ARTICLE

doi:10.1038/nature16965

Genomic analyses identify molecular subtypes of pancreatic cancer

Peter Bailey[1,2], David K. Chang[2,3,4,5], Katia Nones[1,6], Amber L. Johns[3], Ann-Marie Patch[1,6], Marie-Claude Gingras[7,8,9], David K. Miller[1,3], Angelika N. Christ[1], Tim J. C. Bruxner[1], Michael C. Quinn[1,6], Craig Nourse[1,2], L. Charles Murtaugh[10], Ivon Harliwong[1], Senel Idrisoglu[1], Suzanne Manning[1], Ehsan Nourbakhsh[1], Shivangi Wani[1,6], Lynn Fink[1], Oliver Holmes[1,6], Venessa Chin[3], Matthew J. Anderson[1], Stephen Kazakoff[1,6], Conrad Leonard[1,6], Felicity Newell[1], Nick Waddell[1], Scott Wood[1,6], Qinying Xu[1,6], Peter J. Wilson[1], Nicole Cloonan[1,6], Karin S. Kassahn[1,11,12], Darrin Taylor[1], Kelly Quek[1], Alan Robertson[1], Lorena Pantano[13], Laura Mincarelli[2], Luis N. Sanchez[2], Lisa Evers[2], Jianmin Wu[3], Mark Pinese[3], Mark J. Cowley[3], Marc D. Jones[2,3], Emily K. Colvin[3], Adnan M. Nagrial[3], Emily S. Humphrey[3], Lorraine A. Chantrill[3,14], Amanda Mawson[3], Jeremy Humphris[3], Angela Chou[3,15], Marina Pajic[3,16], Christopher J. Scarlett[3,17], Andreia V. Pinho[3], Marc Giry-Laterriere[3], Ilse Rooman[3], Jaswinder S. Samra[18,19], James G. Kench[3,19,20], Jessica A. Lovell[3], Neil D. Merrett[5,21], Christopher W. Toon[3], Krishna Epari[22], Nam Q. Nguyen[23], Andrew Barbour[24], Nikolajs Zeps[25], Kim Moran-Jones[2], Nigel B. Jamieson[2,26,27], Janet S. Graham[2,28], Fraser Duthie[29], Karin Oien[3,29], Jane Hair[30], Robert Grützmann[31]†, Anirban Maitra[32], Christine A. Iacobuzio-Donahue[33], Christopher L. Wolfgang[34,35], Richard A. Morgan[34], Rita T. Lawlor[36,37], Vincenzo Corbo[36], Claudio Bassi[38], Borislav Rusev[36], Paola Capelli[37], Roberto Salvia[38], Giampaolo Tortora[39], Debabrata Mukhopadhyay[40], Gloria M. Petersen[40], Australian Pancreatic Cancer Genome Initiative*, Donna M. Munzy[7,8], William E. Fisher[41], Saadia A. Karim[42], James R. Eshleman[34], Ralph H. Hruban[34], Christian Pilarsky[31], Jennifer P. Morton[42], Owen J. Sansom[42,43], Aldo Scarpa[36,37], Elizabeth A. Musgrove[2], Ulla-Maja Hagbo Bailey[2], Oliver Hofmann[2,13], Robert L. Sutherland[3]‡, David A. Wheeler[7,8], Anthony J. Gill[3,19], Richard A. Gibbs[7,8], John V. Pearson[1,6], Nicola Waddell[1,6], Andrew V. Biankin[2,3,4,5,27] & Sean M. Grimmond[1,2,44]

Integrated genomic analysis of 456 pancreatic ductal adenocarcinomas identified 32 recurrently mutated genes that aggregate into 10 pathways: *KRAS*, *TGF-β*, *WNT*, *NOTCH*, *ROBO/SLIT* signalling, *G1/S* transition, *SWI-SNF*, chromatin modification, DNA repair and RNA processing. Expression analysis defined 4 subtypes: (1) squamous; (2) pancreatic progenitor; (3) immunogenic; and (4) aberrantly differentiated endocrine exocrine (ADEX) that correlate with histopathological characteristics. Squamous tumours are enriched for *TP53* and *KDM6A* mutations, upregulation of the *TP63ΔN* transcriptional network, hypermethylation of pancreatic endodermal cell-fate determining genes and have a poor prognosis. Pancreatic progenitor tumours preferentially express genes involved in early pancreatic development (*FOXA2/3*, *PDX1* and *MNX1*). ADEX tumours displayed upregulation of genes that regulate networks involved in *KRAS* activation, exocrine (*NR5A2* and *RBPJL*), and endocrine differentiation (*NEUROD1* and *NKX2-2*). Immunogenic tumours contained upregulated immune networks including pathways involved in acquired immune suppression. These data infer differences in the molecular evolution of pancreatic cancer subtypes and identify opportunities for therapeutic development.

[1]Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. [2]Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1BD, UK. [3]The Kinghorn Cancer Centre, 370 Victoria St, Darlinghurst, and the Cancer Research Program, Garvan Institute of Medical Research, 384 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia. [4]Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia. [5]South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales 2170, Australia. [6]QIMR Berghofer Medical Research Institute, Herston, Queensland 4006, Australia. [7]Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, Houston, Texas 77030, USA. [8]Michael DeBakey Department of Surgery, Baylor College of Medicine, Houston, Texas 77030, USA. [9]Dan L. Duncan Cancer Center, Baylor College of Medicine, Houston, Texas 77030, USA. [10]Department of Human Genetics, University of Utah, Salt Lake City, Utah 84112, USA. [11]Genetic and Molecular Pathology, SA Pathology, Adelaide, South Australia 5000, Australia. [12]School of Biological Sciences, The University of Adelaide, Adelaide, South Australia 5000, Australia. [13]Harvard Chan Bioinformatics Core, Harvard T. H. Chan School of Public Health, Boston, Massachusetts 02115, USA. [14]Macarthur Cancer Therapy Centre, Campbelltown Hospital, New South Wales 2560, Australia. [15]Department of Pathology. SydPath, St Vincent's Hospital, Sydney, NSW 2010, Australia. [16]St Vincent's Clinical School, Faculty of Medicine, University of New South Wales, New South Wales 2052, Australia. [17]School of Environmental & Life Sciences, University of Newcastle, Ourimbah, New South Wales 2258, Australia. [18]Department of Surgery, Royal North Shore Hospital, St Leonards, Sydney, New South Wales 2065, Australia. [19]University of Sydney, Sydney, New South Wales 2006, Australia. [20]Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown New South Wales 2050, Australia. [21]School of Medicine, University of Western Sydney, Penrith, New South Wales 2175, Australia. [22]Fiona Stanley Hospital, Robin Warren Drive, Murdoch, Western Australia 6150, Australia. [23]Department of Gastroenterology, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia. [24]Department of Surgery, Princess Alexandra Hospital, Ipswich Rd, Woollongabba, Queensland 4102, Australia. [25]School of Surgery M507, University of Western Australia, 35 Stirling Hwy, Nedlands 6009, Australia and St John of God Pathology, 12 Salvado Rd, Subiaco, Western Australia 6008, Australia. [26]Academic Unit of Surgery, School of Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow Royal Infirmary, Glasgow G4 OSF, UK. [27]West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow G31 2ER, UK. [28]Department of Medical Oncology, Beatson West of Scotland Cancer Centre, 1053 Great Western Road, Glasgow G12 0YN, UK. [29]Department of Pathology, Southern General Hospital, Greater Glasgow & Clyde NHS, Glasgow G51 4TF, UK. [30]GGC Bio-repository, Pathology Department, Southern General Hospital, 1345 Govan Road, Glasgow G51 4TY, UK. [31]Department of Surgery, TU Dresden, Fetscherstr. 74, 01307 Dresden, Germany. [32]Departments of Pathology and Translational Molecular Pathology, UT MD Anderson Cancer Center, Houston Texas 77030, USA. [33]The David M. Rubenstein Pancreatic Cancer Research Center and Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York 10065, USA. [34]Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. [35]Department of Surgery, The Sol Goldman Pancreatic Cancer Research Center, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21231, USA. [36]ARC-Net Applied Research on Cancer Centre, University and Hospital Trust of Verona, Verona 37134, Italy. [37]Department of Pathology and Diagnostics, University of Verona, Verona 37134, Italy. [38]Department of Surgery, Pancreas Institute, University and Hospital Trust of Verona, Verona 37134, Italy. [39]Department of Medical Oncology, Comprehensive Cancer Centre, University and Hospital Trust of Verona, Verona 37134, Italy. [40]Mayo Clinic, Rochester, Minnesota 55905, USA. [41]Elkins Pancreas Center, Baylor College of Medicine, One Baylor Plaza, MS226, Houston, Texas 77030-3411, USA. [42]Cancer Research UK Beatson Institute, Glasgow G61 1BD, UK. [43]Institute for Cancer Science, University of Glasgow, Glasgow G12 8QQ, UK. [44]University of Melbourne, Parkville, Victoria 3010, Australia.

