA. [14]Department of Microbiology and Immunology, Weill Cornell Medical School, New York, New York, USA. [15]Department of Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [16]Melanoma and Immunotherapeutics Service, Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [17]Weill Cornell Medical College, Cornell University, New York, New York, USA. [18]Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, New York, USA. [19]Tisch Cancer Institute, Departments of Medicine, Hematology and Medical Oncology, Oncological Sciences, and Pathology, Icahn School of Medicine at Mount Sinai, New York, New York, USA. [20]Dartmouth Norris Cotton Cancer Center, Lebanon, New Hampshire, USA.
*Lists of participants and their affiliations appear in the online version of this paper.
§These authors jointly supervised this work.

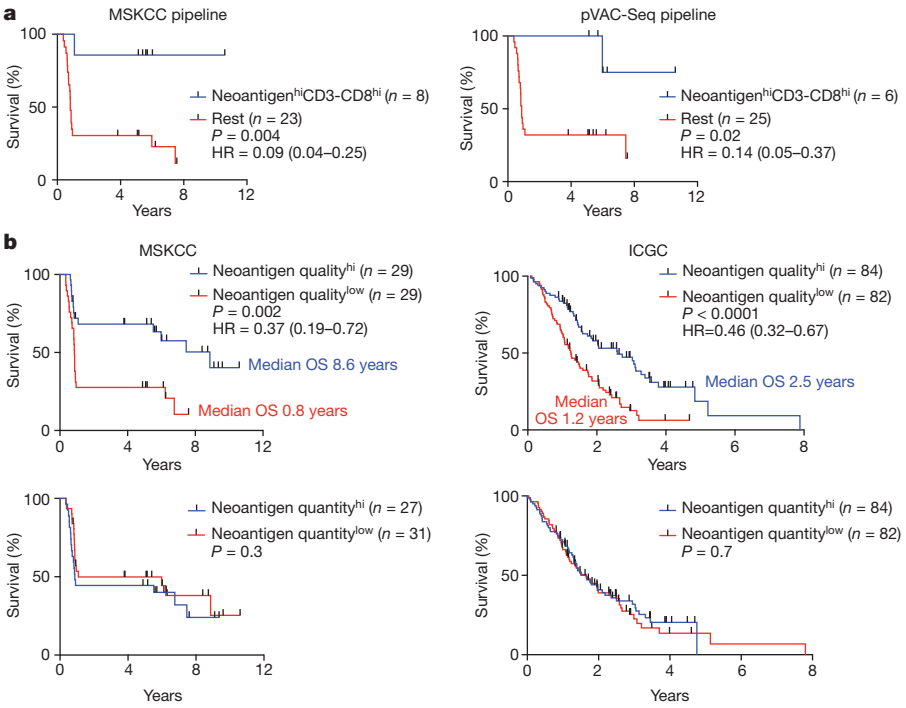


Figure 2 | Neoantigen quality is prognostic of survival. a, Survival of patients with tumours harbouring greater than the median number of both neoantigens (neoantigen^{hi}) and CD3$^+$CD8$^+$ cells (CD3-CD8^{hi}), compared to all other patients (rest), as determined by the MSKCC (left) and pVAC-Seq (right) neoantigen prediction pipelines. **b**, Patient survival stratified by neoantigen quality and quantity in the MSKCC (left) and ICGC (right) cohorts. OS, overall survival. *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test.

B cells and major histocompatibility complex (MHC) class I$^+$ cells (Fig. 1b, Extended Data Fig. 2a). Transcriptomic profiling revealed an immunogenic microenvironment in tumours of LTSs, with upregulation of molecular markers of dendritic cells[4], and antigen experience (PD-1 and TIGIT)[5], as well as downregulation of the immunosuppressive marker STAT3 (Extended Data Fig. 2b). Sequencing of the T-cell receptor (TCR) Vβ chain demonstrated that intratumoural T cells were increased fivefold compared to matched adjacent non-tumour pancreatic tissue and were markedly polyclonal (Fig. 1c). Notably, more than 94% of intratumoural T-cell clones were unique to tumours, consistent with tumour specificity (Fig. 1d). Additionally, in unselected patients, flow cytometry on intratumoural T cells revealed activation and memory marker upregulation compared to draining lymph node and blood T cells, consistent with antigen specificity (Extended Data Fig. 2c). Finally, tumours of LTSs exhibited greater TCR repertoire diversity (Fig. 1e). The association of activated CD8$^+$ T cells and survival was independent of clinicopathological factors and adjuvant chemotherapy (Extended Data Fig. 2d). Collectively, tumours of LTSs exhibited an activated, polyclonal, tumour-specific T-cell infiltrate, suggesting differential antigenic targets.

To determine the neoantigen frequency in PDACs, we performed whole-exome sequencing on macrodissected tumour islands. We detected a median of 38 predicted neoantigens per tumour[6] (Extended Data Fig. 3a, Supplementary Table 1). Notably, patients with both the highest predicted neoantigen number and either the greatest CD3$^+$CD8$^+$, or polyclonal T-cell repertoire, but neither alone, exhibited the longest survival (median survival not reached, Fig. 2a, Extended Data Fig. 3b). We corroborated these findings using a second neoantigen prediction algorithm[7] (Fig. 2a, Extended Data Fig. 3b, c). This association of higher neoantigen quantity and CD8$^+$ T-cell infiltrate with survival was independent of adjuvant chemotherapy (Extended Data Fig. 3d). Furthermore, we found that higher neoantigen quantity and CD8$^+$ T-cell infiltrate together exhibited the strongest association with survival, with other genomic and immune parameters showing weak or no association with survival[8] (Extended Data Fig. 4a–d). Together, these data suggest that neoantigen immunogenicity and quality, and not purely quantity, correlate with survival.

Next, we investigated neoantigen qualities that modulate differential immunogenicity. The theory of molecular mimicry postulates that TCRs that can recognize pathogenic antigens can also recognize non-pathogenic antigens, which has been documented in autoimmunity[9] but not in the cancer context. We theorized that neoantigen homology to infectious disease-derived epitopes, which are recognized by the human TCR repertoire, can serve as a partial surrogate for differential neoantigen immunogenicity or 'non-selfness'. This hypothesis does not assume any associations between pre-existing antimicrobial immunity and survival, but rather aims to develop a strategy to identify candidate neoantigens based on defined immunogenic pathogen-derived epitopes. To test this hypothesis, we developed neoantigen quality and quantity models (Extended Data Fig. 5a, b). For each neoantigen in the quality model, we derived sequence alignment scores to human infectious disease-derived, class I-restricted peptide sequences with positive immune assays (hereby referred to as 'microbial') from the Immune Epitope Database (IEDB, Supplementary Table 2). We then inferred its probability of TCR recognition using a nonlinear logistic dependence on alignment score, and calculated quality scores by amplifying these binding probabilities by inferred relative wild-type and mutant peptide-MHC class I affinities (Extended Data Fig. 5a, b). For the quantity model, the neoantigen score equalled the total number of neoantigens within a clone. Using an evolutionary model, we interpreted the score of a clone as calculated by each model as its fitness cost due to immune interactions[10]. We then recreated the clonal tree structure for each tumour based on mutant allele frequencies, and projected the change to the effective cancer cell population size of each tumour. Of these models, the quality model, but not the quantity model, significantly stratified short- and long-term survivors independent of confounding factors and adjuvant chemotherapy (Fig. 2b, Extended Data Fig. 6a). Notably, all tumours with the highest neoantigen load in combination with the most abundant CD8$^+$ T-cell infiltrates harboured high-quality neoantigens (Extended Data Fig. 3e). Testing these models

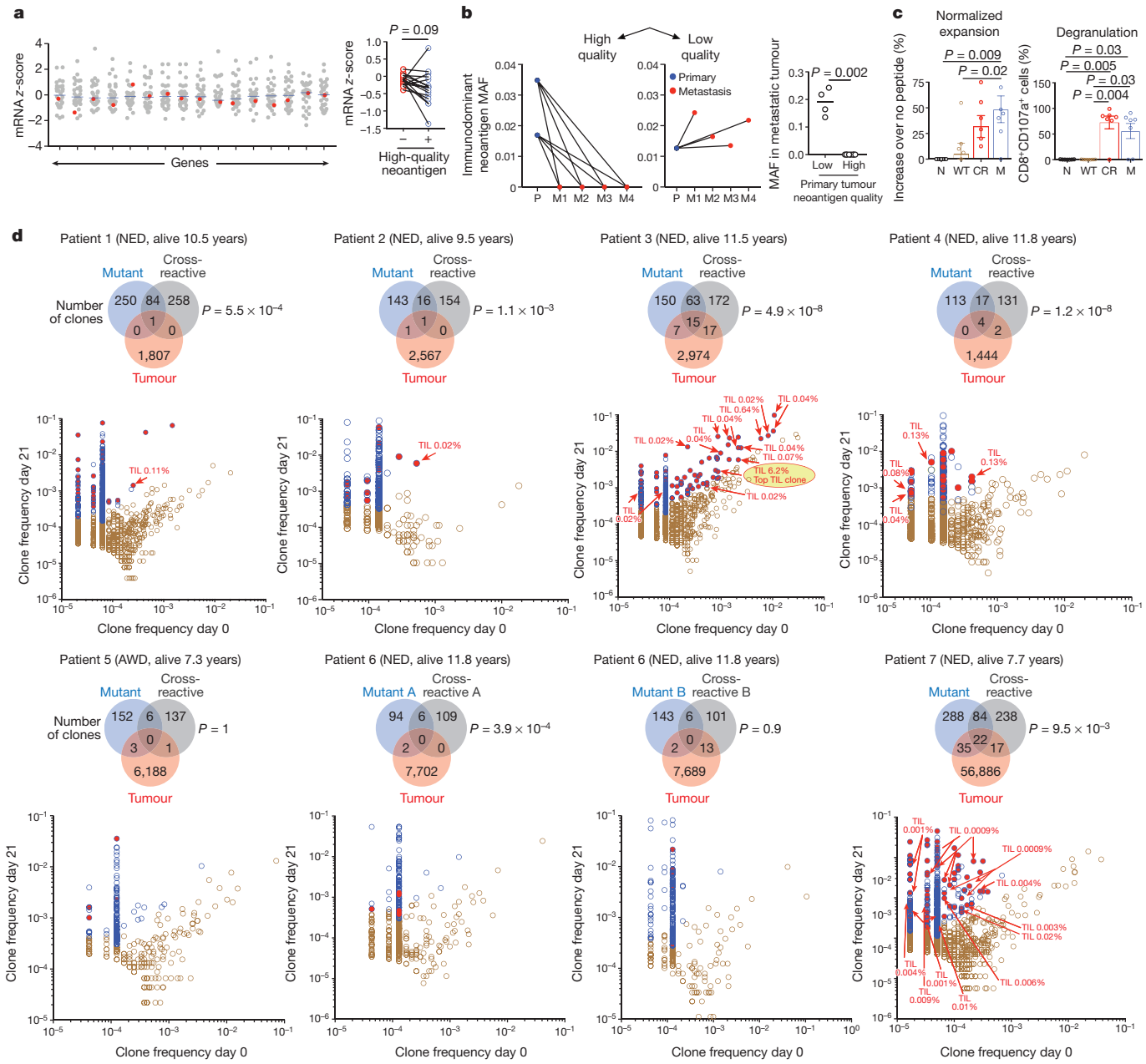


Figure 3 | Neoantigen and cross-reactive microbial peptide T cells detected in blood and tumours. a, Left, gene expression in the presence (red) or absence (grey) of high-quality neoantigenic mutations. *x* axis, genes; shaded circles, biologically independent samples in individual patients ($n = 30$). Right, median non-neoantigenic and neoantigenic expression. All high-quality neoantigenic genes with available mRNA expression are shown. **b**, Metastatic propagation of all clones in the primary tumour stratified by neoantigen quality. Mutant allele frequencies in matched primary–metastatic tumours (left) and metastatic tumours alone (right) are shown in biologically independent samples in one patient. MAF, mutant allele frequency; M, metastasis; P, primary tumour. **c**, **d**, Peripheral blood mononuclear cells (PBMCs) pulsed with no (N), wild-type control (WT), cross-reactive (CR), and high-quality neo (M) peptides ($n = 7$). **c**, $CD8^+$ T-cell expansion and degranulation. **d**, Clonal overlap of expanded T-cell clones in **c** and archival tumours by TCR Vβ sequencing. Arrows indicate clones in archival primary tumours with rank frequencies. Venn diagrams show the number of T-cell clones expanding with mutant peptides, with cross-reactive peptides, their respective clonal overlap, and clonal overlap with archival primary tumours. Note the presence of clones recognizing both neopeptides and cross-reactive peptides in archival tumours. Years surviving after surgery are shown for each individual patient. AWD, alive with disease; NED, no evidence of disease. Horizontal bars indicate median values, error bars represent the s.e.m. *n* is the number of biologically independent samples in individual patients in **a** and **c**. *P* values were determined using a two-tailed Student's *t*-test (**a**), a two-tailed Mann–Whitney *U*-test (**b**), a one-way analysis of variance (ANOVA) with Tukey's multiple comparison test (**c**) and as described in the Methods (**d**).

in a larger cohort unselected by survival (International Cancer Genome Consortium (ICGC); $n = 166$), neoantigen quality, but not quantity, was significantly prognostic of survival and independent of confounding variables (Fig. 2b, Extended Data Fig. 6b, c), with a stable association with survival in subsampled datasets in both cohorts (Extended Data Fig. 7a, b). We conclude that neoantigen quality is a biomarker of survival in PDAC.

Recent data have shown that T-cell-recognized neoantigens can be selectively lost from the tumour cell population by either mutant allelic loss or overall reduced gene expression[11]. Consistently, genes with high-quality neoantigens evidenced a modest trend to lower mRNA expression compared to gene expression in the absence of high-quality neoantigens (Fig. 3a). To explore further possible *in vivo* high-quality neoantigen immunoediting, we examined neoantigen clonal dynamics

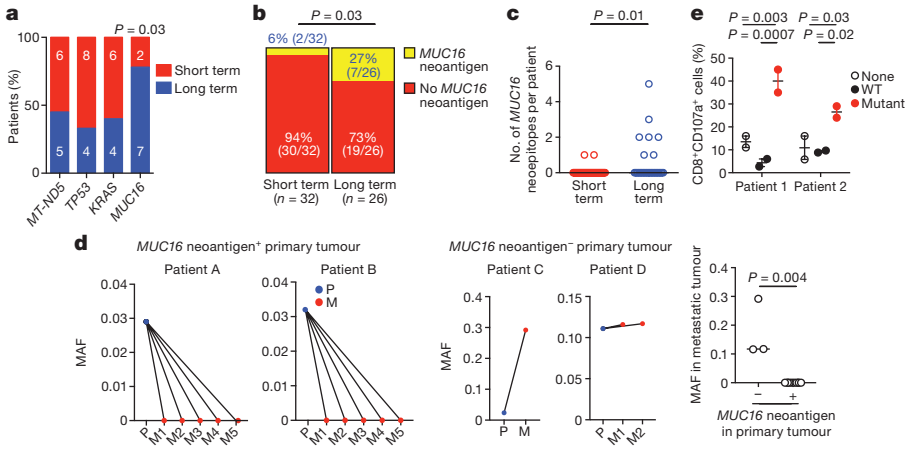


Figure 4 | *MUC16* is a neoantigenic hotspot in survivors of pancreatic cancer. a, Genes with neoantigens in more than 15% of patients. *MT-NDS* is also known as *MTND5P1*. **b**, *MUC16* neoantigen frequency. **c**, Number of *MUC16* neoepitopes per patient. Short term $n = 32$, long term $n = 26$ in **a**–**c**. **d**, Metastatic propagation of all clones in the primary tumour stratified by the presence or absence of MUC16 neoantigens. Mutant allele frequencies in matched primary–metastatic tumours (left) and metastatic tumours alone (far right) are shown in biologically independent samples in four patients. **e**, $CD8^+$ T-cell degranulation in PBMCs pulsed with no peptide (none), control wild-type peptide (WT), and MUC16 neopeptide (mutant). Data in **e** are representative of two independent experiments with similar results. *n* is the number of biologically independent samples in individual patients. Horizontal bars indicate median values, error bars represent the s.e.m. *P* values were determined using a two-sided χ^2 test (**a**, **b**), a two-tailed Mann–Whitney *U*-test (**c**, **d**) and a two-way ANOVA with Tukey's multiple comparison test (**e**).

on primary-to-metastatic tumour progression in one patient obtained through rapid autopsy. Of the three clones in the primary tumour, both clones with high-quality neoantigens were lost in multiple metastatic samples, in contrast to the clone with a low-quality neoantigen that was propagated to multiple metastatic sites (Fig. 3b, Extended Data Fig. 1f). These findings suggest differential immune fitness of clones bearing high- versus low-quality neoantigens within the same primary tumour.

Next, we sought to detect *in vivo* T-cell responses to high-quality neoantigens. We identified 7 very long-term survivors of PDAC (median overall survival 10.5 years) that normally account for less than 2% of all patients with PDAC (Extended Data Fig. 1g), and pulsed their peripheral blood mononuclear cells (PBMCs) with antigens predicted by the quality model. Notably, we observed selective $CD8^+$ T-cell expansion and degranulation (as measured by CD107a expression) to neopeptides and their homologous infectious disease-derived peptides (deemed the 'cross-reactive' peptide relative to a neoantigen) but not to their wild-type peptides (Fig. 3c, Extended Data Fig. 8a, b, Supplementary Table 3). Moreover, in all patients, identical TCR clones were significantly expanded to both the neopeptides and cross-reactive peptides (Fig. 3d). In five out of seven patients, we identified neoantigen and microbial cross-reactive peripheral T-cell clones, and in seven out of seven patients high-quality neoantigen-reactive T-cell clones that were also present in their respective archival primary tumours. Patient 3, alive and disease-free 12 years after primary tumour removal, illustrated the most extreme instance—15 neoantigen and microbial cross-reactive T-cell clones that persisted in the peripheral blood were found in the primary tumour, including the most frequent intratumoural T-cell clone with a rank frequency of 6.2% (Fig. 3d). We conclude that our quality model identifies bona fide neoantigens targeted by T cells, and that tumour-infiltrating T cells can cross-reactively recognize both cancer neoantigens and homologous non-cancer microbial antigens.

In exploring whether select genetic loci or 'immunogenic hotspots' were preferentially enriched for neoantigens, we detected four loci harbouring neoantigens in more than 15% of all patients, with one locus preferentially enriched in LTSs: the tumour antigen *MUC16*, which is a common ovarian cancer biomarker and an established T-cell immunotherapy target[12] (Fig. 4a). In tumours of LTSs, we found a fourfold higher frequency of *MUC16* neoantigens and several *MUC16* neoantigens in the same tumour, whereas non-antigenic *MUC16* mutation frequency was no different (Fig. 4b, c, Extended Data Fig. 9a). Only one patient with *MUC16* neoantigens had a hypermutated phenotype (more than 500 mutations), and exclusion of this patient did not alter the results (Fig. 4c). Supporting possible *in vivo* anti-MUC16 immunity, tumours in LTSs had lower levels of mRNA (6.6-fold), protein, and a lower mutant allele frequency (4-fold) in non-hypermutated tumours compared to short-term survivors (Extended Data Fig. 9b). Consistently, the pVAC-Seq pipeline identified *MUC16* as the most common locus generating neoantigens, after the frequently mutated oncogenes (*KRAS* and *TP53*) (Extended Data Fig. 9c). We found no differences between the two cohorts in cell-autonomous regulators of MUC16 expression, mediators of MUC16-dependent effects on tumour progression, other mucin or tumour antigen expression, or evidence of *MUC16* mutations altering RNA or protein expression[13–16] (Extended Data Fig. 9d). One interpretation of these results is that MUC16 neoantigen-specific T-cell immunity induces immunoediting of *MUC16*-expressing clones in primary tumours, and prolongs survival, given the cell-autonomous roles of MUC16 in promoting metastases[14,16]. Notably, MUC16 protein expression was low yet not absent in the tumours of LTSs, indicating antigen availability, and mutations did not affect cell intrinsic expression (Extended Data Figs 9e, 10a–c). Consistent with possible *MUC16* immunoediting, *MUC16* neoantigens in primary tumours had complete neoantigenic mutational loss in matched metastases ($n = 10$), in contrast to *MUC16* non-neoantigenic mutations that demonstrated mutation enrichment on metastatic progression (Fig. 4d, Extended Data Fig. 1f). *MUC16* was also the only locus recurrently harbouring neoantigens in both MSKCC and ICGC cohorts, outside of the most frequently mutated genes (oncogenes *KRAS*, *TP53*; largest human gene *TTN*) (Extended Data Fig. 9c). Although the propensity to generate MUC16 neoantigens may be related to its large size, we did not detect trends towards neoantigen formation based on gene size alone across cohorts or pipelines. Additionally, as the number of patients with MUC16 neoantigens in the LTS cohort was small, validation in a larger LTS dataset is warranted. Hence *MUC16* is a candidate immunogenic hotspot in PDAC.

We next stimulated peripheral blood from two LTSs (both disease-free 8 years after surgery) with predicted MUC16 neoantigens. In both patients, we observed $CD8^+$ T-cell expansion and degranulation, with expanded clones detected in archival surgically resected primary tumours (Fig. 4e, Extended Data Fig. 9f). We confirmed peripheral blood $CD8^+$ T-cell recognition of two additional MUC16 neoantigen–MHC complexes using peptide–MHC multimers in human leukocyte

antigen (HLA)-matched healthy donors (Extended Data Fig. 8c), consistent with putative MUC16 neoantigen binding by the human TCR repertoire. Hence we present evidence of *in vivo* T-cell reactivity to neoantigens in the tumour antigen MUC16, with lasting MUC16-specific T-cell immunity in survivors of PDAC.

Our results do not indicate causal associations of pre-existing microbial and anti-tumour immunity in LTSs. Instead, our data suggest that embedding microbial homology in the context of our neoantigen quality model can help to create an effective surrogate for immunogenic neoantigens. We posit two non-mutually exclusive mechanisms to explain these findings. The first is that although the naive human TCR repertoire is theoretically vast, the observed TCR recombination products are notably restricted[17], possibly representing sculpting of the TCR recombination space under the evolutionary selection pressures of pathogens, thereby skewing the repertoire to recognize their common protein features. The second is that as microbial antigens are by definition non-self sequences, enriched with documented human T-cell clones surviving thymic selection, homologous tumour neoantigens are similarly non-self sequences enriched with bona fide human T-cell clones. However, given recent evidence of intratumoural bacterial polarization of T-cell phenotypes[18], and microbial dependence of immunotherapy efficacy[9], whether homology to patient-derived microbiomes enhances relevant neoantigen identification remains unknown yet timely. Notably, we identified no associations with survival when neoantigen quality was calculated using alignment to immunogenic allergy or autoimmune-derived IEDB peptides (Extended Data Fig. 7c). However, the significant size discrepancy between the number of infectious disease-derived (Supplementary Table 2) and allergy or autoimmune-derived (Supplementary Table 4) peptides would warrant confirmation of these results in a larger allergy or autoimmune peptide dataset.

Although viral-specific T cells have been detected in human tumours[19], the presence of identical circulating and intratumoural T-cell clones reactive to both high-quality neoantigens and infectious disease-derived sequences offers proof-of-principle of neoantigenic molecular mimicry in a cancer context. Although patients with high-quality neoantigen tumours exhibited prolonged survival, it remains unproven whether these cross-reactive T-cell clones contribute to this outcome, or whether this reflects the degeneracy of the human TCR repertoire.

Our results provide insight into the heterogeneous immunobiology of PDAC, a presumed poorly immunogenic and checkpoint blockade-refractory tumour, demonstrating that neoantigens may be T-cell targets in LTSs. We propose that neoantigen quality, and not merely quantity, modulates immunogenicity, clonal fitness, and immunoselection during tumour evolution, with neoantigens in immunogenic residues such as *MUC16* emerging as apparent hotspots. Our data suggest that neoantigen-specific immunity gained during primary tumour outgrowth could be associated with decreased relapse and prolonged survival, comparable to classical mouse studies of prior tumour exposure protecting against tumour rechallenge[20]. Our findings support the development of strategies to harness neoantigen-specific immunity to treat checkpoint blockade-refractory cancers, and the identification of immunogenic hotspots for directed neoantigen targeting.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2016; accepted 2 October 2017.
Published online 8 November 2017.

1. Ino, Y. *et al.* Immune cell infiltration as an indicator of the immune microenvironment of pancreatic cancer. *Br. J. Cancer* **108,** 914–923 (2013).
2. Hiraoka, N. *et al.* Intratumoral tertiary lymphoid organ is a favourable prognosticator in patients with pancreatic cancer. *Br. J. Cancer* **112,** 1782–1790 (2015).
3. Remark, R. *et al.* In-depth tissue profiling using multiplexed immunohistochemical consecutive staining on single slide. *Sci. Immunol.* **1,** aaf6925 (2016).
4. Bindea, G. *et al.* Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* **39,** 782–795 (2013).
5. Gros, A. *et al.* Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. *Nat. Med.* **22,** 433–438 (2016).
6. Rizvi, N. A. *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348,** 124–128 (2015).
7. Hundal, J. *et al.* pVAC-Seq: A genome-guided *in silico* approach to identifying tumor neoantigens. *Genome Med.* **8,** 11 (2016).
8. Witkiewicz, A. K. *et al.* Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* **6,** 6744 (2015).
9. Zitvogel, L., Ayyoub, M., Routy, B. & Kroemer, G. Microbiome and anticancer immunosurveillance. *Cell* **165,** 276–287 (2016).
10. Łuksza, M. & Lässig, M. A predictive fitness model for influenza. *Nature* **507,** 57–61 (2014).
11. Verdegaal, E. M. E. *et al.* Neoantigen landscape dynamics during human melanoma–T cell interactions. *Nature* **536,** 91–95 (2016).
12. Chekmasova, A. A. *et al.* Successful eradication of established peritoneal ovarian tumors in SCID-Beige mice following adoptive transfer of T cells genetically targeted to the MUC16 antigen. *Clin. Cancer Res.* **16,** 3594–3606 (2010).
13. Morgado, M. *et al.* Tumor necrosis factor-α and interferon-γ stimulate MUC16 (CA125) expression in breast, endometrial and ovarian cancers through NFκB. *Oncotarget* **7,** 14871–14884 (2016).
14. Das, S. *et al.* Carboxyl-terminal domain of MUC16 imparts tumorigenic and metastatic functions through nuclear translocation of JAK2 to pancreatic cancer cells. *Oncotarget* **6,** 5772–5787 (2015).
15. Shukla, S. K. *et al.* MUC16-mediated activation of mTOR and c-Myc reprograms pancreatic cancer metabolism. *Oncotarget* **6,** 19118–19131 (2015).
16. Muniyan, S. *et al.* MUC16 contributes to the metastasis of pancreatic ductal adenocarcinoma through focal adhesion mediated signaling mechanism. *Genes Cancer* **7,** 110–124 (2016).
17. Elhanati, Y., Murugan, A., Callan, C. G., Mora, T. & Walczak, A. M. Quantifying selection in immune receptor repertoires. *Proc. Natl Acad. Sci. USA* **111,** 9875–9880 (2014).
18. Saito, T. *et al.* Two FOXP3^{+}CD4^{+} T cell subpopulations distinctly control the prognosis of colorectal cancers. *Nat. Med.* **22,** 679–684 (2016).
19. Andersen, R. S. *et al.* Dissection of T-cell antigen specificity in human melanoma. *Cancer Res.* **72,** 1642–1650 (2012).
20. Gross, L. Intradermal immunization of C3H mice against a sarcoma that originated in an animal of the same line. *Cancer Res.* **3,** 326–333 (1943).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Rudensky, A. Snyder-Charan, C. Callan, Y. Elhanati, Z. Sethna, J. Leung, J. Ruan, C. Crabtree, P. Garcia, M. Singh, A. McNeil, D. Haviland, J. Melchor and J. Tsoi for discussions, technical and editorial assistance. This work was supported by National Institutes of Health (NIH) R01DK097087-01 Pancreatic Cancer Action Network-AACR Research Acceleration Network Grant (S.D.L.), P30 CA008748-50S4 administrative supplement (S.D.L., V.P.B.), Suzanne Cohn Simon Pancreatic Cancer Research Fund (S.D.L.), National Cancer Institute K12CA184746-01A1 (V.P.B.), Damon Runyon Clinical Investigator Award (V.P.B.), Stand Up to Cancer, Lustgarden Foundation, and the National Science Foundation (J.D.W., B.D.G.), the V Foundation (V.P.B., J.A.M., J.D.W., B.D.G.), the Phil A. Sharp Innovation Award (B.D.G., J.D.W.), Swim Across America, and the Ludwig Institute for Cancer Research (J.D.W., T.M.), and the Parker Institute for Cancer Immunotherapy (D.K.W., C.I.O.C., J.D.W., T.M.). Services by the Integrated Genomics Core were funded by the National Cancer Institute Cancer Center Support Grant (P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

Author Contributions V.P.B., M.Ł., P.J.A., D.T.F., J.D.W., R.P.D., B.D.G., T.A.C., T.M. and S.D.L. conceived the study and V.P.B., J.D.W., T.A.C., B.D.G., T.M. and S.D.L. designed all experiments. V.P.B., M.Ł., J.N.Z., V.M., J.A.M., R.R., B.H., G.A., U.B., Y.S., D.K.W., C.I.O.C., O.G.-H., M.A., B.M., J.Z., J.L., J.S., M.A.-A., R.Z., N.R., M.S., Z.L.K., O.B., A.J.L., P.J.A., D.T.F., M.M., S.G., C.A.I.-D. and members of the Australian Pancreatic Cancer Genome Initiative acquired and analysed data. G.A., U.B. and O.B. performed the histopathological analyses. M.A. and C.A.I.D. performed tissue acquisition, and mutational identification for rapid autopsy tissues. V.M., N.R., T.A.C., D.K.W. and C.I.O.C. performed the neoantigen identification. M.Ł. and B.D.G. constructed the neoantigen fitness models. V.P.B., J.N.Z., M.A.-A. and J.A.M. performed the *in vitro* T-cell assays. O.G.-H. performed the transfections, immunocytochemistry, and western blots. M.G. provided statistical oversight. V.P.B., M.Ł., J.N.Z., D.T.F., J.D.W., R.P.D., B.D.G., T.A.C., T.M. and S.D.L. interpreted the data. V.P.B., M.Ł., J.D.W., B.D.G., T.M. and S.D.L. drafted the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to V.P.B. (balachav@mskcc.org).

Reviewer Information *Nature* thanks E. Verdegaal, R. Vonderheide and the other anonymous reviewer(s) for their contribution to the peer review of this work.