†Present address: Universitätsklinikum Erlangen, Department of Surgery, 91054 Erlangen, Germany.

*Lists of participants and their affiliations appear in the Supplementary Information.

‡Deceased.

Pancreatic cancer (PC) is the fourth leading cause of cancer death in Western societies, and projected to be the second within a decade[1]. It has a median survival measured in months and a five-year survival of <5%. Advances in therapy have only achieved incremental improvements in overall outcome, but can provide notable benefit for undefined subgroups of patients. As a consequence, there is an urgent need to better understand the molecular pathology of PC in order to improve patient selection for current treatment options, and to develop novel therapeutic strategies.

Genomic analyses of pancreatic cancer reveal a complex mutational landscape with four common oncogenic events in well-known cancer genes (*KRAS, TP53, SMAD4* and *CDKN2A*), amongst a milieu of genes mutated at low prevalence. Despite this heterogeneity, oncogenic point mutations of individual genes aggregate into core molecular pathways including DNA damage repair, cell cycle regulation, TGF-β signalling, chromatin regulation and axonal guidance[2–5]. Increasingly sophisticated analyses are revealing biologically important events with clinical significance, including whole-genome sequencing, which sub-classifies PC into 4 subtypes based on the frequency and distribution of structural variation. Those termed unstable due to a large number of structural variants correlate with defects in DNA maintenance and therapeutic responsiveness to platinum based therapies[2]. Aberrations in other features that characterize cancer genomes, including mutational signatures[6], and differential methylation[7] are providing deeper insights into disease pathophysiology.

Here we performed a comprehensive integrated genomic analysis of 456 PCs and their histopathological variants using a combination of whole-genome and deep-exome sequencing, with gene copy number analysis to determine the mutational mechanisms and candidate genomic events important in pancreatic carcinogenesis. RNA expression profiles were used to define four subtypes and the different transcriptional networks that underpin them. These subtypes are associated with distinct histopathological characteristics and differential survival. Genomic and epigenetic features that characterize each subtype infer different mechanisms of molecular evolution.

Mutational landscape of PC

Study participants were recruited and consent for genomic sequencing obtained through the Australian Pancreatic Cancer Genome Initiative (APGI; http://www.pancreaticcancer.net.au) as part of the International Cancer Genome Consortium (ICGC; http://www.icgc.org). The 382 APGI group consisted of participants with primarily treatment-naive resected PC, which were pancreatic ductal adenocarcinoma (PDAC) and its variants (adenosquamous, colloid, PDAC associated with intraductal papillary mucinous neoplasm (IPMN)) and a small number of rare acinar cell carcinomas (Supplementary Table 1). We detected 23,538 high confidence coding mutations[2,8,9], of which, 7,377 were verified using orthogonal approaches (Supplementary Tables 1, 2 and 19). A total of 21,208 high confidence genomic rearrangements were also identified (Supplementary Tables 3 and 4)[2,8]. To maximize the power to define coding driver mutations, 74 previously published PC exomes[3–5] were included to yield a final cohort of 456 tumours. OncodriverFM detected 32 significantly mutated genes (false discovery rate (FDR) ≤ 0.1), 22 of which were also identified by MutsigCV2 ($Q < 0.1$) and/or were supported by HOTNET2 analysis (Methods and Supplementary Table 5). These significantly mutated genes aggregated into 10 molecular mechanisms (Extended Data Fig. 1): with activating mutations of *KRAS* in 92%; disruption of *G1/S* checkpoint machinery (*TP53*, *CDKN2A* and *TP53BP2*) in 78%; *TGF-β* signalling (*SMAD4*, *SMAD3*, *TGFBR1*, *TFGBR2*, *ACVR1B* and *ACVR2A*) in 47%; histone modification (*KDM6A*, *SETD2* and ASCOM complex members *MLL2* and *MLL3*) in 24%; the *SWI/SNF* complex (*ARID1A*, *PBRM1* and *SMARCA4*) in 14%; the BRCA pathway (*BRCA1*, *BRCA2*, *ATM* and *PALB2*: 5% germline, 12% somatic); WNT signalling defects through *RNF43* mutation (5%); and RNA processing genes, *SF3B1*, *U2AF1* and *RBM10* (16%). *RBM10* is implicated in lung cancer[10], where inactivating mutations influence expression of oncogenic isoforms of *NUMB*[11]. *SF3B1* mutations in PC were aggregated at the K700E mutation hotspot common in myelodysplastic syndrome, breast and lung cancer[12] and presents a potential therapeutic target[13]. Mutations in other genes encoding splicing machinery: *SF3A1*, *U2AF2*, *SF1* and *RBM6* were also identified (Extended Data Fig. 2 and Supplementary Table 6).

GISTIC2 identified 50 regions of recurrent gain (43 focal, 7 chromosomal arms) and 73 regions of loss (61 focal, 12 chromosomal arms) (Supplementary Tables 7–9). These regions included known oncogenes *MET*, *NOTCH1* and *GATA6* and tumour suppressor genes *CDKN2A*, *SMAD4, TP53*, *BRCA1, ARID1A, PBRM1* and *SMARCA4*. Integrating copy number and expression data identified a number of genes/amplicons implicated in the progression of other cancer types that exhibited concordant gene expression changes (Supplementary Table 10). These included: amplification of *MIB1*, a known mediator of NOTCH signalling and pancreas development[14] and the *CCNE1-URI1* amplicon at 19q12 (Extended Data Fig. 2b). *CCNE1* is a marker of poor prognosis in ovarian, breast and lung cancers and is associated with resistance to platinum based therapy[15]. Recent small interfering RNA (siRNA) screening of PC cell lines provides supportive evidence for *CCNE1* amplification as an important mechanism in pancreatic carcinogenesis, and may represent a therapeutic opportunity using CDK inhibitors[16].

DNA deamination, ectopic APOBEC activity, BRCA-deficiency and mismatch repair were re-affirmed as the predominant mutational mechanisms in PC. Chromothriptic and break-fusion-bridge related genomic catastrophes were uncommon (12%; Supplementary Table 11). Somatic LINE-1 retro-transposition of known HotL1 elements was present in 35% of patients[17] (Supplementary Table 12). As only one of these events directly affected a known cancer gene (insertion into *ROBO2*), it appears unlikely that this is a major mutational mechanism in PC. No recurrent fusion events were detected (Supplementary Table 13).

Transcriptional networks and subtypes of PC

We used bulk tumour tissue to better understand the transcriptional networks and molecular mechanisms that underpin the tumour microenvironment. Initial unsupervised clustering of RNA-seq data for 96 tumours with high epithelial content (≥40%) to balance stromal gene expression resolved four stable classes (Fig. 1a and Extended Data Fig. 3). These four subtypes were also present in the extended set of 232 PCs using array-based mRNA expression profiles encompassing the full range of tumour cellularity (from 12–100%) (Extended Data Fig. 4). We named these subtypes: (1) squamous; (2) pancreatic progenitor; (3) immunogenic; and (4) aberrantly differentiated endocrine exocrine (ADEX) on the basis of the differential expression of transcription factors and downstream targets important in lineage specification and differentiation during pancreas development and regeneration. Transcriptional network analysis identified 26 coordinately expressed gene programmes representing distinct biological processes, 10 of which discriminated the 4 PC classes (Fig. 1b, Extended Data Fig. 5 and Supplementary Tables 14–16). These 4 subtypes were associated with specific histological characteristics: (1) squamous with adenosquamous carcinomas (6/25 in squamous versus 1/71 in the rest, $P = 0.0011$ Fisher's exact test); (2) pancreatic progenitor and (3) immunogenic with mucinous non-cystic (colloid) adenocarcinomas and carcinomas arising from IPMN, which are mucinous ($P = 0.0005$); and (4) ADEX with rare acinar cell carcinomas (although numbers were small, both cases clustered with the ADEX class) (Fig. 1a). Squamous subtype was an independent poor prognostic factor (Fig. 1c and Supplementary Table 21).