Australian Pancreatic Cancer Genome Initiative
Garvan Institute of Medical Research Amber L. Johns[1], R. Scott Mead[1,2], Anthony J. Gill[1,3], David K. Chang[1,4], Skye H. McKay[1], Lorraine A. Chantrill[1,5], Venessa T. Chin[1], Angela Chou[1,5], Jeremy L. Humphris[1], Marina Pajic[1], Angela Steinmann[1], Mehreen Arshi[1], Ali Drury[1], Danielle Froio[1], Ashleigh Morgan[1], Paul Timpson[1], David Hermann[1], Claire Vennin[1], Sean Warren[1], Mark Pinese[1], Jianmin Wu[1], Andreia V. Pinho[1]

Prince of Wales Hospital, R. Scott Mead[1,2], Katherine Tucker[2], Lesley Andrews[2]

Royal North Shore Hospital Anthony J. Gill[1,3], Jaswinder S. Samra[3], Jennifer Arena[3], Nick Pavlakis[3], Hilda A. High[3], Anubhav Mittal[3]

University of Glasgow David K. Chang[1,4], Andrew V. Biankin[4], Peter Bailey[4], Sancha Martin[4], Elizabeth A. Musgrove[4], Marc D. Jones[4], Craig Nourse[4], Nigel B. Jamieson[4]

St Vincent's Hospital Angela Chou[1,5], Lorraine A. Chantrill[1,5], Alina Stoita[5], David Williams[5], Allan Spigelman[5]

QIMR Berghofer Medical Research Institute Nicola Waddell[6], John V. Pearson[6], Ann-Marie Patch[6], Katia Nones[6], Felicity Newell[6], Pamela Mukhopadhyay[6], Venkateswar Addala[6], Stephen Kazakoff[6], Oliver Holmes[6], Conrad Leonard[6], Scott Wood[6], Christina Xu[6]

University of Melbourne, Centre for Cancer Research Sean M. Grimmond[7], Oliver Hofmann[7]

University of Queensland, Institute for Molecular Bioscience Peter J. Wilson[8], Angelika Christ[8], Tim Bruxner[8]

Bankstown Hospital Ray Asghari[9], Neil D. Merrett[9], Darren Pavey[9], Amitabha Das[9]

Liverpool Hospital Annabel Goodwin[10,11], Peter H. Cosman[10], Kasim Ismail[10], Chelsie O'Connor[10]

Royal Prince Alfred Hospital, Chris O'Brien Lifehouse Caroline L. Cooper[11], Annabel Goodwin[10,11], Peter Grimison[11], James G. Kench[11], Charbel Sandroussi[11]

Westmead Hospital Vincent W. Lam[12], Duncan McLeod[12], Adnan M. Nagrial[12], Judy Kirk[12], Virginia James[12]

Fremantle Hospital Michael Texler[13], Cindy Forest[13], Krishna P. Epari[13], Mo Ballal[13], David R. Fletcher[13], Sanjay Mukhedkar[13]

St John of God Healthcare Nikolajs Zeps[14], Maria Beilin[14], Kynan Feeney[14]

Royal Adelaide Hospital Nan Q. Nguyen[15], Andrew R. Ruszkiewicz[15], Chris Worthley[15]

Flinders Medical Centre John Chen[16], Mark E. Brooke-Smith[16], Virginia Papangelis[16]

Envoi Pathology Andrew D. Clouston[17], Patrick Martin[17]

Princess Alexandria Hospital Andrew P. Barbour[18], Thomas J. O'Rourke[18], Jonathan W. Fawcett[18], Kellee Slater[18], Michael Hatzifotis[18], Peter Hodgkinson[18]

Austin Hospital Mehrdad Nikfarjam[19]

Johns Hopkins Medical Institutes James R. Eshleman[20], Ralph H. Hruban[20], Christopher L. Wolfgang[20], Mary Hodgin[20]

ARC-Net Centre for Applied Research on Cancer Aldo Scarpa[21], Rita T. Lawlor[21], Stefania Beghelli[21], Vincenzo Corbo[21], Maria Scardoni[21] & Claudio Bassi[21]

[1]The Kinghorn Cancer Centre, Garvan Institute of Medical Research, 370 Victoria Street, Darlinghurst, Sydney, New South Wales 2010, Australia. [2]Prince of Wales Hospital, Barker Street, Randwick, New South Wales 2031, Australia. [3]Royal North Shore Hospital, Westbourne Street, St Leonards, New South Wales 2065, Australia. [4]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK. [5]St Vincent's Hospital, 390 Victoria Street, Darlinghurst, New South Wales, 2010 Australia. [6]QIMR Berghofer Medical Research Institute, 300 Herston Road, Herston, Queensland 4006, Australia. [7]University of Melbourne, Centre for Cancer Research, Victorian Comprehensive Cancer Centre, 305 Grattan Street, Melbourne, Victoria 3000, Australia. [8]Institute for Molecular Bioscience, University of Queensland, St Lucia, Queensland 4072, Australia. [9]Bankstown Hospital, Eldridge Road, Bankstown, New South Wales 2200, Australia. [10]Liverpool Hospital, Elizabeth Street, Liverpool, New South Wales 2170, Australia. [11]Royal Prince Alfred Hospital, Missenden Road, Camperdown, New South Wales 2050, Australia. [12]Westmead Hospital, Hawkesbury and Darcy Roads, Westmead, New South Wales 2145, Australia. [13]Fremantle Hospital, Alma Street, Fremantle, Western Australia 6959, Australia. [14]St John of God Healthcare, 12 Salvado Road, Subiaco, Western Australia 6008, Australia. [15]Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. [16]Flinders Medical Centre, Flinders Drive, Bedford Park, South Australia 5042, Australia. [17]Envoi Pathology, 1/49 Butterfield Street, Herston, Queensland 4006, Australia. [18]Princess Alexandria Hospital, Cornwall Street & Ipswich Road, Woolloongabba, Queensland 4102, Australia. [19]Austin Hospital, 145 Studley Road, Heidelberg, Victoria 3084, Australia. [20]Johns Hopkins Medical Institute, 600 North Wolfe Street, Baltimore, Maryland 21287, USA. [21]ARC-NET Center for Applied Research on Cancer, University of Verona, Via dell'Artigliere, 19 37129 Verona, Province of Verona, Italy.

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The investigators were not blinded to allocation during experiments and outcome assessment.

Patient samples. *MSKCC PDAC cohort.* All tissues were collected at MSKCC following study protocol approval by the MSKCC Institutional Review Board. Informed consent was obtained for all patients. The study was in strict compliance with all institutional ethical regulations. All tumour samples were surgically resected primary pancreatic ductal adenocarcinomas. Patients treated with neoadjuvant therapy were excluded. All tumours were subjected to pathological re-review and histological confirmation by two expert PDAC pathologists before analysis. LTSs were defined as patients with overall survival of >3 years from surgery, short-term survivors as patients with survival >3 months and <1 year from surgery, to exclude perioperative mortalities.

ICGC cohort. Clinical characteristics of the ICGC cohort have been described previously[21].

Rapid autopsy cohort. Primary and metastatic tumour samples were collected posthumously from four patients as part of the Gastrointestinal Cancer Rapid Medical Donation program at Johns Hopkins Hospital[22]. Informed consent was obtained from all subjects. This program was deemed in accordance with the Health Insurance Portability and Accountability Act and the study protocol was approved by the Johns Hopkins Hospital Institutional Review Board. The study was in strict compliance with all institutional ethical regulations.

Tissue microarray. Tissue microarrays were constructed from tumour and adjacent non-tumour cores from formalin-fixed, paraffin-embedded tissue blocks in short- ($n=45$ tumours, 5 normal tissue) and long-term ($n=51$ tumours, 5 normal tissue) survivors of PDAC. Histology sections were reviewed by two expert PDAC pathologists and the most representative areas were selected and marked on haematoxylin and eosin slides. 1 mm diameter cores were sampled from three different tumour regions per patient using an automated TMA Grand Master (Perkin Elmer). 5-μm sections were prepared from tissue microarray blocks for immunohistochemistry. Patient subsets were randomly selected to undergo tissue microarray construction.

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 μg ml^{-1}) were purchased from Abcam. Antibodies to MUC1 (clone M695, dilution 1:100), and mesothelin (clone 5B2, dilution 1:50) were purchased from Vector Laboratories. Immunohistochemistry was performed using standard techniques. MUC16 expression was scored as described[23]. For each core, a cumulative MUC16 expression score was calculated as the product of a score for the frequency of tumour cells expressing MUC16 (0–25% = 1; 26–50% = 2; 51–75% = 3; 76–100% = 4) and a score for the intensity of staining (0 = negative; 1 = weak; 2 = moderate; and 3 = strong). The median expression score across triplicate cores is reported as the final score for each patient.

Multiplexed consecutive immunohistochemistry on the same slide was performed as described[3]. Tissue microarray slides were incubated overnight at 37 °C. Then, paraffin was removed using xylene and the tissue rehydrated before incubation in antigen-retrieval solution at 95 °C for 45 min (pH 9 Target Retrieval Solution, Dako). After endogenous peroxidase inhibition and FcR blocking, granzyme B was stained with anti-granzyme B monoclonal antibody (clone GrB-7, Dako) for 1 h at room temperature. After signal amplification with a horseradish peroxidase (HRP)-labelled polymer (Dako), the revelation was performed using 3-amino-9-ethylcarbazole (AEC, Vector Laboratories). Then slides were immersed in haematoxylin, rinsed in distilled water and mounted in aqueous-based mounting medium (Glycergel, Dako). After imaging using a whole-slide scanner, the slides were subjected to the Multiplexed Immunohistochemical Consecutive Staining on Single Slide protocol (MICSSS) and stained for T cells (CD3, clone 2GV6, Ventana and CD8, clone C8/144b, Dako), regulatory T cells (FoxP3, clone 236A/E7, Abcam), B cells (CD20, clone L26, Dako), macrophages (CD68, clone KP1, Dako), mature dendritic cells (DC-LAMP, clone 1010E1.01, Novus Biologicals), MHC class I cells (HLA-ABC, clone EMR8-5, Abcam) and tumour cells (CK19, clone EP1580Y, Abcam).

Immunofluorescence. For CD4, FoxP3, and CK19 staining, sections were first incubated with anti-CD4 (Ventana, 790-4423, 0.5 μg ml^{-1}) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Laboratories, PK6101) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 488 (Invitrogen, T20922) prepared according to the manufacturer's instructions with predetermined dilutions. Next, slides were incubated with anti-FoxP3 (Abcam, ab20034, 5 μg ml^{-1}) for 4 h, followed by a 60 min incubation with biotinylated horse anti-mouse IgG (Vector Labs, MKB-22258) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 568 (Invitrogen, T20914) prepared according to the manufacturer's instructions with predetermined dilutions. Finally, sections were incubated with anti-CK19 (Abcam, ab52625, 1 μg ml^{-1}) for 5 h, followed by a 60 min incubation with biotinylated goat anti-rabbit IgG (Vector, PK6101) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 647 (Invitrogen, T20936) prepared according to the manufacturer's instructions with predetermined dilutions. After staining, slides were counterstained with 4′,6-diamidino-2-phenylindole (DAPI, Sigma Aldrich, D9542, 5 μg ml^{-1}) for 10 min and coverslips mounted with Mowiol.

For CD3, CD8 and CK19 staining, slides were first incubated with anti-CD3 (DAKO, A0452, 1.2 μg ml^{-1}) for 4 h, followed by a 60 min incubation with biotinylated goat anti-rabbit IgG (Vector Labs, PK6101) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 488 (Invitrogen, T20922) prepared according to the manufacturer's instructions with predetermined dilutions. Next, slides were incubated with anti-CD8 (Ventana, 790-4460, 0.35 μg ml^{-1}) for 5 h, followed by a 60 min incubation with biotinylated goat anti-rabbit IgG (Vector, PK6101) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 568 (Invitrogen, T20914) prepared according to the manufacturer's instructions with predetermined dilutions. Finally, sections were incubated with anti-CK19 (Abcam, ab52625, 1 μg ml^{-1}) for 5 h, followed by 60 min incubation with biotinylated goat anti-rabbit IgG (Vector, PK6101) at 1:200 dilution. The detection was performed with streptavidin–HRP D (part of DABMap kit, Ventana Medical Systems), followed by incubation with tyramide Alexa Fluor 647 (Invitrogen, T20936) prepared according to the manufacturer's instructions with predetermined dilutions. After staining, slides were counterstained with DAPI (Sigma Aldrich, D9542, 5 μg ml^{-1}) for 10 min and coverslips mounted with Mowiol.

Digital image processing and analysis. Tissue microarrays for each immunohistochemical stain were individually digitally scanned using Pannoramic Flash (3DHistech) with a 40×/0.95 numerical aperture (NA) objective. Image registration and alignment was performed using ImageJ (NIH). Regions of interest were drawn for each core and then transferred to others using CaseViewer (3DHistech). Each region from each scan was exported as a TIF file at full resolution (0.243 μm per pixel). Images of the same core from multiple scans were stacked together and aligned using Linear Stack Alignment with SIFT algorithm from FIJI/ImageJ (NIH). Once aligned, the RGB images were colour deconvoluted to separate AEC and haematoxylin stainings and converted into eight-bit pseudo-fluorescent images. Individual immunohistochemical targets were sequentially assigned to fluorescent channels and subsequently merged. Haematoxylin staining was used to segment and count the number of nucleated cells in the core. After processing the images using background subtraction and median filter, staining was thresholded and split using Biovoxxel Watershed Irregular Features plugin. Regions of interest were drawn around each cell and matched to the signals from all other AEC stainings to count the number of positive cells for each staining. Total tissue area was measured by setting a very low threshold for haematoxylin images. For quantification, all nucleated cells were identified, followed by an intensity-based threshold determination of each target to identify positive cells. Triplicate cores were quantified followed by determination of the median number of cells per square millimetre of tissue (ImageJ, NIH). Quantification of cells detected using immunofluorescence was performed in a similar fashion. $CD8^+$ T cells were defined as $CD3^+CD8^+$ cells, cytolytic $CD8^+$ T cells as $CD3^+CD8^+$granzyme B^+ cells, mature dendritic cells as DC-$LAMP^+$ cells, regulatory T cells as $CD3^+FoxP3^+$ cells, macrophages as $CD68^+$ cells, $CD4^+$ T cells as both $CD3^+CD8^-$ and $CD4^+$ cells, B cells as $CD20^+$ cells, and class-I expressing cells as MHC-I^+ cells.

Nucleic acid extraction. *MSKCC PDAC cohort.* 10 μm slides were cut from OCT-embedded frozen tumour and matched normal tissues. Sections were brought to containers with 70% ethanol for OCT removal. After OCT removal, specimens were dissected for subsequent DNA and RNA extraction. For whole-exome sequencing, tumour islands of >70% cellularity were macrodissected based on expert PDAC pathological review, and DNA was extracted using the DNeasy kit. Total RNA from fresh frozen OCT-embedded tissues was extracted using TRIzol RNA Isolation Reagents (15596-026, Life Technologies).

Rapid autopsy cohort. Genomic DNA was extracted using standard phenol–chloroform extraction followed by precipitation in ethanol. Quantification was achieved by LINE assay. Tissue samples confirmed to be of high quality and sufficient concentration were used for subsequent whole-exome sequencing.

Transcriptome analysis. Extracted RNA was qualified on an Agilent BioAnalyzer and quantified by fluorometry (Ribogreen). Preparation of RNA for whole

transcriptome expression analysis was done using the WT Pico Reagent Kit (Affymetrix). Reverse transcription was initiated at the poly-A tail as well as throughout the entire length of RNA to capture both coding and multiple forms of non-coding RNA. RNA amplification was achieved using low-cycle PCR followed by linear amplification using T7 *in vitro* transcription technology. The cRNA was then converted to biotinylated sense-strand DNA hybridization targets. The prepared target was hybridized to GeneChip Human Transcriptome Array 2.0 (Affymetrix). Wash and scan was performed using the GeneChip Hybridization, Wash and Stain Kit using a Fluidics Station 450/250. Arrays were scanned using the GeneChip Scanner 3000. Data analysis for the array was done using Affymetrix Expression Console Software (SST-RMA algorithm to summarize the signal from array probesets). A dendritic cell signature was defined as described previously, using the genes *CCL13*, *CCL17*, *CCL22*, *PPFIBP2*, *NPR1*, *HSD11B1*, and *CD209* (also known as *DC-SIGN*)[4]. Patient subsets were randomly selected to undergo transcriptomic profiling.