Squamous subtype

Four core gene programmes characterized squamous tumours (Fig. 1b), which included gene networks involved in inflammation, hypoxia response, metabolic reprogramming, TGF-β signalling, MYC pathway activation, autophagy and upregulated expression of *TP63ΔN* and its target genes. Many of these genes are highly expressed in the C2-squamous-like class of tumours of breast, bladder, lung and

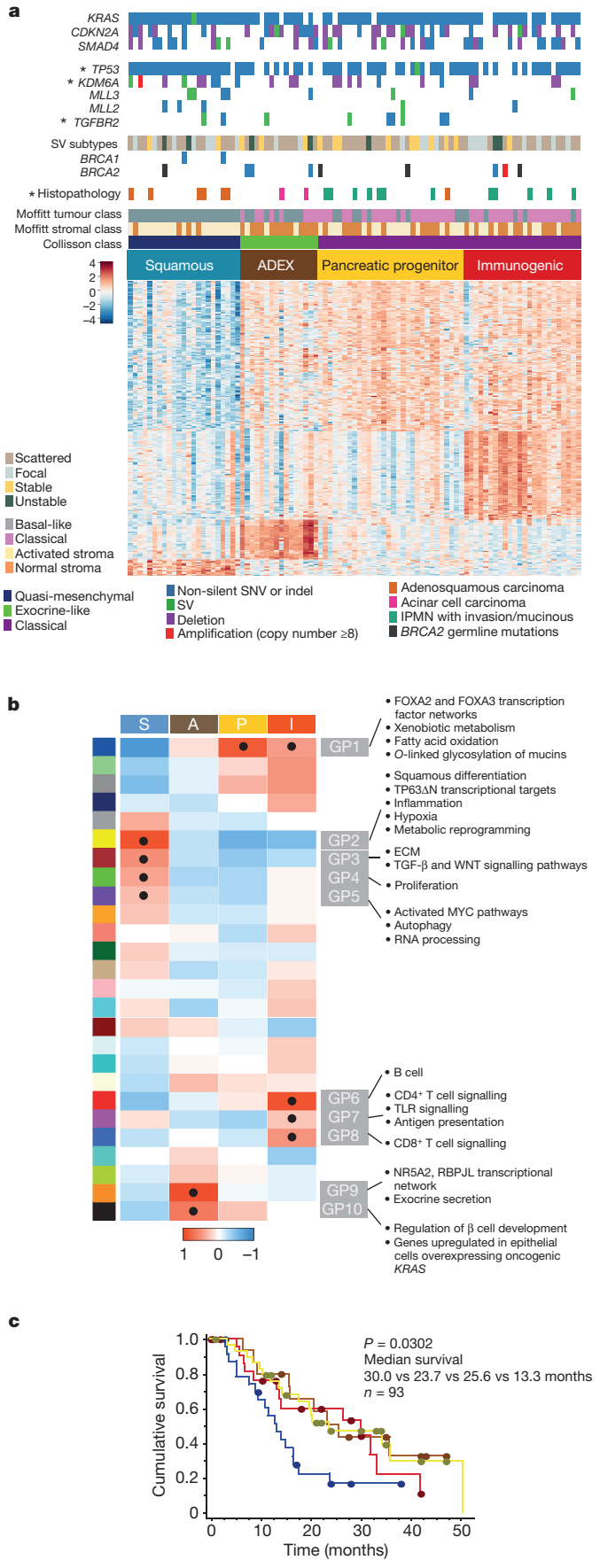


Figure 1 | Molecular classes and transcriptional networks defining PDAC. a, Unsupervised analysis of RNA-seq identified 4 PDAC classes: squamous (blue); ADEX (abnormally differentiated endocrine exocrine; brown); pancreatic progenitor (yellow); and immunogenic (red). *$P < 0.05$, Fisher's exact test. **b**, Heatmap of gene programmes significantly enriched in PDAC. Black dot denotes transcriptional networks showing highest significance for an individual class. **c**, Kaplan–Meier analysis of patient survival stratified by class.

head and neck cancer defined in the Cancer Genome Atlas (TCGA) pan-cancer studies[18], which was the reason we termed them squamous (Fig. 2a). As in these other cancer types, the pancreatic squamous subtype was associated with mutations in *TP53* ($P = 0.01$) and *KDM6A* ($P = 0.02$), which interacts with ASCOM complex constituents *MLL2* and *MLL3* (Figs 1a and 2b). Although previous immunohistochemical studies have identified increased *TP63* expression in adenosquamous pancreatic tumours[19], RNA-seq identified high *TP63ΔN* expression and its target genes as a key feature (Fig. 2c). *TP63ΔN*, in the presence of *TP53* mutation, is known to regulate epithelial cell plasticity, tumorigenicity and epithelial to mesenchymal transition in a variety of solid tumours[20]. Squamous tumours were enriched for activated α6β1 and α6β4 integrin signalling, and activated EGF signalling, (Extended Data Fig. 6 and Supplementary Table 16). The squamous subtype is associated with hypermethylation and concordant downregulation of genes that govern pancreatic endodermal cell-fate determination (for example, *PDX1*, *MNX1*, *GATA6*, *HNF1B*) leading to a complete loss of endodermal identity (Fig. 2d, e and Supplementary Table 17).

Pancreatic progenitor subtype

Transcriptional networks containing transcription factors *PDX1*, *MNX1*, *HNF4G*, *HNF4A*, *HNF1B*, *HNF1A*, *FOXA2*, *FOXA3* and *HES1* primarily define the pancreatic progenitor class (Extended Data Fig. 7). These transcription factors are pivotal for pancreatic endoderm cell-fate determination towards a pancreatic lineage and are linked to maturity onset diabetes of the young (MODY). *PDX1*, in particular, is critical for pancreas development with ductal, exocrine and endocrine cells all derived from embryonic progenitor cells that express *PDX1* (ref. 21). Gene programmes regulating fatty acid oxidation, steroid hormone biosynthesis, drug metabolism and *O*-linked glycosylation of mucins also define pancreatic progenitor tumours. Importantly, apomucins *MUC5AC* and *MUC1*, but not *MUC2* or *MUC6*, are preferentially co-expressed in pancreatic progenitor tumours. The expression of these apomucins defines the pancreatobiliary subtype of IPMN and is consistent with PDAC-associated IPMN clustering within this class (Supplementary Tables 14–16). *TGFBR2* inactivating mutations were also enriched in this subtype ($P = 0.029$).

ADEX subtype

The ADEX class is defined by transcriptional networks that are important in later stages of pancreatic development and differentiation, and is a subclass of pancreatic progenitor tumours. Transcriptional networks that characterize both exocrine and endocrine lineages at later stages are upregulated, rather than one or the other as is the case in normal pancreas development. The key networks identified include upregulation of: (i) transcription factors *NR5A2, MIST1* (also known as *BHLHA15A*) and *RBPJL* and their downstream targets that are important in acinar cell differentiation and pancreatitis/regeneration[22,23]; and (ii) genes associated with endocrine differentiation and MODY (including *INS, NEUROD1, NKX2-2* and *MAFA* (Extended Data Fig. 8 and Supplementary Table 16)). Importantly, several patient-derived pancreatic cancer cell lines were enriched with gene programmes associated with the ADEX class. Moreover, these cell lines expressed multiple genes associated with terminally differentiated pancreatic tissues, including *AMY2B*, *PRSS1*, *PRSS3*, *CEL* and *INS*. In addition, the methylation pattern of ADEX tumours was distinct from normal pancreas and clustered with other PCs (Extended Data Fig. 9).

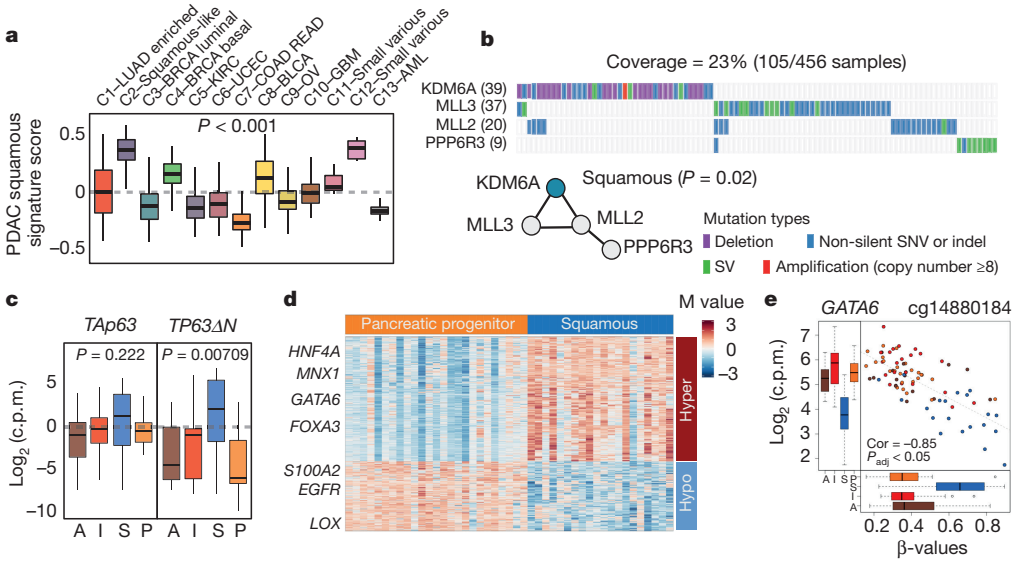


Figure 2 | Molecular characterization of the squamous class. a, Boxplot of PDAC squamous class signature scores generated using pan-cancer 12 expression data and stratified by class. **b**, Mutual exclusivity plot of a mutated gene sub-network identified by HotNet2. **c**, Boxplot of *TAp63* and *TP63ΔN* expression levels stratified by class. **d**, Heatmap of differentially methylated genes. **e**, Hypermethylation of GATA6 is associated with the concordant down regulation of *GATA6* gene expression. Pearson correlation and adjusted *P* values are as indicated. In **a** and **c** the boxplots are annotated by a Kruskall–Wallis *P* value.

Immunogenic subtype

The immunogenic class shares many of the characteristics of the pancreatic progenitor class, but is associated with evidence of a significant immune infiltrate. Associated immune gene programmes included B cell signalling pathways, antigen presentation, $CD4^+$ T cell, $CD8^+$ T cell and Toll-like receptor signalling pathways (Extended Data Fig. 10 and Supplementary Table 16). Enrichment analysis identified upregulated expression of genes associated with nine different immune cell types and/or phenotypes[24] (Fig. 3a). The predominant expression profiles were those related to infiltrating B and T cells, with both cytotoxic ($CD8^+$) and regulatory T cells ($CD4^+CD25^+FOXP3^+$ T_{regs}). Upregulation of CTLA4 and PD1 acquired tumour immune suppression pathways in the immunogenic subtype inferred therapeutic opportunities with novel immune modulators (Fig. 3c).

Immune mechanisms in pancreatic cancer

To better define candidate molecular mechanisms active in the tumour microenvironment, we correlated enrichment of expression patterns that characterize specific immune cell populations with each gene programme (Fig. 3a and Supplementary Tables 15, 16 and 18). Of all gene

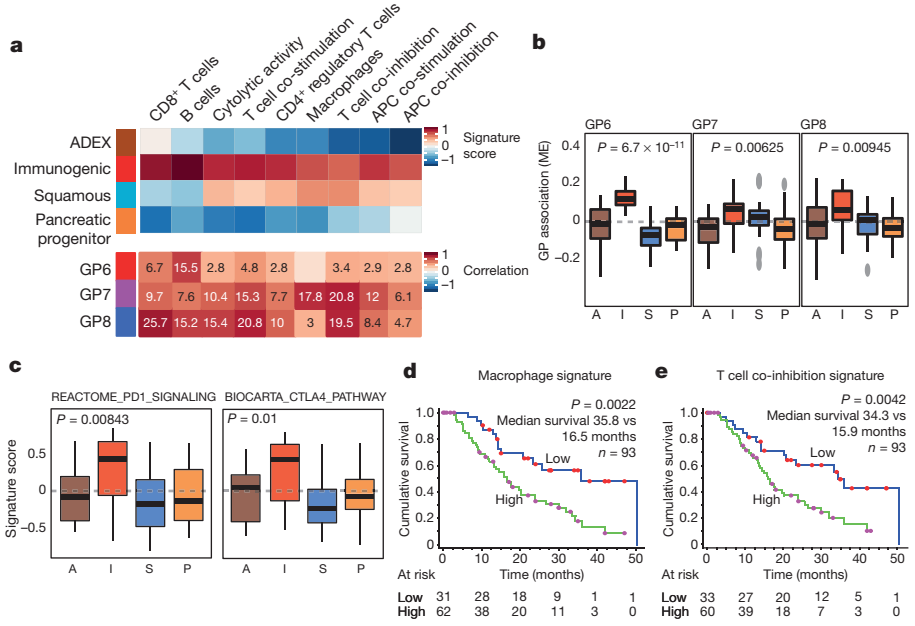


Figure 3 | Immune pathways in PDAC. a, Heatmap showing enrichment of immune cell/phenotype gene signatures in PDAC (top panel). Heatmap showing correlation of immune cell/phenotype gene signatures with the identified PDAC GPs (bottom panel). Numbers in cells represent $-\log_{10}$ of correlation significance. **b**, Boxplot of GP module eigengene (ME) scores (a measure of sample gene programme relatedness) stratified by class and showing GP class associations. **c**, Boxplot of *PD1* (also known as *PDCD1*) and *CTLA4* gene signature scores stratified by class. **d**, **e**, Kaplan–Meier analysis comparing survival of patients having either high or low immune cell/phenotype signature scores. In **b** and **c**, the boxplots are annotated by a Kruskall–Wallis *P* value.