TCR Vβ sequencing. Frozen tumour (short term $n = 30$, long term $n = 30$) and paired non-tumour adjacent pancreas tissue (short term $n = 30$, long term $n = 30$) samples were processed (Adaptive Biotechnologies). Genomic DNA was extracted according to the manufacturer's instructions (QIAsymphony, Qiagen). The quantity and quality of extracted DNA was verified before sequencing. Using a standard quantity of input DNA, the TCR Vβ CDR3 regions were amplified and sequenced using the survey multiplexed PCR ImmunoSeq assay. The ImmunoSeq platform combines multiplex PCR with high-throughput sequencing to selectively amplify the rearranged complementarity-determining region 3 (CDR3) of the TCR, producing fragments sufficiently long to identify the VDJ region spanning each unique CDR3. 45 forward primers specific for TCR Vβ gene segments and 13 reverse primers specific to TCR Jβ gene segments were used (Adaptive Biotechnologies). Read lengths of 156 bp were obtained using the Illumina HiSeq System. The ImmunoSeq assay allows for quantitative assessment of both total and unique TCRs in a sample, as it uses a complete synthetic repertoire of TCRs to establish an amplification baseline and adjust the assay chemistry to correct for primer bias. Barcoded, spiked-in synthetic templates were also used to measure and correct for sequencing coverage and residual PCR bias. Output data were then filtered and clustered using the relative frequency ratio between similar clones and a modified nearest-neighbour algorithm, to merge closely related sequences and remove PCR and sequencing errors. The number of rearranged TCRs per diploid genome in the input material (total number of T cells) was estimated as described previously[24]. The frequency of T cells was determined as the total number of T cells per total number of sequenced cells in the input material. A T-cell clone was defined as a T cell with a unique TCR Vβ CDR3 amino acid sequence. Clonality was defined as (1 − normalized entropy). Normalized entropy was calculated as the Shannon entropy divided by the logarithm of the number of unique productive (exonic) TCR sequences. Shannon entropy equals the clonal abundance of all productive TCR sequences in the input material. For *in vitro* stimulated cells, clones with identical amino acid sequences that expanded >2-fold on day 21 compared to day 0, and fulfilled the Fisher's exact test and Storey's *Q* value for false discovery rate were defined as expanded. Data analysis was performed using Adaptive Biotechnologies ImmunoSeq Analyzer (Analyzer 3.0).

Whole-exome sequencing. For all MSKCC PDAC patients, 500 ng of genomic DNA was fragmented to a target size of 150–200 bp on the Covaris LE220 system. Barcoded libraries (Kapa Biosystems) were subjected to exon capture by hybridization using the SureSelect Human All Exon 51MB V4 kit (Agilent). DNA libraries were subsequently sequenced on a HiSeq 4000 (Illumina) in a Paired End 100/100, using the TruSeq SBS Kit v3 (Illumina) with a target coverage of 150× for tumour samples and 70× for matched normal (MSKCC Center for Molecular Oncology). Sequence data were demultiplexed using CASAVA, and after removal of adaptor sequences using cutadapt (v1.6), reads were aligned to the reference human genome (hg19) using the Burrows–Wheeler Alignment tool (bwa mem v0.7.12). Duplicate-read removal, InDel realignment and base quality score recalibration were performed using the Genome Analysis Toolkit (GATK) according to GATK best practices, as described previously[6]. Variants were identified on processed data using Mutect, Mutect rescue (SNPs) and HaplotypeCaller (insertions/deletions). A mean unique sequence coverage of 167.45× was achieved for tumour samples and 84.75 for normal samples. All *MUC16* mutations were manually reviewed by 3 investigators using the Integrated Genomics Viewer v2.3.72. Whole-genome and whole-exome sequencing for ICGC[21] patients has been described previously. For all ICGC and rapid autopsy samples, BAM files were re-processed and mutations identified as per the above outlined MSKCC protocol. Depth of sequencing for the rapid autopsy samples ranged from 150× to 250×.

HLA typing. HLA typing for patients with PDAC was performed *in silico* using the tool Short Oligonucleotide Analysis Package (SOAP)-HLA (http://soap.genomics.org.cn/SOAP-HLA.html).

Somatic mutation immunogenicity predictions. *MSKCC pipeline.* Immunogenicity of somatic mutations was estimated using a previously described bioinformatics tool called NASeek[6]. In brief, NASeek is a computational algorithm that first translates all mutations in exomes to strings of 17 amino acids, for both the wild-type and mutated sequences, with the amino acid resulting from the mutation centrally situated. Second, it evaluates putative MHC class I binding for both wild-type and mutant nonamers using a sliding window method using NetMHC3.4 (http://www.cbs.dtu.dk/services/NetMHC/) for patient-specific HLA types, to generate predicted binding affinities for both peptides. NASeek finally assesses for similarity between nonamers that were predicted to be presented by patient-specific MHC class I. All mutations with binding scores below 500 nM are defined as neoantigens (Supplementary Table 1). As the MSKCC pipeline was, on average, more stringent with respect to the number of neoantigens identified (in comparison to the pVAC-Seq pipeline below), all neoantigen predictions were performed with the MSKCC pipeline unless otherwise specified.

pVAC-Seq pipeline. As an independent algorithm to identify neoantigens, we used the pVAC-Seq pipeline[7] with the NetMHCpan binding strength predictor (<500 nM binding strength). As recommended, we used the variant effect predictor from Ensembl[25] to annotate variants for downstream processing by pVAC-Seq.

Neoantigen fitness modelling. The fitness of a clone is defined as

$$\frac{dN_\alpha}{dt} = F_\alpha N_\alpha$$

where N_α is the effective population size of tumour clone α and F_α is the fitness of clone α. We assume that the fitness effects due to the immune system are separable from other tumour fitness effects; that is, $F_\alpha = F_\alpha^0 + F_\alpha^I$, where F_α^I denotes the contribution to fitness due to selection on neoantigens and F_α^0 denotes contributions from other factors, such as classical tumour driver mutations. Our model assumes the two components are separable, which is essentially equivalent to assuming that targeting of neoantigens is emanating largely from passenger mutations. We test the hypothesis that, owing to T-cell-mediated immune recognition, for some tumours, the deleterious effects of immune pressure due to mutation-derived neoantigens can become a dominant fitness effect, either counteracting or substantially slowing the growth rate of the tumour.

In general, the predicted effective total tumour population size at time τ, denoted as $n(\tau)$, is

$$n(\tau) = \frac{N(\tau)}{N(0)} = \sum_\alpha X_\alpha(0)\ \exp(F_\alpha \tau) = \sum_\alpha X_\alpha(0)\ \exp((F_\alpha^0 + F_\alpha^I)\tau)$$

where $N(0) = \sum_\alpha N_\alpha(0)$ is the initial total effective population size of all clones within the tumour, and $X_\alpha(0) = N_\alpha(0)/N(0)$ is the frequency of clone α. The initial frequency of clone α is the size of a clone estimated from a tumour's phylogenetic tree, using the PhyloWGS software package (https://github.com/morrislab/phylowgs)[26]. For the tumours in our cohort there was not a discernable difference in the distributions of pancreatic cancer driver mutations across clones. As a result we assume $F_\alpha^0 \approx F_0$. Therefore,

$$n(\tau) = e^{F_0 \tau} \sum_\alpha X_\alpha(0)\ \exp(F_\alpha^I \tau)$$

We rank samples according to the value of $n(\tau)$, on the basis of their neoantigen–immune interactions only. This is equivalent to the assumption that $e^{F_0\tau}$ also does not vary as considerably across samples when compared with $\sum_\alpha X_\alpha(0)\ \exp(F_\alpha^I \tau)$. As a result

$$n(\tau) \propto \sum_\alpha X_\alpha(0)\ \exp(F_\alpha^I \tau)$$

when considered across our cohort, and we use the immune component as a basis for ranking. Our hypothesis is that there are tumours in which typically $|F_0| \leq |F_\alpha^I|$, and the deleterious effects of neoantigen recognition sufficiently counterbalance the growth of the tumour. By ranking tumours according to $n(\tau)$, we therefore test whether tumours with the strongest deleterious effects owing to neoantigen fitness have better survival.

For a given neoantigen with sequence $\boldsymbol{s}$ we calculate R as the probability that neoantigen $\boldsymbol{s}$ is recognizable by the T-cell receptor repertoire. We do so by calculating the probability that a neoantigen sufficiently aligns to an epitope $\boldsymbol{e}$ from the IEDB via a thermodynamic model using the alignment score between the two peptides, $|\boldsymbol{s}, \boldsymbol{e}|$:

$$R = Z(k)^{-1} \sum_{\boldsymbol{e} \in \text{IEDB}} \exp[-k(a - |\boldsymbol{s}, \boldsymbol{e}|)]$$

where a represents the horizontal displacement of the binding curve and k sets the slope of the curve at a. The partition function then becomes

$$Z(k) = 1 + \sum_{\boldsymbol{e} \in \mathrm{IEDB}} \exp[-k(a - |\boldsymbol{s}, \boldsymbol{e}|)]$$

The set of known positive epitopes were derived from the IEDB, restricting the search to all human infectious disease, class I-restricted targets with positive immune assays (http://www.iedb.org/). As the peptides in IEDB can change over time, the version of IEDB used in our study is included (Supplementary Table 2). The alignments between all neoantigens and IEDB epitope sequences are found with the blastp algorithm using BLOSUM62 matrix (gap opening penalty $= -11$, gap extension penalty $= -1$). For the identified alignments the alignment scores are then computed with Biopython Bio.pairwise2 package (http://biopython.org).

The amplitude due to relative MHC dissociation constants between a neoantigen and its wild-type counterpart is $A \approx K_{\mathrm{d}}^{\mathrm{WT}}/K_{\mathrm{d}}^{\mathrm{MT}}$, approximated here by the ratio of their inferred MHC binding affinities, which are inferred for neoantigens and their wild-type counterpart using NetMHC3.4 (http://www.cbs.dtu.dk/services/NetMHC/) as described above. We use the standard cutoff for $K_{\mathrm{d}}^{\mathrm{MT}}$, the mutant dissociation constant, used in the literature, that is $K_{\mathrm{d}}^{\mathrm{MT}} < 500$ nM. For a given neoantigen, the quantity $A \times R$ is referred to as the recognition potential of a neoantigen, and is a measure of neoantigen quality in regard to T-cell receptor recognition. Fitness due to T-cell-mediated neoantigen recognition within a clone is defined as

$$F_{\alpha}^{I} = -\max_{i \in \mathrm{Clone}\,\alpha} (A_i \times R_i)$$

where i is an index running over neoantigens within a clone; that is, within a clone α, the maximal product of the amplitude A_{i} and the recognition probability R_{i} for a neoantigen. As an alternative hypothesis, fitness in the neoantigen load hypothesis is defined as $F_{\alpha}^{I} = -L_{\alpha}$, where L_{α} is the simple number of neoantigens in clone α. For all cases, we computed the neoantigen load without clonal phylogeny, which was the standard benchmark, and the neoantigen load with clonal phylogeny, by taking into account the effective size of clones in which neoantigens were contained. We also compared our results to those obtained using the wild-type recognition potential alone—in which case our MHC amplitude was 1—and the fitness model without clonality, which essentially just scores the best neoantigen across the tumour.

We split samples by the median value of the cohort, with samples below this value designated as a low fitness group (neoantigen quality$^{\mathrm{high}}$ group), and those above as a high fitness group (neoantigen quality$^{\mathrm{low}}$ group). We then compared survival for high- versus low-quality groups, expecting high-quality tumours to be related to longer patient survival times. Our model has three parameters: a, k and τ. We observe significant separation of patients at a sufficiently steep probability binding function; we therefore set the slope parameter $k = 1$. We report values for the shift and time parameters, a and τ, which optimized survival, as quantified by the log-rank test score using the lifelines software package (https://lifelines.readthedocs.io/en/latest/). To test the stability of this choice, we derived the optimal value for subsampled datasets, with subsampling frequencies of 0.5, 0.7, and 0.9. The optimal parameters obtained on the full dataset, $a = 26$ and $\tau \in [0.02, 0.04]$ were the most frequent choice for all subsampling frequencies (as shown on the distribution of optimal parameters in Extended Data Fig. 7a).

We repeated the same analysis on the larger ICGC cohort with 166 patients. The optimal parameters obtained on the full dataset were $a = 23$ and $\tau \in [0.19, 0.21]$; however the optimum depended only marginally on τ, raising significant patient segregation for a very broad interval of values (in particular at $\tau \in [0.02, 0.04]$, optimal for the MSKCC cohort, we obtain $P < 0.01$, see Extended Data Fig. 7a). We observe that the horizontal displacement parameter a is lower than in the MSKCC cohort. We attribute this trend to the difference in the survival time distribution in the two cohorts and the MSKCC cohort containing extreme LTSs, with probably 'higher quality' neoantigens.

***In vitro* T-cell assays.** Fresh blood was collected from seven PDAC LTSs whose tumours were identified on the basis of whole-exome sequencing and *in silico* predictions to harbour neoantigens. PBMCs were isolated by density centrifugation over Ficoll-Paque Plus (GE Healthcare). Peptides were generated for immunodominant neoantigens as predicted by the neoantigen quality algorithm (the neoantigen with the maximum quality score within a tumour, driven by maximal TCR recognition probability was defined as the immunodominant neoantigen), MUC16 neoantigens, and the corresponding wild-type nonamers (Peptide 2.0, Supplementary Table 2). *In vitro* peptide stimulation was performed as described with minor modifications[6]. In brief, 1×10^6 PBMCs were cultured with mutant or wild-type peptides ($10\,\mu\mathrm{g\,ml^{-1}}$) on day 1. IL-2 ($50\,\mathrm{U\,ml^{-1}}$) and IL-15 ($10\,\mathrm{ng\,ml^{-1}}$) were added on day 2 and every subsequent 2–3 days. Mutant and wild-type peptides were added to respective cultures on day 7, and day 14 for second and third rounds of restimulation. On day 21, cells were restimulated in the presence of peptide for 5 h and cells were subsequently stained as per the manufacturer's instructions or subject to sequencing. A peptide pool of 23 class-I-restricted viral peptides from human cytomegalovirus and influenza virus (CEF, Cellular Technology Limited) served as a positive control (data not shown). Normalized expansion on day 21 of culture was defined as (absolute CD8$^+$ T-cell expansion day 21 − absolute CD8$^+$ T-cell expansion day 0)/(absolute CD8$^+$ T-cell expansion day 0).

Flow cytometry. Fresh blood and tumour samples from six individual patients undergoing elective surgery at MSKCC were collected. Informed consent was obtained according to a MSKCC Institutional Review Board-approved protocol. Blood was drawn at the time of surgery, and PBMCs were isolated by density centrifugation over Ficoll-Paque Plus (GE Healthcare). Tumour and draining lymph-node tissues were processed immediately after removal from the patient and single-cell suspensions were prepared. To assess whether T cells bind *in silico* predicted neoantigen–HLA complexes, T cells of PBMCs from HLA-specific healthy donors (Precision For Medicine) were assessed for binding to MUC16-neoantigen–MHC multimers. MUC16-MHC–fluorescein isothiocyanate (FITC) multimers were designed to HLA-B0801 (Immudex) with nonamer peptide sequences derived based on mutated MUC16 sequences identified on whole-exome sequencing that were *in silico* predicted to be immunogenic. Single-cell PBMC suspensions were surface stained for anti-human CD45, CD3, CD56, CD8, CD4, CD107a, and MHC-multimers according to the manufacturer's instructions. 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). Flow cytometry was performed on an LSRFortessa (BD Biosciences) and data were analysed using FlowJo Software (Tree Star).