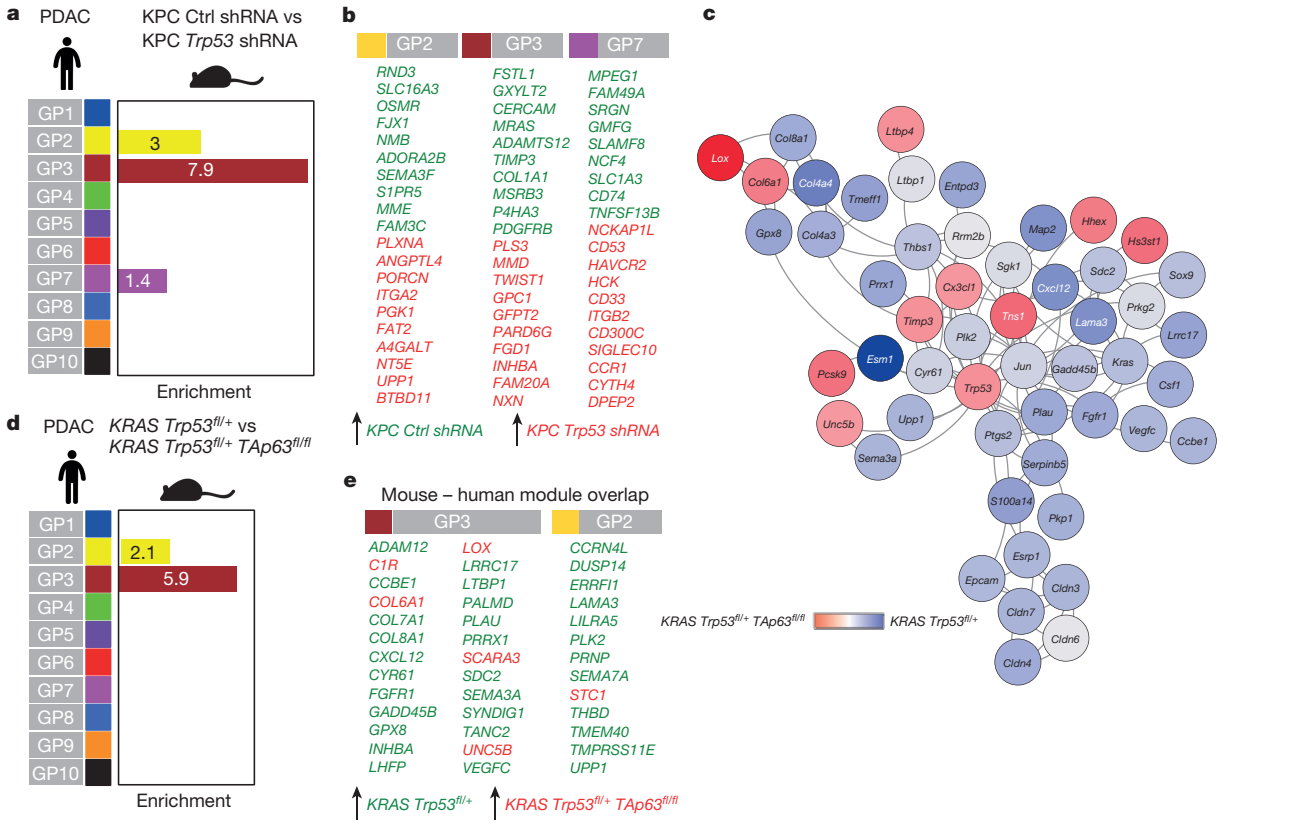


Figure 4 | Gain of function TP53 mutations and loss of TAp63 regulate key GPs associated with the squamous class. a, Significant GP enrichment of genes deregulated in KPC-mouse-derived cell lines treated with Trp53 specific short hairpin RNAs (shRNAs). **b**, Trp53 regulated genes enriched in either GP 2, 3 or 7. **c**, Sub-network of genes differentially expressed between *KRAS Trp53*$^{fl/+}$ and *KRAS Trp53*$^{fl/+}$ *Trp63*$^{fl/fl}$ cell lines. Node colour represents change in gene expression. **d**, Genes differential expressed between *KRAS Trp53*$^{fl/+}$ and *KRAS Trp53*$^{fl/+}$ *Trp63*$^{fl/fl}$ cell lines significantly enriched in GPs 2 and 3. **e**, Trp63 regulated genes enriched in GPs 2 and 3. In **a** and **d**, bars are annotated with significance values $-\log_{10}$ (*P* value). In **b** and **e**, the arrows and colour represent upregulation of gene expression in the indicated cell types.

programmes (GP), GP6, GP7 and GP8 were enriched with immune cell specific gene expression signatures (Fig. 3b). Specifically, GP6 and GP8 were associated with B cell and CD8$^+$ T cell signatures, respectively, with GP8 associated with the T cell co-inhibitory phenotype (Extended Data Fig. 10). GP7 was associated with both the macrophage signature and T-cell co-inhibition, which co-segregated with poor survival (Fig. 3d, e). Importantly, pathway analysis of GP7, also showed enrichment for antigen processing and presentation, and Toll-like receptor cascade(s) including high expression of *TLR4, TLR7, TLR8, PDCD1LG2* (*PD-L2*) and *CSF1R*. The latter are known mediators of tumour associated macrophage immunosuppression and inflammation.

TP53 and *TP63* modulation of squamous PDAC

Based on the association of *TP53* mutation and upregulated *TP63* expression in the squamous subtype, we used cell lines derived from genetically engineered mouse models of pancreatic cancer (*Kras*$^{G12D/+}$; *Trp53*$^{fl/+}$; *TAp63*$^{fl/fl}$ KPC mice) to begin to unravel the functional consequences of these events in defining squamous tumours. Mice with mutations in the DNA binding domain as compared to *TP53*-null animals have more aggressive disease with increased metastatic potential, primarily mediated through platelet-derived growth factor receptor β (*PDGFRB*)[25]. Analyses of transcriptome data from previous mutant *TP53* knockdown experiments from ref. 25 showed that mutant *TP53* regulates the expression of transcriptional networks associated with the squamous subtype, particularly GPs 2 and 3, including *PDGFRB* (Fig. 4a, b; Supplementary Table 20). *Kras*$^{G12D/+}$; *Trp53*$^{fl/+}$; *TAp63*$^{fl/fl}$ mice have more aggressive metastatic pancreatic cancer than their *Kras*$^{G12D/+}$; *Trp53*$^{fl/+}$ counterparts and also show deregulation of GPs 2 and 3, inferring that TAp63 plays an important role in squamous PC (Fig. 4c–e). Transcriptional network analysis identified additional key factors involved in metastasis that were upregulated in the squamous subtype for example, *LOX*[26].

Transcriptomic classification of PDAC

We compared our transcriptome classification with those of 2 previously published studies that had either physically[27] or virtually[28] micro-dissected tumour epithelium to define PC subtypes (Fig. 1a and Extended Data Fig. 9). Using their classifiers to subtype our data, 3 of the classes we defined directly overlap with the Collisson classification, with the exception of the novel immunogenic subtype. We altered Collisson's nomenclature to better reflect the insights into the molecular pathology and candidate mechanisms that our integrated analysis generated. The Collisson 'quasimesenchymal' subtype was renamed 'squamous' to reflect the molecular characteristics of squamous tumours across multiple tissue types, as defined by the TCGA pan-cancer analysis. 'Classical' was termed 'pancreatic progenitor' based on the prominence of transcriptional networks vital for early pancreas development, and the predominant discriminator from the squamous subtype. The Collisson 'exocrine-like' also contained transcriptional networks characteristic of committed endocrine differentiation and as a consequence were renamed ADEX. Although approximately 50% of squamous subtype tumours fell within the 'basal' subgroup of Moffitt *et al.*[28], the remainder were composed of a mixture of other Bailey/Collisson subtypes.

More sophisticated analyses using larger numbers of tumours continues to reveal novel insights into pancreatic cancer pathophysiology. In particular, integrated analysis of genomic, epigenomic and transcriptomic characteristics is generating biological insights with

potential therapeutic relevance. The increased appreciation of the role of the immune system in cancer development and progression has led to new classes of therapeutics that specifically target mechanisms through which the tumour evades immune destruction. Therapeutics that target some of these mechanisms are currently in clinical trials in many cancer types, including pancreatic cancer. Early clinical trial data suggest that, similar to most targeted therapies, patient selection will also be important for drugs that target the immune system. The novel immunogenic subtype of pancreatic cancer is characterized by specific mechanisms that can potentially be targeted using immune modulators, and testing in clinical trials is encouraged.

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 3 May; accepted 30 December 2015.
Published online 24 February; corrected online 2 March 2016 (see full-text HTML version for details).

1. Rahib, L. *et al.* Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. *Cancer Res.* **74,** 2913–2921 (2014).
2. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518,** 495–501 (2015).
3. Jones, S. *et al.* Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. *Science* **321,** 1801–1806 (2008).
4. Wang, L. *et al.* Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by *MLH1* haploinsufficiency and complete deficiency. *Genome Res.* **22,** 208–219 (2012).
5. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491,** 399–405 (2012).
6. Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Campbell, P. J. & Stratton, M. R. Deciphering signatures of mutational processes operative in human cancer. *Cell Rep.* **3,** 246–259 (2013).
7. Nones, K. *et al.* Genome-wide DNA methylation patterns in pancreatic ductal adenocarcinoma reveal epigenetic deregulation of SLIT-ROBO, ITGA2 and MET signaling. *Int. J. Cancer* **135,** 1110–1118 (2014).
8. Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nature Commun.* **5,** 5224 (2014).
9. Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521,** 489–494 (2015).
10. The Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511,** 543–550 (2014).
11. Bechara, E. G., Sebestyen, E., Bernardis, I., Eyras, E. & Valcarcel, J. RBM5, 6, and 10 differentially regulate NUMB alternative splicing to control cancer cell proliferation. *Mol. Cell* **52,** 720–733 (2013).
12. Scott, L. M. & Rebel, V. I. Acquired mutations that affect pre-mRNA splicing in hematologic malignancies and solid tumors. *J. Natl. Cancer Inst.* **105,** 1540–1549 (2013).
13. Maguire, S. L. *et al. SF3B1* mutations constitute a novel therapeutic target in breast cancer. *J. Pathol.* **235,** 571–580 (2015).
14. Horn, S. *et al. Mind bomb 1* is required for pancreatic β-cell formation. *Proc. Natl Acad. Sci. USA* **109,** 7356–7361 (2012).
15. Scaltriti, M. *et al.* Cyclin E amplification/overexpression is a mechanism of trastuzumab resistance in HER2$^+$ breast cancer patients. *Proc. Natl Acad. Sci. USA* **108,** 3761–3766 (2011).
16. Shain, A. H., Salari, K., Giacomini, C. P. & Pollack, J. R. Integrative genomic and functional profiling of the pancreatic cancer genome. *BMC Genomics* **14,** 624 (2013).
17. Tubio, J. M. *et al.* Mobile DNA in cancer. Extensive transduction of nonrepetitive DNA mediated by L1 retrotransposition in cancer genomes. *Science* **345,** 1251343 (2014).
18. Hoadley, K. A. *et al.* Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* **158,** 929–944 (2014).
19. Brody, J. R. *et al.* Adenosquamous carcinoma of the pancreas harbors KRAS2, DPC4 and TP53 molecular alterations similar to pancreatic ductal adenocarcinoma. *Mod. Pathol.* **22,** 651–659 (2009).
20. Engelmann, D. & Putzer, B. M. Emerging from the shade of p53 mutants: N-terminally truncated variants of the p53 family in EMT signaling and cancer progression. *Sci. Signal.* **7,** re9 (2014).
21. Hale, M. A. *et al.* The homeodomain protein PDX1 is required at mid-pancreatic development for the formation of the exocrine pancreas. *Dev. Biol.* **286,** 225–237 (2005).
22. von Figura, G., Morris, J. P. IV, Wright, C. V. & Hebrok, M. Nr5a2 maintains acinar cell differentiation and constrains oncogenic Kras-mediated pancreatic neoplastic initiation. *Gut* **63,** 656–664 (2014).
23. Hale, M. A. *et al.* The nuclear hormone receptor family member NR5A2 controls aspects of multipotent progenitor cell formation and acinar differentiation during pancreatic organogenesis. *Development* **141,** 3123–3133 (2014).
24. Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160,** 48–61 (2015).
25. Weissmueller, S. *et al.* Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor β signaling. *Cell* **157,** 382–394 (2014).
26. Miller, B. W. *et al.* Targeting the LOX/hypoxia axis reverses many of the features that make pancreatic cancer deadly: inhibition of LOX abrogates metastasis and enhances drug efficacy. *EMBO Mol. Med.* **7,** 1063–1076 (2015).
27. Collisson, E. A. *et al.* Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. *Nature Med.* **17,** 500–503 (2011).
28. Moffitt, R. A. *et al.* Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. *Nature Genet.* **47,** 1168–1178 (2015).

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

Acknowledgements We would like to thank C. Axford, M.-A. Brancato, S. Rowe, M. Thomas, S. Simpson and G. Hammond for central coordination of the Australian Pancreatic Cancer Genome Initiative, data management and quality control; M. Martyn-Smith, L. Braatvedt, H. Tang, V. Papangelis and M. Beilin for biospecimen acquisition; and Deborah Gwynne for support at the Queensland Centre for Medical Genomics. We also thank M. Hodgins, M. Debeljak and D. Trusty for technical assistance at Johns Hopkins University. Funding support was from: National Health and Medical Research Council of Australia (NHMRC; 631701, 535903, 427601); Queensland Government (NIRAP); University of Queensland; Australian Government: Department of Innovation, Industry, Science and Research (DIISR); Australian Cancer Research Foundation (ACRF); Cancer Council NSW: (SRP06-01, SRP11-01. ICGC); Cancer Institute NSW: (10/ECF/2-26; 06/ECF/1-24; 09/CDF/2-40; 07/CDF/1-03; 10/CRF/1-01, 08/RSA/1-15, 07/CDF/1-28, 10/CDF/2-26,10/FRL/2-03, 06/RSA/1-05, 09/RIG/1-02, 10/TPG/1-04, 11/REG/1-10, 11/CDF/3-26); Garvan Institute of Medical Research; Cancer Research UK Glasgow Centre Program, A18076; Avner Nahmani Pancreatic Cancer Research Foundation; R.T. Hall Trust; Petre Foundation; Philip Hemstritch Foundation; Gastroenterological Society of Australia (GESA); American Association for Cancer Research (AACR) Landon Foundation—INNOVATOR Award; Wellcome Trust Senior Investigator Award 103721/Z/14/Z; Cancer Research UK Programme Grant C29717/A17263; Cancer Research UK Programme Grant A12481; Pancreatic Cancer UK; The Howat Foundation; University of Glasgow; European Research Council Starting Grant, 311301, Italian Ministry of University and Research (Cancer Genome Project FIRB RBAP10AHJB), Associazione Italiana Ricerca Cancro (n.12182) , Fondazione Italiana Malattie Pancreas – Ministry of Health (CUP_J33G13000210001), European Community Grant FP7 Cam-Pac, grant agreement number 602783.

Author Contributions Investigator contributions are as follows: P.J.B., J.V.P., N.W., A.V.B., S.M.G. (concept and design); P.J.B., D.A.W., R.A.G., A.S., D.K.C., J.V.P., N.W., A.V.B., S.M.G. (project leaders); P.J.B., D.K.C., A.V.B., S.M.G. (writing team); D.K.M., A.N.C., T.J.C.B., C.N., K.N., S.W., D.M.M., N.W., L.E., L.M., L.S., S.M.G., I.H., S.I., S.M., E.N., K.Q., S.M.G. (genomics); P.J.B., D.K.M., K.S.K., N.W., P.J.W., O. H., A.M.P., F.N., O.H., C.L., D.T., S.W., Q.X., K.N., N.C., M.Q., M.A., A.R., M.G., S.K., K.Q., L.P., J.M., M.C., L.C.M., O.S., L.F., U.B., N.W., J.V.P., S.M.G. (data analysis); D.K.C., A.L.J., A.M.N., A.M., A.V.P., C.W.T., E.K.C., E.S.H., I.R., M.G., J.H., J.A.L., K.E., L.A.C., M.D.J., A.J.G., N.Q.N., A.B., N.Z., C.P., R.G., J.R.E., R.H.H., A.M., C.A.I., C.L.W., B.R., V.C., P.C., C.B., R.S., G.T., D.M., G.M.P., J.H., M.P., J.W., V.C., C.J.S., J.G.K., R.T.L., N.D.M., N.B.J., J.S.G., J.D.S., R.A.M., J.H., S.A.K., K.M., R.L.S., A.V.B. (sample acquisition and processing, clinical annotation, interpretation and analysis); A.J.G., A.C., R.H.H., F.D., K.O., A.S., W.F., J.G.K., C.T. (pathology assessment).

Author Information All DNA sequencing and RNA-seq data have been deposited in the European Genome-phenome Archive (EGA): accession code EGAS00001000154. All gene expression, genotyping, and methylome data used in this study has been deposited in the NCBI Gene Expression Omnibus (GEO) under accession codes GSE49149 and GSE36924. Mouse cell line expression data are available in the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-4415. 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. Correspondence and requests for materials should be addressed to A.V.B. (andrew.biankin@glasgow.ac.uk) or S.M.G. (sean.grimmond@unimelb.edu.au).

METHODS

Human research ethical approvals. APGI: Sydney South West Area Health Service Human Research Ethics Committee, western zone (protocol number 2006/54); Sydney Local Health District Human Research Ethics Committee (X11-0220); Northern Sydney Central Coast Health Harbour Human Research Ethics Committee (0612-251M); Royal Adelaide Hospital Human Research Ethics Committee (091107a); Metro South Human Research Ethics Committee (09/QPAH/220); South Metropolitan Area Health Service Human Research Ethics Committee (09/324); Southern Adelaide Health Service/Flinders University Human Research Ethics Committee (167/10); Sydney West Area Health Service Human Research Ethics Committee (Westmead campus) (HREC2002/3/4.19); The University of Queensland Medical Research Ethics Committee (2009000745); Greenslopes Private Hospital Ethics Committee (09/34); North Shore Private Hospital Ethics Committee. Johns Hopkins Medical Institutions: Johns Hopkins Medicine Institutional Review Board (NA00026689). ARC-Net, University of Verona: approval number 1885 from the Integrated University Hospital Trust (AOUI) Ethics Committee (Comitato Etico Azienda Ospedaliera Universitaria Integrata) approved in their meeting of 17 November 2010 and documented by the ethics committee 52070/CE on 22 November 2010 and formalized by the Health Director of the AOUI on the order of the General Manager with protocol 52438 on 23 November 2010. Ethikkommission an der Technischen Universität Dresden (Approval numbers EK30412207 and EK357112012).

Patient material acquisition and extraction. Samples were acquired through the Australian Pancreatic Cancer Genome Initiative (APGI) as part of the International Cancer Genome Consortium (ICGC). Informed consent was obtained from all subjects. Tissue dissection of primary material, RNA and DNA extraction was performed using previously published methods[2]. Tumour cellularity was estimated for each sample using a combination of qPure analysis of high-density SNP profiles and *KRAS* amplicon sequencing[2]. Primary tumours ($n=342$) and 41 patient-derived cell lines (representing low cellularity tumours) (Supplementary Table 1) underwent whole genome sequencing when tumour cellularity was >40% (mean coverage 75×, $n=179$), or deep-exome sequencing (mean coverage: 400×, $n=204$) for samples with a cellularity of 12–40%.