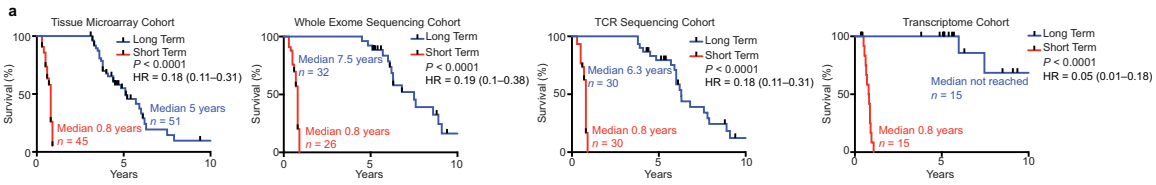
Plasmids and transfection. phrGFP II-C vector (Stratagene) expressing human wtMUC16$^{\mathrm{c354N206}}$–GFP fragment was a gift from the D. Spriggs laboratory. A point mutation at position R15C in the C-terminal portion of wtMUC16 was introduced using QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) and validated by sequencing. HEK293T cells were transiently transfected with phrGFP II-C vector (empty vector), wtMUC16$^{\mathrm{c354N206}}$–GFP, mtMUC16$^{\mathrm{R16C}}$–GFP using Lipofectamine 3000 (Invitrogen) according to the manufacturer's instructions. Protein extracts from transiently transfected cells were analysed by western blot (clone 4H11, anti-MUC16-carboxy-terminal monoclonal antibody[27]) and β-actin–HRP antibody (Sigma Aldrich). Transfected cells were also analysed by immunocytochemistry (clone 4H11).

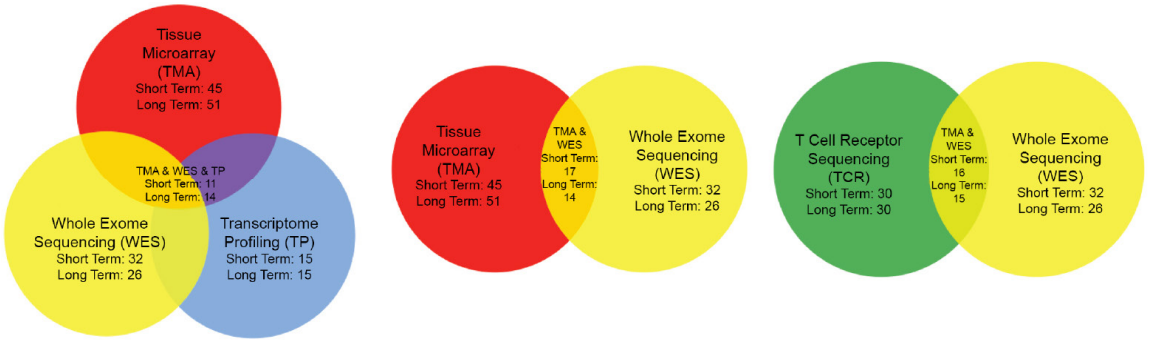
Statistics. Comparisons between two groups were performed using an unpaired two-tailed Mann–Whitney U-test (unpaired samples), a paired two-tailed Mann–Whitney U-test (paired samples), and a two-tailed Student's t-test (normally distributed parameters). Multiple samples were compared using a Kruskal–Wallis test (non-grouped) and ANOVA with Tukey's post-test for multiple comparisons (grouped). Survival curves were compared using a log-rank test (Mantel–Cox). Categorical variables were compared using a χ^2 test. All comparison groups had equivalent variances. Specific clonal expansion on day 21 compared to day 0 (>2-fold expansion by Fisher's exact test and Storey's Q value for false discovery rate) was assessed by TCR Vβ sequencing. In Fig. 3d, we tested whether the number of clones expanding under more than one condition (mutant, cross reactive, tumour) is significantly different than that which would be observed by chance using Poisson-family saturated log-linear models with the canonical link function and the corresponding P values for the third-degree term in these models. We used a hypothetical value of 10^7 for the total number of clones in the blood for this calculation. The glm function in R version 3.4 was used for fitting these models. $P < 0.05$ was considered to be statistically significant. Data analysis was performed using statistical software (Prism 7.0, GraphPad software; Cox regression using STATA 13.1).

Data availability. Data presented in this study can be downloaded from https://dcc.icgc.org/repositories under the identifier PACA-AU. Transcriptomic data are available under Gene Expression Omnibus (GEO) accession number 89997. Source Data are provided for all experiments, including Figs 1b, c, e, 3a–d, 4a–e, and Extended Data Figs 2a–c, 3a, 3c, 3e, 4c, 6b, 8e, 9a–f, 10a. All other data are available from the corresponding author upon reasonable request.

Code availability. The data (Supplementary Table 1), the computational algorithm and software used, (Extended Data Fig. 7), and the source code (Supplementary Data 1) allowing for reproduction of the neoantigen quality evaluation in this manuscript are included as indicated above.

21. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531,** 47–52 (2016).
22. Embuscado, E. E. *et al.* Immortalizing the complexity of cancer metastasis: genetic features of lethal metastatic pancreatic cancer obtained from rapid autopsy. *Cancer Biol. Ther.* **4,** 548–554 (2005).
23. Haridas, D. *et al.* Pathobiological implications of MUC16 expression in pancreatic cancer. *PLoS One* **6,** e26839 (2011).
24. Tumeh, P. C. *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **515,** 568–571 (2014).
25. McLaren, W. *et al.* The Ensembl Variant Effect Predictor. *Genome Biol.* **17,** 122 (2016).
26. Deshwar, A. G. *et al.* PhyloWGS: reconstructing subclonal composition and evolution from whole-genome sequencing of tumors. *Genome Biol.* **16,** 35 (2015).
27. Dharma Rao, T. *et al.* Novel monoclonal antibodies against the proximal (carboxy-terminal) portions of MUC16. *Appl. Immunohistochem. Mol. Morphol.* **18,** 462–472 (2010).





b Clinicopathologic characteristics of patients in tissue microarray cohort.

Variable	Short Term (*n* = 45) *n* (%)	Long Term (*n* = 51) n (%)	P- value
Gender			
Male	24 (53)	26 (51)	0.84
Female	21 (47)	25 (49)	
Age			
Median (Range)	78 (54-91)	74 (38-95)	0.33
Tumor Location			
Head	32 (71)	37 (73)	1
Body/Tail	13 (29)	14 (27)	
Procedure			
Distal Pancreatectomy	13 (29)	13 (25)	0.93
Pancreaticoduodenectomy	31 (69)	37 (73)	
Total Pancreatectomy	1 (2)	1 (2)	
Pathological Stage			
I	0 (0)	2 (4)	0.09
II	38 (84)	47 (92)	
III	4 (9)	2 (4)	
IV*	3 (7)	0 (0)	
pT			
1	0 (0)	0 (0)	0.35
2	0 (0)	2 (4)	
3	42 (96)	47 (92)	
4	3 (4)	2 (4)	
pN			
0	11 (24)	21 (41)	0.13
1	34 (76)	30 (59)	
pM			
0	42 (93)	51 (100)	1
1*	3 (7)	0 (0)	
Margin			
Positive	8 (18)	2 (4)	0.04
Negative	37 (82)	49 (96)	
Adjuvant Treatment			
Yes	24 (53)	42 (82)	0.008
No	20 (45)	9 (18)	
Unknown	1 (2)	0 (0)	

c Clinicopathologic characteristics of patients in transcriptome profiling cohort.

Variable	Short Term (*n* = 15) *n* (%)	Long Term (*n* = 15) *n* (%)	P- value
Gender			
Male	8 (53)	5 (33)	0.46
Female	7 (47)	10 (67)	
Age			
Median (Range)	76 (54-84)	65 (51-95)	0.20
Tumor Location			
Head	10 (67)	12 (80)	0.69
Body/Tail	5 (33)	3 (20)	
Procedure			
Distal Pancreatectomy	5 (33)	3 (20)	0.69
Pancreaticoduodenectomy	10 (67)	12 (80)	
Pathological Stage			
I	0 (0)	0 (0)	0.34
II	12 (80)	14 (93)	
III	1 (7)	1 (7)	
IV*	2 (13)	0 (0)	
pT			
1	0 (0)	0 (0)	1
2	0 (0)	0 (0)	
3	13 (87)	14 (93)	
4	2 (13)	1 (7)	
pN			
0	5 (33)	6 (40)	1
1	10 (67)	9 (60)	
pM			
0	13 (87)	15 (100)	0.48
1*	2 (13)	0 (0)	
Margin			
Positive	4 (27)	1 (7)	0.33
Negative	11 (73)	14 (93)	
Adjuvant Treatment			
Yes	10 (67)	13 (87)	0.36
No	4 (26)	2 (13)	
Unknown	1 (7)	0 (0)	

d Clinicopathologic characteristics of patients in TCR sequencing cohort.

Variable	Short Term (*n* = 30) *n* (%)	Long Term (*n* = 30) *n* (%)	P- value
Gender			
Male	17 (57)	12 (40)	0.30
Female	13 (43)	18 (60)	
Age			
Median (Range)	73 (45-91)	75 (54-95)	0.28
Tumor Location			
Head	17 (57)	23 (77)	0.17
Body/Tail	13 (43)	7 (23)	
Procedure			
Distal Pancreatectomy	13 (43)	7 (23)	0.17
Pancreaticoduodenectomy	17 (57)	23 (77)	
Pathological Stage			
I	0 (0)	2 (7)	0.17
II	26 (87)	27 (90)	
III	1 (3)	1 (3)	
IV*	3 (10)	0 (0)	
pT			
1	0 (0)	2 (7)	0.31
2	0 (0)	0 (0)	
3	28 (93)	27 (90)	
4	2 (7)	1 (3)	
pN			
0	14 (47)	17 (57)	0.61
1	16 (53)	13 (43)	
pM			
0	28 (93)	30 (100)	0.49
1*	2 (7)	0 (0)	
Margin			
Positive	1 (3)	2 (7)	1
Negative	29 (97)	28 (93)	
Adjuvant Treatment			
Yes	18 (60)	26 (86)	0.04
No	9 (30)	2 (7)	
Unknown	3 (10)	2 (7)	

e Clinicopathologic characteristics of patients in whole exome sequencing cohort.

Variable	Short Term (*n* = 32) *n* (%)	Long Term (*n* = 26) *n* (%)	P- value
Gender			
Male	15 (47)	8 (31)	0.28
Female	17 (53)	18 (69)	
Age			
Median (Range)	73 (48-91)	75 (51-95)	1
Tumor Location			
Head	18 (56)	21 (81)	0.06
Body/Tail	14 (44)	5 (19)	
Procedure			
Distal Pancreatectomy	13 (41)	5 (19)	1
Pancreaticoduodenectomy	19 (59)	21 (81)	
Pathological Stage			
I	0 (0)	1 (4)	0.29
II	28 (88)	24 (92)	
III	1 (3)	1 (4)	
IV*	3 (9)	0 (0)	
pT			
1	0 (0)	1 (4)	0.5
2	0 (0)	0 (0)	
3	29 (91)	24 (92)	
4	3 (9)	1 (4)	
pN			
0	9 (28)	11 (42)	0.28
1	23 (72)	15 (58)	
pM			
0	30 (94)	26 (100)	0.25
1*	2 (6)	0 (0)	
Margin			
Positive	5 (16)	3 (12)	0.72
Negative	27 (84)	23 (88)	
Adjuvant Treatment			
Yes	22 (63)	24 (92)	0.08
No	9 (28)	2 (8)	
Unknown	1 (9)	0 (0)	

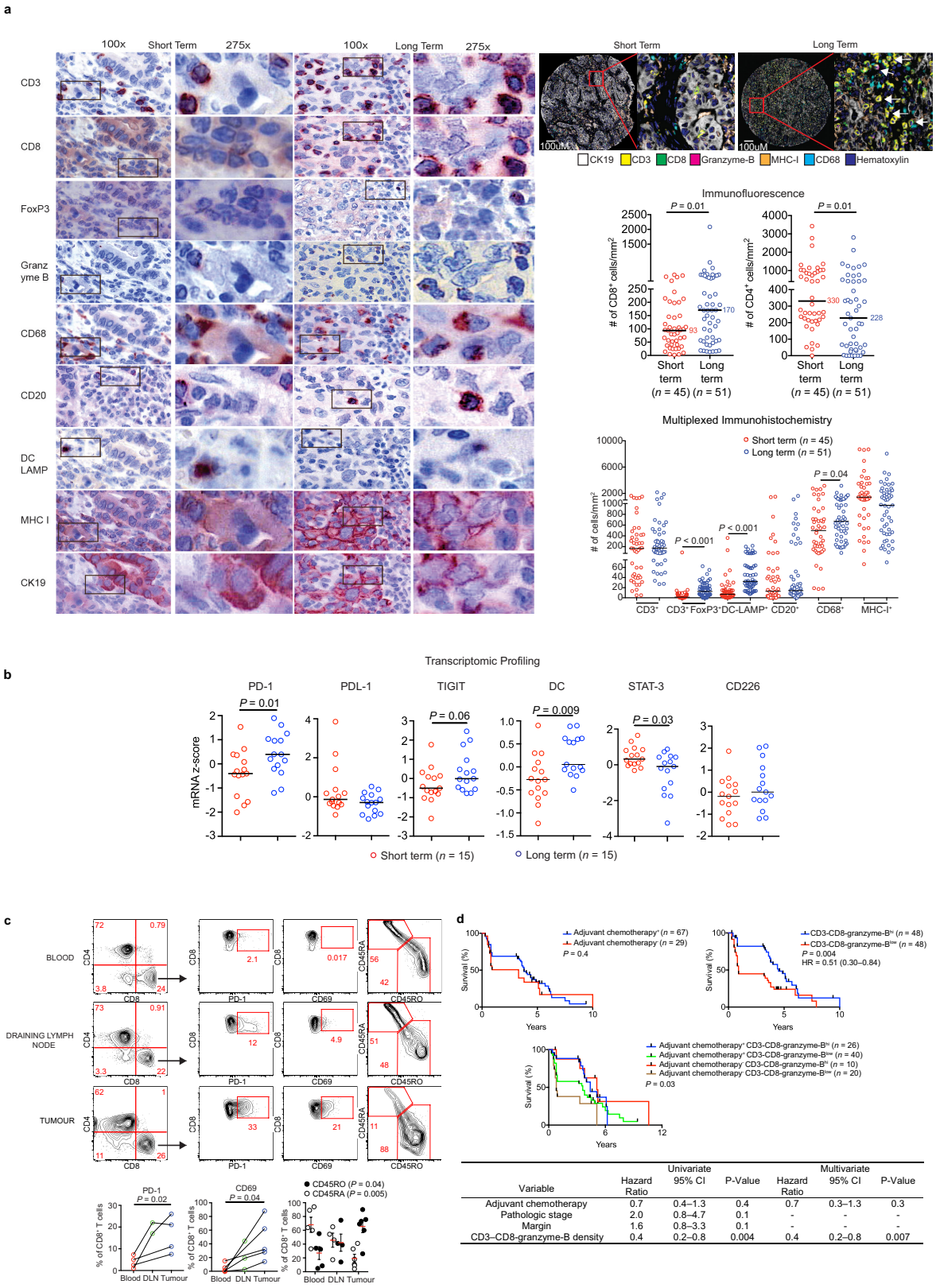
f Clinicopathologic characteristics of patients in matched primary and metastatic tumor cohort.

Variable	MUC16 Neoantigenic (*n* = 2), n (%)	Non- MUC16 Neoantigentic (*n* = 2), n (%)	P- value
Gender			
Male	0 (0)	1 (50)	1
Female	2 (100)	1 (50)	
Age			
Median (Range)	71 (57-85)	52 (50-54)	0.31
Pathological Stage			
I	0 (0)	0 (0)	1
II	0 (0)	0 (0)	
III	1 (50)	0 (0)	
IV	1 (50)	2 (100)	
Chemotherapy			
Yes	2 (100)	1 (50)	1
No	0 (0)	1 (50)	

g Clinicopathologic characteristics of very long term pancreatic cancer survivors.

	Long Term (*n* = 7) *n* (%)	
Gender		
Male	1 (14)	
Female	6 (86)	
Age		
Median (Range)	73 (60-88)	
Tumor Location		
Head	4 (57)	
Body/Tail	3 (43)	
Procedure		
Distal Pancreatectomy	3 (43)	
Pancreaticoduodenectomy	4 (57)	
Pathological Stage		
I	1 (14)	
II	6 (86)	
III	0 (0)	
IV	0 (0)	
pT		
1	1 (14)	
2	0 (0)	
3	6 (86)	
4	0 (0)	
pN		
0	5 (71)	
1	2 (29)	
pM		
0	7 (100)	
1	0 (0)	
Margin		
Positive	1 (14)	
Negative	6 (86)	
Adjuvant Treatment		
Yes	7 (100)	
No	0 (0)	
Unknown	0 (0)	
Recurrence and Survival	Recurrence	Survival (years)
Patient 1	No	10.5
Patient 2	No	9.5
Patient 3	No	11.5
Patient 4	No	11.8
Patient 5	Yes	7.3
Patient 6	Yes	11.8
Patient 7	No	7.7

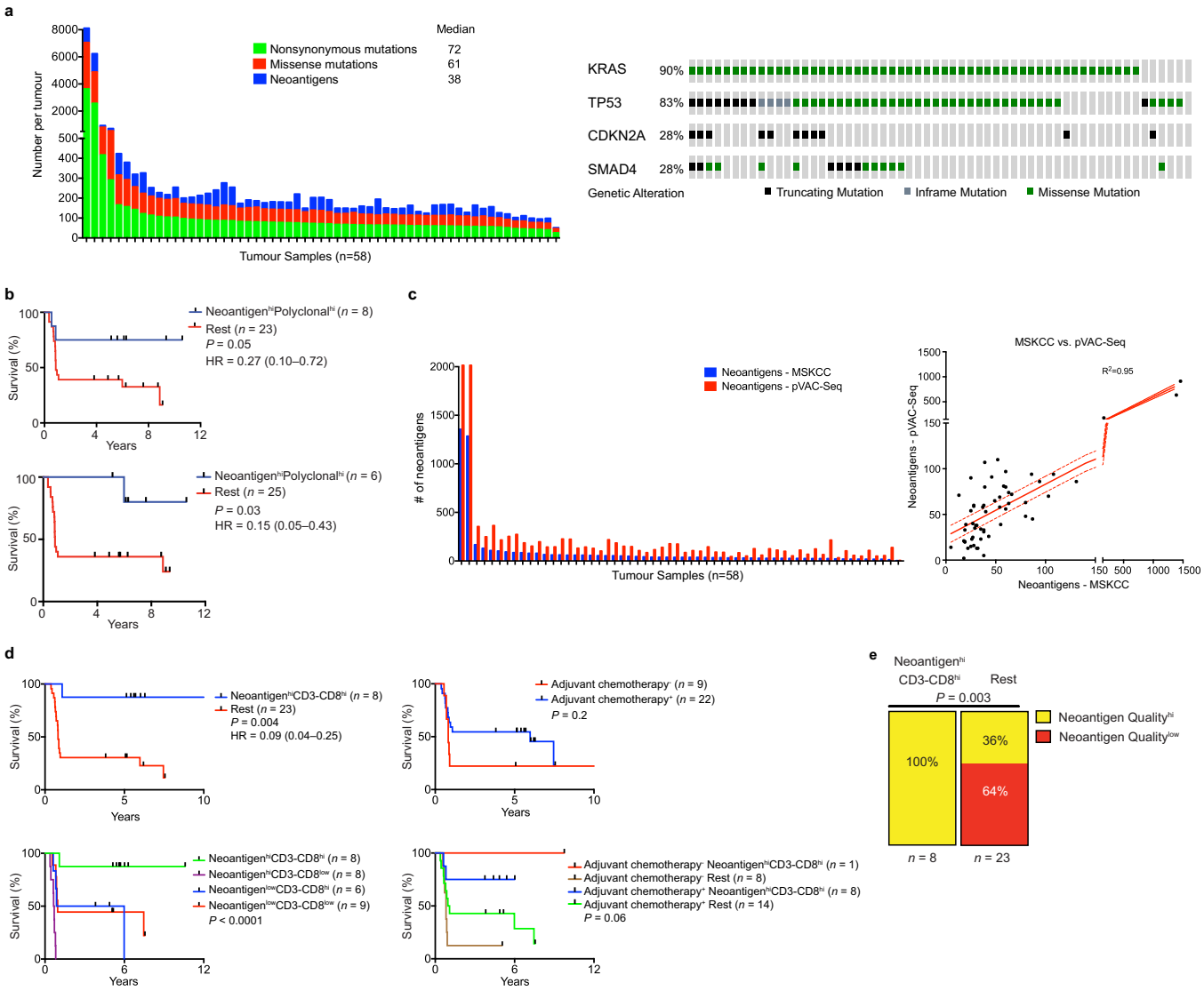
Extended Data Figure 1 | Clinicopathological characteristics of the MSKCC cohort. a, Overall survival and patient overlap of short- and long-term survivors in tissue microarray, whole-exome sequencing, TCR Vβ sequencing, and bulk tumour transcriptomic profiling cohorts. **b**–**g**, Clinicopathological characteristics of patients in tissue microarray, transcriptome, TCR sequencing, whole-exome sequencing, matched primary–metastatic, and very long-term survivor cohorts. In **b**–**e**, the asterisk indicates three patients with metastases noted on final pathology (one liver metastasis, one metastasis to small bowel/mesentery, one splenic metastasis). *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test (**a**) and a two-sided Fisher's exact test (gender, tumour location, pN, pM, margin, chemotherapy), a two-sided χ^2 test (procedure, pathological stage, pT, adjuvant treatment) and an unpaired, two-tailed Student's *t*-test (age) (**b**–**g**).



Variable	Univariate Hazard Ratio	Univariate 95% CI	Univariate P-Value	Multivariate Hazard Ratio	Multivariate 95% CI	Multivariate P-Value
Adjuvant chemotherapy	0.7	0.4–1.3	0.4	0.7	0.3–1.3	0.3
Pathologic stage	2.0	0.8–4.7	0.1	-	-	-
Margin	1.6	0.8–3.3	0.1	-	-	-
CD3–CD8-granzyme-B density	0.4	0.2–0.8	0.004	0.4	0.2–0.8	0.007

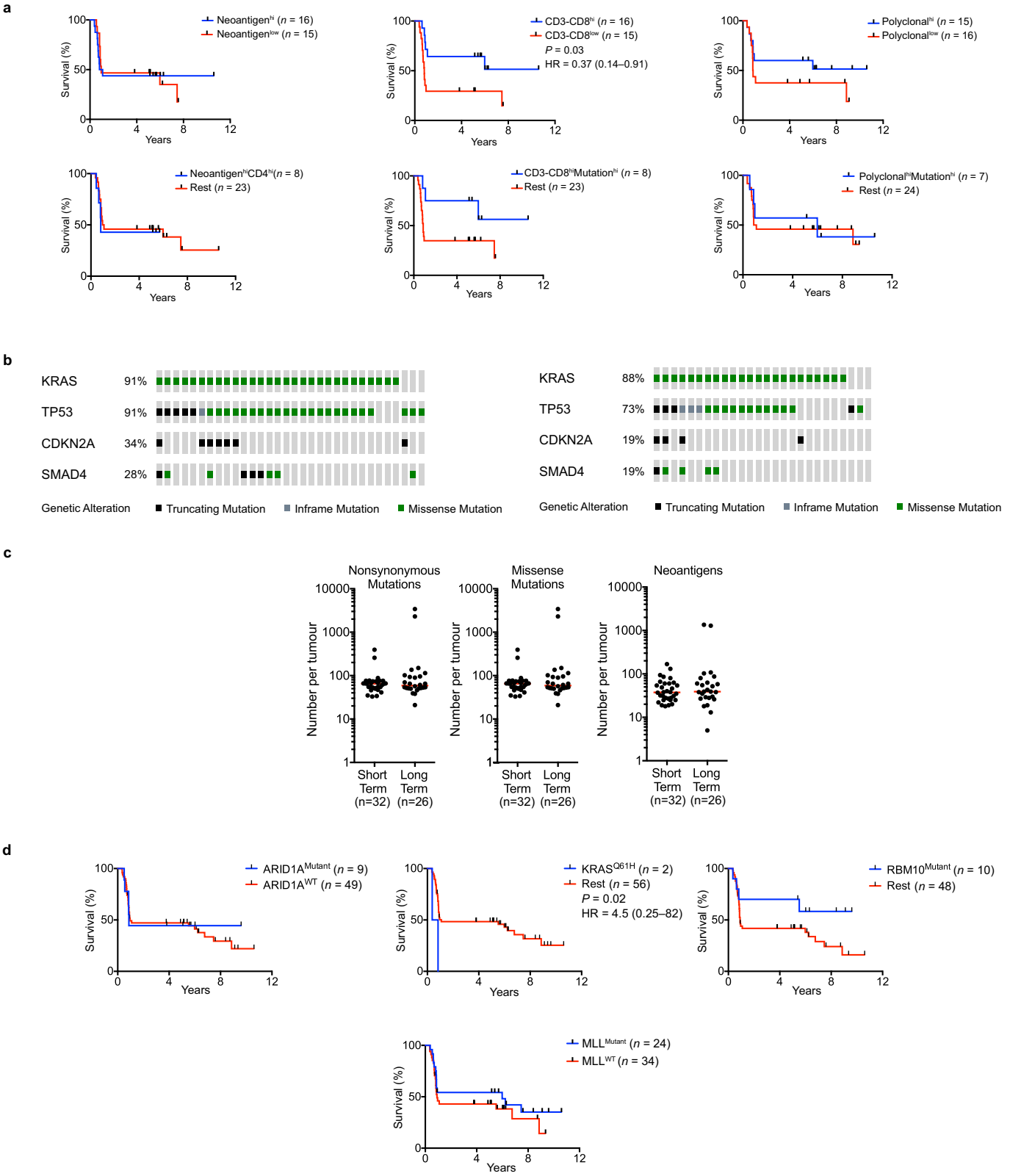
Extended Data Figure 2 | See next page for caption.

Extended Data Figure 2 | Long-term survivors of PDAC display enhanced intratumoural T-cell immunity. a, Left: representative sequential immunohistochemical staining of a single short-term and a single long-term core tumour section. Sections bounded by black rectangles in the 100× panels are magnified to 275× in the panels immediately to the right for each core section. Top right: representative merged images of multiplexed immunohistochemistry are shown. The red rectangular sections are enlarged to 50×. CK19 was used to stain tumour cells. The arrows indicate $CD3^+CD8^+$granzyme B^+ T cells. Middle right: immunofluorescent quantification of $CD8^+$ and $CD4^+$ cells in tumour tissue microarrays of short- and long-term survivors. Slides used were cut from separate sections of the block as those used for sequential immunohistochemistry (in **a** and Fig. 1b). Bottom right: quantification of multiplexed immunophenotyping as shown in **a** (left and top right). All immunofluorescent and immunohistochemical staining was repeated independently in triplicate for each patient. In **a**, short term $n = 45$, long term $n = 51$. **b**, Bulk tumour transcriptomic immune profiling in short- and long-term survivors. Dendritic-cell signature genes include *CCL13*, *CCL17*, *CCL22*, *PPFIBP2*, *NPR1*, *HSD11B1* and *CD209* (also known as *DC-SIGN*)[3]. **c**, Flow-cytometric gating strategy to phenotype human T cells ($n = 7$). The first plot is pre-gated on live cells, followed by $CD45^+$ and $CD3^+CD56^-$ cells. Values indicate the percentage of cells within the red boxes, and are gated based on isotype controls. **d**, Top: overall survival of patients who did or did not receive adjuvant chemotherapy (adjuvant chemotherapy$^{+/-}$ respectively, top left), and of patients with tumours harbouring more or fewer than the median number of CD3-CD8-granzyme B triple-positive cells (CD3-CD8-granzyme $B^{hi/low}$ respectively, top right). Overall survival of all four groups is shown in the bottom panels. The table shows univariate and multivariate Cox regression analysis of clinicopathological features, adjuvant chemotherapy, and CD3-CD8-granzyme B density associations with overall survival. Horizontal bars indicate median values, error bars represent the s.e.m. n is the number of biologically independent samples in individual patients. P values were determined using a two-tailed Mann–Whitney U-test (**a**, **b**), a one-way ANOVA (**c**) and a log-rank test (**d**).

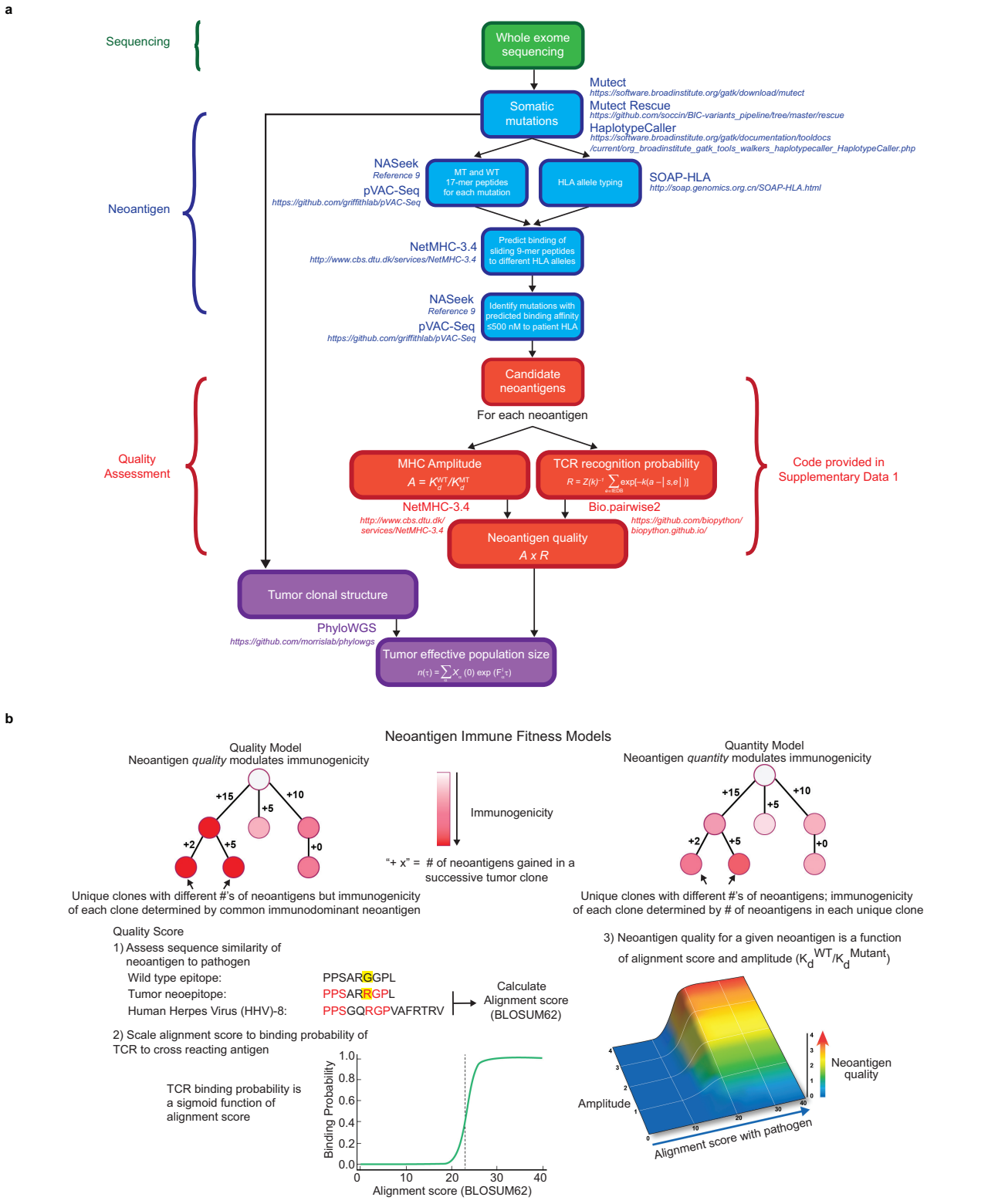


Variable	Univariate Hazard Ratio	Univariate 95% CI	Univariate P-Value	Multivariate Hazard Ratio	Multivariate 95% CI	Multivariate P-Value
Adjuvant chemotherapy	0.5	0.2–1.4	0.2	0.6	0.2–1.7	0.4
Pathologic stage	1.3	0.3–4.6	0.6	-	-	-
Margin	2.2	0.7–6.3	0.1	-	-	-
Neoantigen CD3–CD8 density	0.09	0.01–0.7	0.02	0.09	0.01–0.7	0.02

Extended Data Figure 3 | Neoantigen quantity and CD8$^+$ T-cell infiltrate identify long-term pancreatic cancer survivors. a, Left: number of nonsynonymous, missense and neoantigenic mutations per patient in the MSKCC cohort. The tick marks on the *x* axis correspond to individual tumours. Right: oncoprint demonstrating the frequency of oncogenic driver-gene mutations in the MSKCC cohort. **b**, Overall survival of patients with tumours harbouring more than the median number of neoantigens (neoantigen^{hi}), and greater than the median intratumoural T-cell repertoire polyclonality (polyclonal^{hi}), compared to all other patients (rest). Neoantigens were determined using the MSKCC (top) and the pVAC-Seq (bottom) neoantigen prediction pipelines. **c**, Left: number of neoantigens per tumour, as determined by the MSKCC and pVAC-Seq neoantigen calling pipelines. Tick marks on the *x* axis correspond to individual tumours. Right: correlation matrix of neoantigens as determined by the MSKCC and pVAC-Seq neoantigen calling pipelines. The solid red line indicates the line of best fit, dotted lines indicate 95% confidence intervals. **d**, Top: overall survival of patients with tumours harbouring more or fewer than the median number of neoantigens (neoantigen$^{hi/low}$) and CD3-CD8 double-positive cells (CD3-CD8$^{hi/low}$), compared to all other patients (rest) (top left). Patients who did or did not receive adjuvant chemotherapy (adjuvant chemotherapy$^{+/-}$, respectively) (top right), and all four groups (bottom) are also shown. The table shows univariate and multivariate Cox regression analysis of the associations of clinicopathological features, adjuvant chemotherapy, and neoantigen-CD3-CD8 number with overall survival. **e** Distribution of tumours with high- and low-quality neoantigens in neoantigen^{hi} CD3-CD8^{hi} long-term pancreatic cancer survivors compared to all other patients (rest). *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test (**b**, **d**) and a χ^2 test (**e**).