Exome library preparation. Exome libraries were generated using the Illumina Nextera Rapid Capture Exome kit (Illumina, Part no. FC-140–1003) according to the standard manufacturer's protocol (part no. 15037436 Rev. A February 2013), except they were made in an automated high-throughput fashion using Perkin Elmer's Sciclone G3 NGS Workstation (Product no. SG3-31020-0300). Then 50 ng of gDNA was used as input for tagmentation followed by 10 cycles of PCR to produce sufficient library for exome capture. A total of 500 ng of each library was pooled as a 12-plex reaction for capture using Illumina's Nextera Exome Oligo set. Following two rounds of capture, samples were finally subjected to 10 cycles of PCR to produce exome libraries ready for sequencing. Prior to sequencing, exome libraries were qualified via either the Perkin Elmer LabChip GX with the DNA High Sensitivity LabChip kit (Perkin Elmer, Part no. CLS760672), or the Agilent BioAnalyzer 2100 with the High Sensitivity DNA Kit (Agilent, Part no. 5067–4626). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

Whole-genome library preparation. Whole-genome libraries were generated using either the Illumina TruSeq DNA LT sample preparation kit (Illumina, Part no. FC-121–2001 and FC-121–2001) or the Illumina TruSeq DNA PCR-free LT sample preparation kit (Illumina, Part no. FC-121–3001 and FC-121–3002) according to the manufacturer's protocols with some modifications (Illumina, Part no. 15026486 Rev. C July 2012 and 15036187 Rev. A January 2013 for the two different kits respectively). For the TruSeq DNA LT sample preparation kit, 1 μg of gDNA was used as input for fragmentation to ~300 bp, followed by a SPRI-bead clean up using the AxyPrep Mag PCR Clean-Up kit (Corning, Part no. MAG-PCR-CL-250). After end-repair, 3′ adenylation and adaptor ligation, the libraries were size-selected using a double SPRI-bead method to obtain libraries with an insert size ~300 bp. The size-selected libraries were subjected to 8 cycles of PCR to produce the final whole-genome libraries ready for sequencing. For the TruSeq DNA PCR-free LT sample preparation kit, 1 μg of gDNA was used as input for fragmentation to ~350 bp, followed by an end-repair step and then a size-selection using the double SPRI-bead method to obtain libraries with an insert size ~350 bp. The size-selected libraries then underwent 3′ adenylation and adaptor ligation to produce final whole genome libraries ready for sequencing. Prior to sequencing, whole-genome libraries were qualified via the Agilent BioAnalyzer 2100 with the High Sensitivity DNA Kit (Agilent, Part no. 5067–4626). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

Total RNA library preparation. RNA-Seq libraries were generated using the Illumina TruSeq Stranded Total RNA LT sample preparation kit (with Ribo-Zero Gold) (Illumina, Part no. RS-122–2301 and RS-122–2302), according to the standard manufacturer's protocol (Part no. 15031048 Rev. D April 2013), except they were made in an automated high-throughput fashion using Perkin Elmer's Sciclone G3 NGS Workstation (Product no. SG3-31020-0300). The ribosomal depletion step was performed on 1 μg of total RNA using Ribo-Zero Gold before a heat fragmentation step aimed at producing libraries with an insert size between 120–200 bp. cDNA was then synthesized from the enriched and fragmented RNA using SuperScript II Reverse Transcriptase (Invitrogen, Catalog no. 18064) and random primers. The resulting cDNA was converted into double-stranded DNA in the presence of dUTP to prevent subsequent amplification of the second strand and thus maintain the strandedness of the library. Following 3′ adenylation and adaptor ligation, libraries were subjected to 15 cycles of PCR to produce RNA-seq libraries ready for sequencing. Prior to sequencing, RNA-seq libraries were qualified via the Perkin Elmer LabChip GX with the DNA High Sensitivity LabChip kit (Perkin Elmer, Part no. CLS760672). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kit - Illumina/Universal (KAPA Biosystems, Part no. KK4824) in combination with the Life Technologies Viia 7 real time PCR instrument.

Library sequencing. All libraries were sequenced using the Illumina HiSeq 2000/2500 system with TruSeq SBS Kit v3 - HS (200-cycles) reagents (Illumina, Part no. FC-401-3001), to generate paired-end 101 bp reads.

Sequence alignment and data management. Sequence data was mapped to the Genome Reference Consortium GRCh37 assembly using BWA42. All BAM files have been deposited in the EGA (accession number: EGAS00001000154).

Copy number analysis. Matched tumour and normal patient DNA was assayed using Illumina SNP BeadChips as per manufacturer's instructions (Illumina, San Diego CA) (HumanOmni1-Quad or HumanOmni2.5–8 BeadChips) and analysed as previously described[2,8].

Identification and verification of structural variants. The Somatic structural variant pipeline were identified using the qSV tool. A detailed description of its use has been recently published[2,8].

Identification of and verification of point mutations. Substitutions and indels were called using a consensus calling approach that included qSNP, GATK and Pindel. The details of call integration and filtering, and verification using orthogonal sequencing and matched sample approaches are as previously described[2,8,9]. 97% of *KRAS* mutations identified by *KRAS* deep-amplicon sequencing were detected via WGS and WES, inferring a false negative rate of 3% (Supplementary Table 1).

'Lollipop' plots. Plots showing the location and frequency of inactivating mutations were generated using the MutationMapper web tool hosted at http://www.cbioportal.org/. Available PanCancer mutation data was downloaded from the Cancer Genomic Data Server (CGDS) hosted by the Computational Biology Center (cBio) at the Memorial Sloan-Kettering Cancer Center (MSKCC) using the R package "cgdsr"[29].

Mutational signatures. Mutational signatures were defined for genome-wide somatic substitutions, as previously described[2].

Significantly mutated gene detection. A combination of three robust approaches were used to define significantly mutated genes: (i) MutSigCV2 (ref. 30), which detects genes with point mutations above the background mutation rate; (ii) OncodriverFM[31], which detects point mutated genes with a bias towards pathogenic mutations; and (iii) HOTNET2 (ref. 32), which identifies sub-networks based on protein–protein interactions that contain recurrent point mutations, copy number alterations and structural rearrangements. The HotNet2 (HotNet diffusion-oriented subnetworks) algorithm was used to identify significantly mutated subnetworks in a genome-scale interaction network Heat scores for each protein were calculated as the number of samples having a non-silent SNV, indel, SV or copy number aberration in the corresponding gene[32]. Heat scores were limited to proteins having a corresponding gene mutation in ≥2% of samples. The iRefIndex interaction network was used for the analysis[33]. Supplementary Table 20 contains matrices summarizing all mutations, CNVs and SVs for all samples used in this study.

RNA sequencing library generation and sequencing. RNA-seq libraries were generated using TruSeq Stranded Total RNA (part no. 15031048 Rev. D April 2013) kits, using on a Perkin Elmer's Sciclone G3 NGS Workstation (product no. SG3-31020-0300). Ribosomal depletion step was performed on 1 μg of total RNA using Ribo-Zero Gold before a heat fragmentation step aimed at producing libraries with an insert size between 120–200 bp. cDNA was then synthesized from the enriched and fragmented RNA using Invitrogen's SuperScript II Reverse Transcriptase (catalogue number 18064) and random primers. The resulting cDNA was further converted into double stranded DNA in the presence of dUTP to prevent

subsequent amplification of the second strand and thus maintain the strandedness of the library. Following 3′ adenylation and adaptor ligation libraries were subjected to 15 cycles of PCR to produce RNA-seq libraries ready for sequencing. Prior to sequencing, exome and RNA-seq libraries were qualified and quantified via Caliper's LabChip GX (part no. 122000) instrument using the DNA High Sensitivity Reagent kit (product no. CLS760672). Quantification of libraries for clustering was performed using the KAPA Library Quantification Kits For Illumina sequencing platforms (kit code KK4824) in combination with Life Technologies Viia 7 real time PCR instrument.

RNA-seq analysis. Sequencing reads were mapped to transcripts corresponding to ensemble 70 annotations using RSEM[34]. RSEM data were normalized using TMM (weighted trimmed mean of M-values) as implemented in the R package 'edgeR'. For downstream analyses, normalized RSEM data were converted to counts per million (c.p.m.) and $\log_2$ transformed[35]. Genes without at least 1 c.p.m. in 20% of the sample were excluded from further analysis.

RNA-seq re-analysis of Weismuller *et al.* RNA sequencing data reported in ref. 25 was downloaded from the Sequence Read Archive (SRA): Accession number; SRP033333. The available data was re-analysed using an RNA-seq pipeline implemented in the bcbio-nextgen project (https://bcbio-nextgen.readthedocs.org/en/latest/). Briefly, after quality control and adaptor trimming, reads were aligned to the UCSC mouse mm10 genome build using STAR[36] Counts for known genes were generated using the function featureCounts in the R/Bioconductor package "Rsubread"[37]. The R/Bioconductor package "DESeq2" was used to identify differentially expressed genes[38].