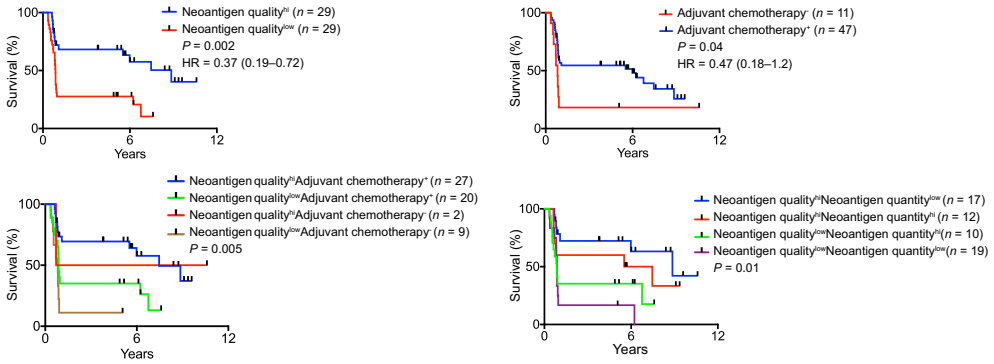


Extended Data Figure 4 | Unique genomic features alone do not identify long-term survivors of PDAC. a, Overall survival of patients with tumours harbouring more or fewer than the median number of neoantigens (neoantigen$^{hi/low}$), CD3-CD8 double-positive cells (CD3-CD8$^{hi/low}$), polyclonality (polyclonal$^{hi/low}$), mutations (mutation$^{hi/low}$), and CD4 single positive cells (CD4$^{hi/low}$). **b**, Oncoprint demonstrating no difference in the frequency of oncogenic driver mutations in short- and long-term tumours. **c**, No difference was found in the number of nonsynonymous, missense, and immunogenic mutations (neoantigens) in short- and long-term PDAC tumours. **d**, Overall survival stratified by mutations in *ARID1A*, *KRAS*^{Q61H}, *RBM10* and MLL-related genes (*KMT2A*, *KMT2B*, *KMT2C* and *KMT2E* (also known as *MLL*, *MLL2*, *MLL3* and *MLL5*, respectively)). Horizontal bars indicate median values. *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test (**a**, **d**).



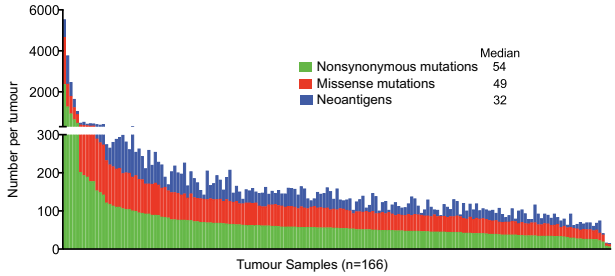
Extended Data Figure 5 | Neoantigen immune fitness models. **a**, Comprehensive flowchart of neoantigen quality identification pipeline. Software programs used for each step are indicated in italicized text. Mathematical formulae for the calculation of individual components of neoantigen quality are defined in the Methods. All software components of the pipeline are published and/or publicly available. **b**, Top: schematic of neoantigen immune fitness models. Each circle represents a tumour clone in an evolutionary tree. Clones in both models are identical with respect to the number of mutations and neoantigens. The numbers represent hypothetical neoantigens gained in a successive tumour clone. Shades of red indicate the immunogenicity of each clone, as ascribed by the two models, namely neoantigen quality or neoantigen quantity. Bottom: parameters defining the quality score in the quality model (1)–(3). In (1), amino acid sequences of a hypothetical wild-type epitope, tumour neoepitope, and a homologous microbial epitope are shown. Yellow highlights the changing amino acid between the wild-type and the tumour sequence as a consequence of a tumour-specific mutation. The amino acids in red indicate homology between the tumour neoepitope and the microbial epitope.

a

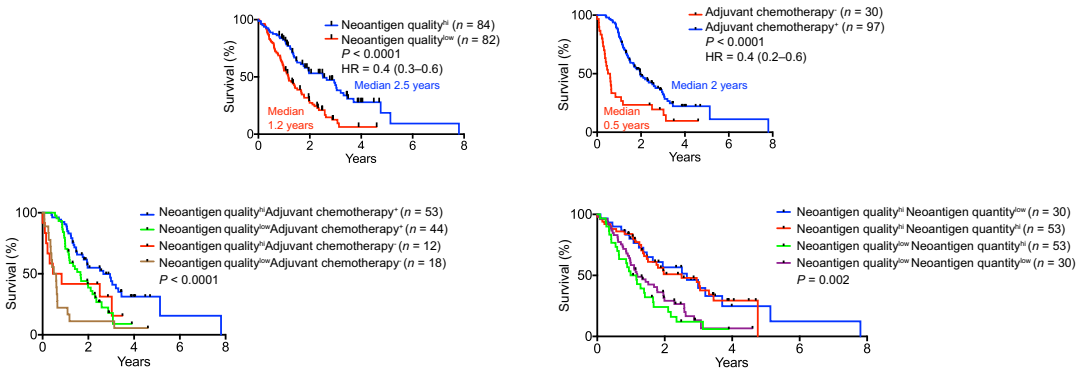


Variable	Univariate			Multivariate		
	Hazard Ratio	95% CI	P-Value	Hazard Ratio	95% CI	P-Value
Adjuvant chemotherapy	0.4	0.1–0.9	0.02	0.5	0.2–1.2	0.1
Pathologic stage	1.4	0.5–4.1	0.4	-	-	-
Margin	1.5	0.7–3.4	0.2	-	-	-
Neoantigen quality	0.3	0.1–0.7	0.003	0.3	0.1–0.7	0.008

b

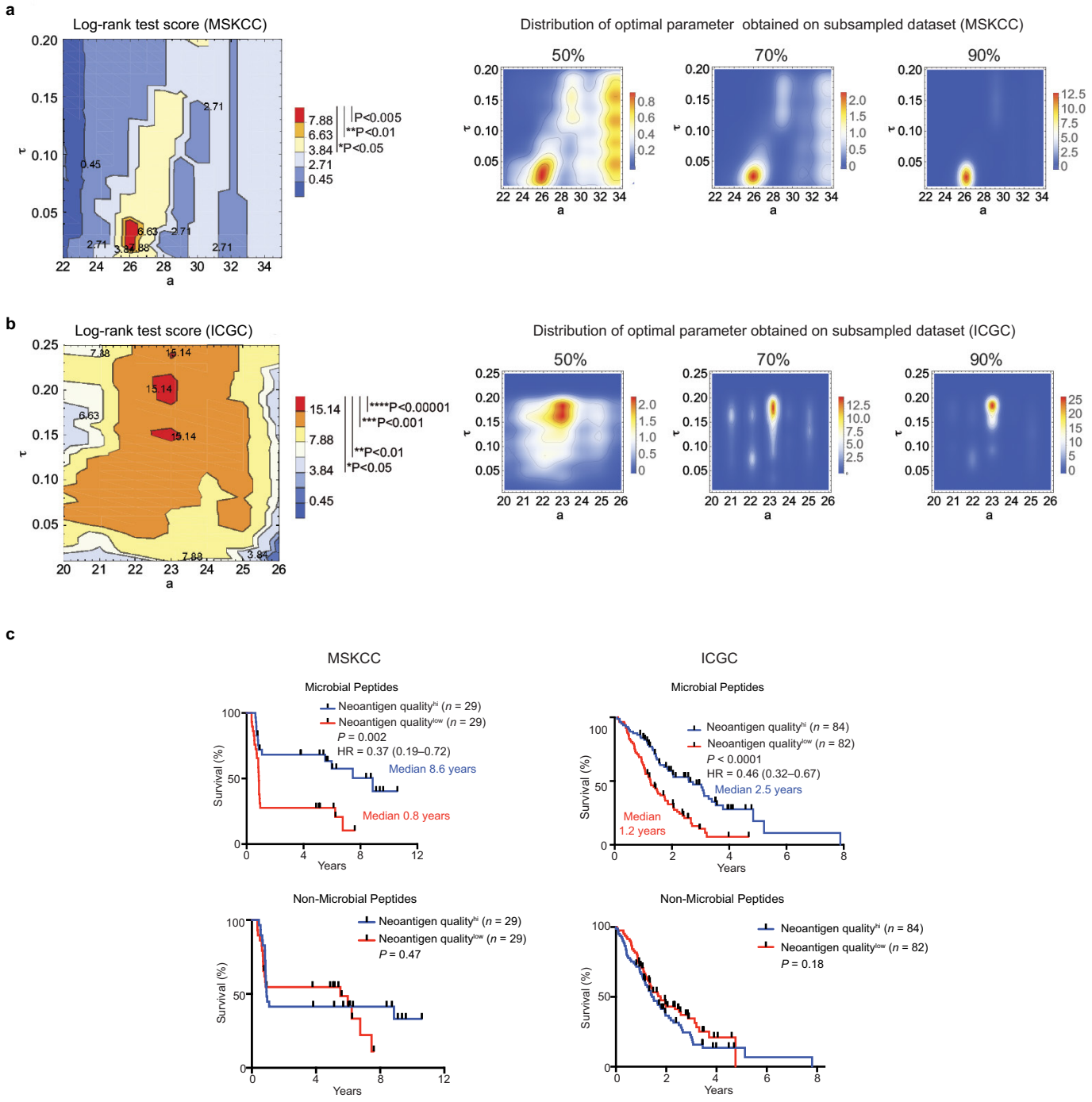


c

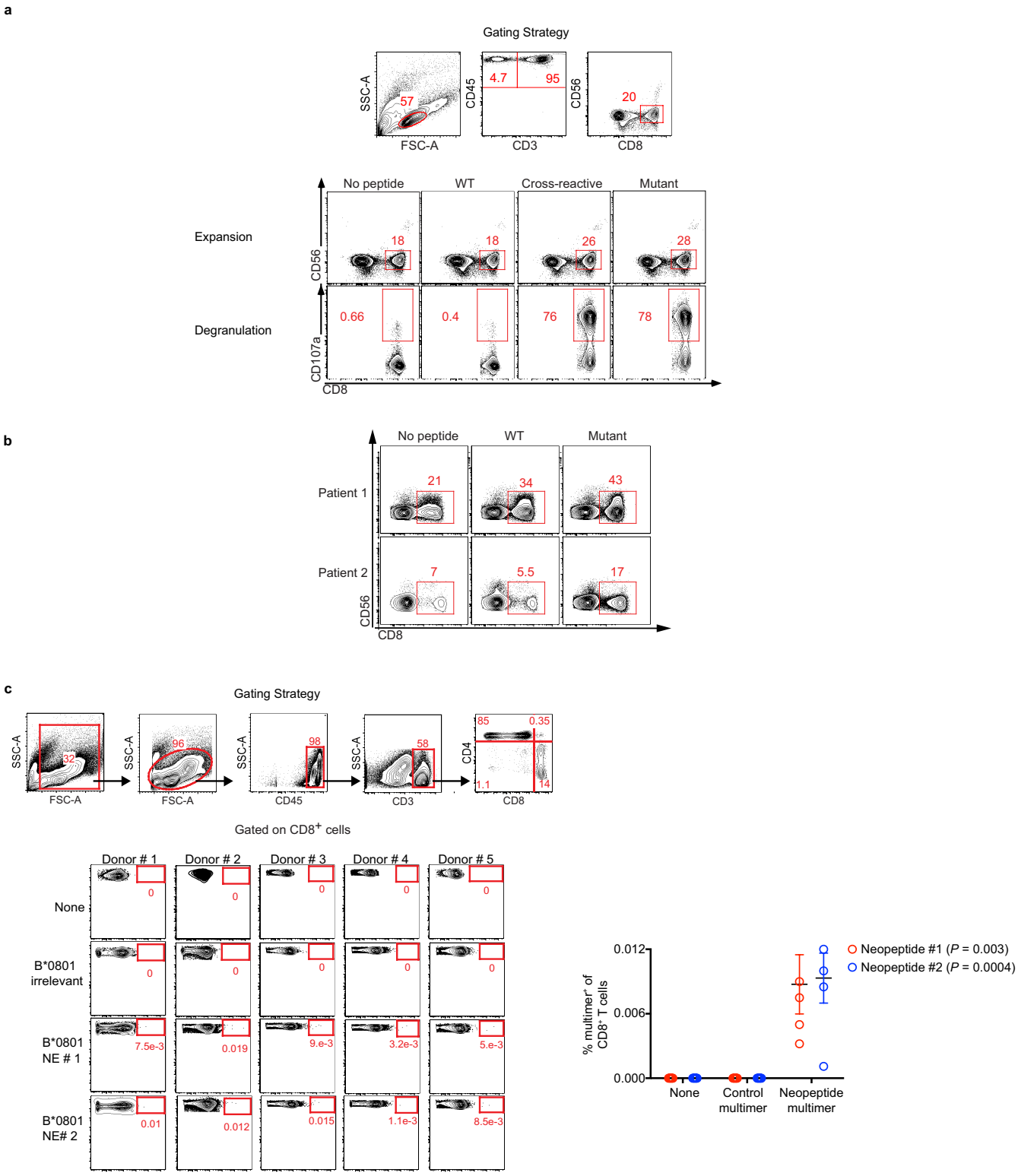


Variable	Univariate			Multivariate		
	Hazard Ratio	95% CI	P-Value	Hazard Ratio	95% CI	P-Value
Adjuvant chemotherapy	0.4	0.2–0.5	0.0001	0.4	0.2–0.7	0.001
Pathologic stage						
-Stage I	-	-	-	-	-	-
-Stage II	2.5	0.9–7	0.07	3.7	0.9–15.9	0.07
Stage III/IV	6.8	1.9–24.8	0.004	12.7	1.1–146	0.04
Margin						
-R0	-	-	-	-	-	-
-R1	2.3	1.4–3.7	0.0001	2.7	1.6–4.5	0.001
-R2	3.6	1.6–8	0.001	0.9	0.09–9.5	0.9
Neoantigen Quality	0.4	0.3–0.64	0.0001	0.5	0.3–0.9	0.01

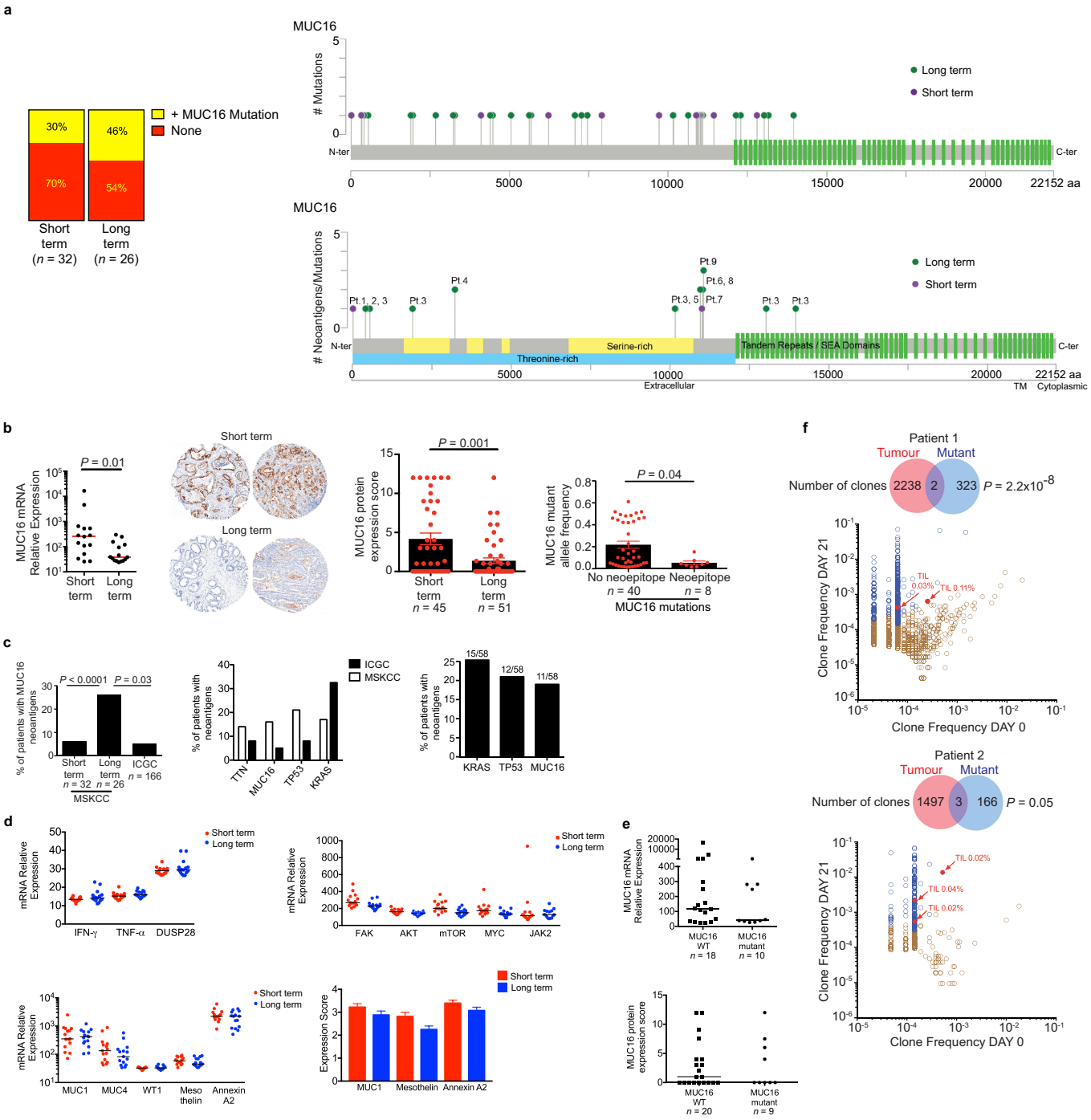
Extended Data Figure 6 | Neoantigen quality is independently prognostic of survival. a, Top: overall survival of patients whose tumours displayed high compared to low neoantigen quality (neoantigen quality$^{hi/low}$) (left), and overall survival of patients who did or did not receive adjuvant chemotherapy (right). Bottom: overall survival of all four groups. Neoantigen quality defined by pipeline and schema as defined in Extended Data Fig. 5a, b. The table shows univariate and multivariate Cox regression analyses of the associations of clinicopathological features, adjuvant chemotherapy and neoantigen quality with overall survival. Data include all patients in the whole-exome sequencing MSKCC cohort. **b**, Number of nonsynonymous, missense and neoantigenic mutations per patient in the ICGC cohort ($n = 166$). **c**, Top: overall survival of patients in the ICGC cohort whose tumours displayed high compared to low neoantigen quality (neoantigen quality$^{hi/low}$) (left), and overall survival of patients in the ICGC cohort stratified by adjuvant chemotherapy administration (right). Bottom: overall survival of all four groups. Neoantigen quality defined by pipeline and schema as defined in Extended Data Fig. 5a, b. The table shows univariate and multivariate Cox regression analyses of the associations of clinicopathological features, adjuvant chemotherapy and neoantigen quality with overall survival in the ICGC cohort. Data on adjuvant chemotherapy is included for patients whose treatment status was available. *n* is the number of biologically independent samples in individual patients. *P* values were determined using a log-rank test (**a**, **c**).



Extended Data Figure 7 | Stability of neoantigen quality model parameters on subsampled cohorts and prognostic dependence of neoantigen quality on infectious disease-derived peptides. **a**, **b**, Parameters of the neoantigen fitness quality model for the MSKCC (**a**) and the ICGC (**b**) cohorts. Left: log-rank test score landscape as a function of the model parameters (the horizontal alignment score displacement a, and the characteristic time τ); the significance of the score is denoted in the legend. Right: two-dimensional histograms showing distributions of optimal parameters obtained on subsampled datasets with 50, 70, and 90% of patients left, over 500 iterations of subsampling at each frequency. **c**, Overall survival of patients in the MSKCC and ICGC cohorts; whole tumours displayed high compared to low neoantigen quality (neoantigen quality$^{hi/low}$). Neoantigen quality was calculated using alignment to immunogenic infectious disease-derived IEDB peptides (microbial peptides) or using alignment to immunogenic non-infectious disease-derived allergy or autoimmune peptides in the IEDB database (non-microbial peptides). n is the number of biologically independent samples in individual patients. P values were determined using a log-rank test (**c**).

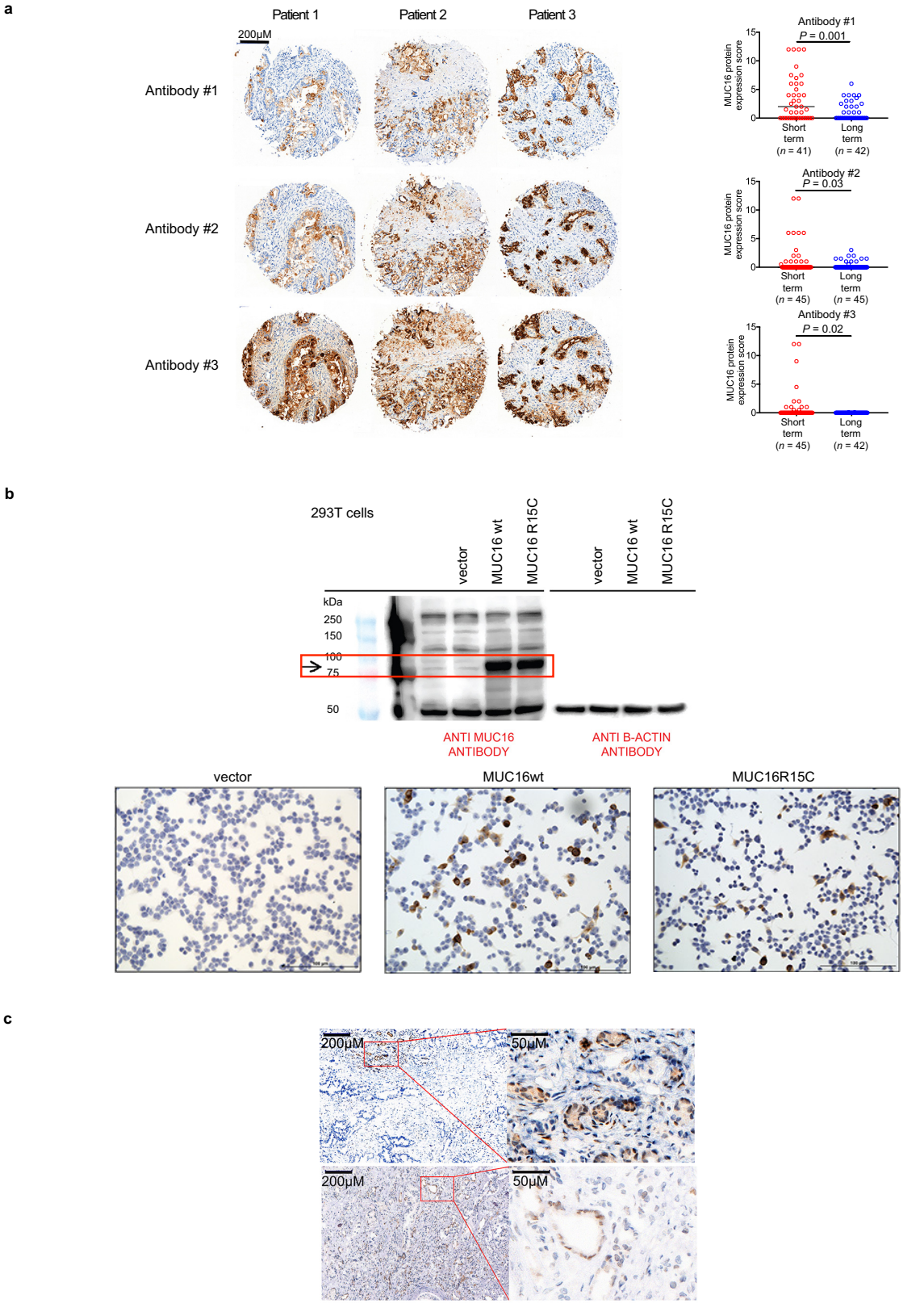


Extended Data Figure 8 | Predicted MUC16 neoantigens are recognized by the human TCR repertoire. a, PBMCs pulsed with no peptide, wild-type control peptide, cross-reactive peptide or high-quality neopeptide (mutant). Representative gating strategies for $CD8^+$ T-cell expansion and degranulation are shown. **b**, PBMCs pulsed with no peptide, wild-type control peptide, and MUC16 neopeptides (mutant). Representative gating strategies for $CD8^+$ T-cell expansion are shown. **c**, Representative gating strategy to identify $CD8^+$ T cells in peripheral blood of healthy donors (top panel). Identification of $CD8^+$ T cells in healthy donors reactive to unique MUC16 neoepitopes predicted to bind to the B*0801 HLA-allele, using MUC16-neoepitope–HLA multimers. Quantification of all healthy donors (neoepitope no. 1, 2, $n = 5$) is shown (right). Multimer staining is shown on the x axis, CD8 is shown on the y axis. Peptide information is provided in Supplementary Table 3. n is the number of biologically independent samples in individual patients. Horizontal bars indicate median values, error bars represent the s.e.m. P values were determined using a one-way ANOVA (**c**).



Extended Data Figure 9 | See next page for caption.

Extended Data Figure 9 | Long-term survivors do not display differences in *MUC16* mutations, transcriptional regulators or downstream targets of MUC16, or differences in other mucins and tissue expression antigens. a, The frequency of *MUC16* mutations in short- and long-term PDAC tumours. Lollipop plot showing the location of *MUC16* mutations and neoantigens in short- and long-term survivors of pancreatic cancer. **b**, Expression of bulk tumour *MUC16* mRNA (left) and protein (middle) by immunohistochemistry. Immunohistochemical staining was repeated independently in triplicate for each patient. Right: *MUC16* mutant allele frequency in non-hypermutated tumours with *MUC16* mutations. **c**, Left: frequency of patients with MUC16 neoantigens in MSKCC and ICGC cohorts. Middle: frequency of patients with neoantigens in genes recurrently harbouring neoantigens in >5% of patients in both MSKCC and ICGC cohorts. Right: genes most frequently harbouring neoantigens in the MSKCC cohort as determined by pVAC-Seq. Frequency of patients (*y* axis) and raw numbers (above bar graphs) are indicated. **d**, mRNA expression of transcriptional activators of *MUC16* (top left), mediators implicated in MUC16-dependent tumour progression (top right), and mRNA (bottom left) and protein (bottom right) levels of tissue expression antigens MUC1, MUC4, WT1, mesothelin and annexin A2 in short- and long-term tumours. WT1 protein was undetectable in both short- and long-term survivors. $n = 15$ per group in top left, top right, and bottom left; short term $n = 45$, long term $n = 51$ in bottom right. **e**, *MUC16* mRNA and protein expression in *MUC16* non-mutated (WT; $n = 18$ (top), $n = 20$ (bottom)) and mutated (mutant; $n = 10$ (top), $n = 9$ (bottom)) tumours. **f**, TCR Vβ sequencing of T-cell product following the pulse of peripheral blood T cells with MUC16 neopeptides as in Fig. 4e. Brown open circles indicate stable or contracted clones with mutant neopeptide; blue open circles indicate expanded clones with mutant neopeptide; red solid circles indicate expanded clones with mutant neopeptide detected in archival primary tumours. Arrows indicate clones in archival primary tumours with rank frequencies; Venn diagrams show clonal overlap in respective compartments. Horizontal bars indicate median values, error bars represent the s.e.m. *n* is the number of biologically independent samples in individual patients. *P* values were determined using two-tailed Mann–Whitney *U*- and Student's *t*-tests (**b**), a χ^2 test (**c**) and as described in the Methods (**f**).



Extended Data Figure 10 | See next page for caption.

Extended Data Figure 10 | *MUC16* mutations do not alter tumour cell-intrinsic MUC16 protein expression. a, Representative immunohistochemical staining (left) and quantification (right) of MUC16 expression in tissue microarrays of short- and long-term survivors of pancreatic cancer, as assessed using three independent anti-MUC16 antibodies. Ab no. 1: clone EPSISR23, purchased from Abcam; Ab no. 2: polyclonal, purchased from Abcam ab133419; Ab no. 3: clone 4H11[27]. Each open circle represents the median expression of independent immunohistochemical staining performed in triplicate for each patient. **b**, Western blot (top) and immunocytochemistry (bottom) of untransfected (−), empty vector (vector), MUC16 wild-type (MUC16 WT) and MUC16 mutant (MUC16 R15C) HEK293 cells. The left blot was probed with anti-MUC16-specific antibody (clone 4H11[27]) and the right blot with anti-β-actin. The red rectangle indicates the MUC16-specific band. All cells in the bottom panels were probed with anti-MUC16 antibody (clone 4H11[27]). The inserted mutation was identical to a neoantigenic MUC16 mutation (detected in patient 1 shown in Extended Data Fig. 9a). Data are representative of two independent experiments with similar results. **c**, MUC16 immunohistochemistry on samples from two long-term survivors of pancreatic cancer with MUC16 neoepitopes in primary resected tumours. Areas in rectangular low-power fields are shown magnified in the right panels. Immunohistochemical staining was performed independently in triplicate for each patient in tissue microarrays, and confirmed with immunohistochemical staining on whole tumour sections (shown). Horizontal bars indicate median values, error bars represent the s.e.m. *n* is the number of biologically independent samples in individual patients. *P* values were determined using a two-tailed Student's *t*-test (**a**).