***KRAS Trp53*$^{fl/+}$ and *KRAS Trp53*$^{fl/+}$ *Trp63*$^{fl/fl}$ mouse derived cell lines.** Cell lines were generated in house from pancreatic tumours harvested from *Pdx1-Cre*, *LSL-Kras*$^{G12D/+}$, *Trp53*$^{fl/+}$ mice or *Pdx1-Cre*, *LSL-Kras*$^{G12D/+}$, *Trp53*$^{fl/+}$, *TAp63*$^{fl/fl}$ mice described previously[49]. Low passage cell lines were used and authenticated by morphology. Mycoplasma testing confirmed that all cell lines were mycoplasma negative. Independently derived cell-lines representing either the *KRAS Trp53*$^{fl/+}$ ($n=3$) or *KRAS Trp53*$^{fl/+}$ *Trp63*$^{fl/fl}$ ($n=3$) genotype were used for RNA-seq analysis. RNA-seq libraries were generated using the KAPA stranded RNaseq Kit with RiboErase (HMR) (KAPA Biosystems; kit ref. KR1151 – v3.15) according to the manufacturer's instructions. Briefly, samples were fragmented for 6 min at 94 °C with 10 cycles of library amplification. Library quality control was performed using an Agilent BioAnalyzer 2100 in combination with a High Sensitivity DNA Kit (Agilent, Part no. 5067-4626). Samples were evenly pooled to a 2 nM concentration and a 1% PhiX control spike-in was used for sequencing quality control. Libraries were run on the NextSeq 500 platform according to the manufacturer's instructions (Illumina, San Diego CA). Sequenced libraries were mapped to UCSC mouse mm10 genome build using TopHat and differential gene expression determined using Cufflinks 2.1.1 and Cuffdiff 2.1.1 as implemented in BaseSpace (https://basespace.illumina.com/home/indexIllumina, San Diego CA).

Microarray analysis. Tumour RNA was assayed using HumanHT-12 v4 Expression BeadChips as per manufacturer's instructions (Illumina, San Diego CA) and analysed as previously described[7]. Batch correction was performed using the R package 'sva'[39].

Clustering. Non-negative matrix factorization (NMF) was employed to identify stable sample clusters[40] The top 2,000 most variable genes were used as input. NMF parameters: Brunet algorithm; $k=1$ to $k=7$ clusters; number of clusterings to build consensus matrix = 20; error function = Euclidean; and 500 iterations. The preferred clustering result was determined using the observed cophenetic correlation between clusters and the average silhouette width of the consensus membership matrix as determined by the R package 'cluster'. The R package 'ConsensusClusterPlus'[41] was also employed to verify sample clustering. Similar sample clusters were obtained using both methods (data not shown). The package 'ConsensusClusterPlus' was also used to subtype PC samples according to the expression signatures defined in Moffitt *et al.*[28]

Differential gene expression (DGE). To identify the most representative samples within each cluster, we computed silhouette widths using the R 'cluster' package. Samples with positive silhouette widths were retained for DGE analysis. DGE analysis between representative samples was performed using the function 'voom' as implemented in the R package 'edgeR'[42]. To define genes differentially expressed between all classes we used the function 'sam' as implemented in the R package 'siggenes'.

Gene sets. Gene sets representing immune cell-type expression markers and immune meta-genes were obtained from a recent publication[24]. Gene sets representing PDAC classes were generated by selecting significantly upregulated genes in a given class versus all other classes. An adjusted *P* value of 0.01 was used as the cut-off in each case.

Gene set enrichment. Gene set enrichment was performed using the R package 'GSVA' (function gsva - arguments: method = "gsva", mx.diff = TRUE)[43]. GSVA implements a non-parametric unsupervised method of gene set enrichment that allows an assessment of the relative enrichment of a selected pathway across the sample space. The output of GSVA is a gene-set by sample matrix of GSVA enrichment scores that are approximately normally distributed. GSVA enrichment scores were generated for each gene set using the transformed RSEM data unless otherwise indicated. For survival analyses, sample GSVA enrichment scores were stratified into quantiles (for example, lower 33% or upper 66% of values).

WGCNA. Weighted gene co-expression network analysis (WGCNA) was used to generate a transcriptional network from the normalized and transformed RSEM[44]. Briefly, WGCNA clusters genes into network modules using a topological overlap measure (TOM). The TOM is a highly robust measure of network interconnectedness and essentially provides a measure of the connection strength between two adjacent genes and all other genes in a network. Genes are clustered using 1-TOM as the distance measure and gene modules are defined as branches of the resulting cluster tree using a dynamic branch-cutting algorithm[45].

The module eigengene is used as a measure of module expression in a given sample and is defined as the first principle component of a module. To relate sample traits of interest to gene modules, sample traits were correlated to module eigengenes and significance determined by a Student asymptotic *P* value for the given correlations. For gene module survival analyses, module eigengenes were stratified into quantiles (for example, lower 33% or upper 66% of values). To relate gene modules to PDAC classes, PDAC class gene set GSVA enrichment scores were used as sample traits and correlated with the module eigengenes as discussed above. Similarly, to relate the immune cell-type expression markers and immune meta-genes to the gene modules each immune GSVA enrichment score was correlated with the module eigengenes as before.

To determine the enrichment of differentially expressed mouse genes in modules generated by WGCNA, mouse identifiers were first mapped to their corresponding human HGCN Symbol using the R/Bioconductor package "biomaRt". Module gene enrichment was then determined using the function userListEnrichment in the WGCNA package. We considered, as significant, only those modules showing both significant enrichment and significant gene expression/gene module correlations.

Pathway analysis. Ontology and pathway enrichment analysis was performed using the R package 'dnet'[46] and/or the Reactome FI Cytoscape plugin 4.1.1 (ref. 47) as indicated. The R package 'dnet' was also used to identify significant sub-networks of differentially expressed genes.

Pan-cancer 12 data and squamous assignment. Platform corrected input data was obtained from Synapse as part of the Pan-Cancer 12 data freeze (syn1715755)[18]. Pan-cancer 12 subtype assignments were also obtained from Synapse (syn1889916) and sample sizes, as indicated, used for statistical comparisons. To determine the relationship between the PDAC classes and the pan-cancer 12 subtypes, PDAC class gene sets were used in combination with the pan-cancer 12 expression data to generated GSVA enrichment scores as discussed above. Sample GSVA enrichment scores representing each PDAC class were then stratified according to the pan-cancer 12-subtype assignments. A Kruskal–Wallis test was applied to the stratified scores to determine whether the distributions differed.

Methylation analysis. Sample methylation was determined using Illumina 450K arrays as previously described[7] with the following modifications. Probe-level Illumina GeneStudio output files were imported into R package 'lumi'[48] and data filtered to remove failed hybridizations, probes comprising SNPs and probes located on sex chromosomes. The filtered methylation values were then colour balance corrected and normalized using Shift and Scaling Normalization (SSN) as implemented by lumi. Gene methylation values were obtained by collapsing probe level values for a given gene loci (that is, probes located 1,500 bp upstream of the transcriptional start site (TSS) through to the end of transcription) using the function collapseRows(method = "maxRowVariance") from the package WGCNA[44]. Probes were mapped to gene loci using the R package "genomicFeatures". Differential gene methylation between representative samples (selected as above under heading Differential gene expression (DGE)) was determined using the R package 'limma'. M-values were used for differential gene methylation analysis. Concordant changes in methylation and expression were calculated as follows. Probes were mapped to a given gene using the R package "genomicFeatures". As above, probes located 1,500 bp upstream of the TSS through to the end of transcription were considered for each gene. The correlation between a probe β-value and the corresponding gene $\log_2$ (CPM) expression value was then calculated using Pearson's correlation coefficient. The statistical significance of each probe/gene correlation was calculated by permuting the data 10,000 times and comparing the correlation coefficients obtained before and after permutation. The resulting *P* values were adjusted for multiple testing using the approach of Benjamini and Hochberg.

Concordant copy number expression analysis. Analysis of variance was used to identify significant changes in gene expression between samples exhibiting corresponding gene copy number aberrations. Accordingly, gene expression

values were stratified on the basis of sample copy number change for a given gene. Deleted genes and genes having copy number ≥ 8 were considered for the analysis. *P* values were adjusted using the R package 'qvalue' for which adjusted values <0.05 were considered statistically significant. Variance was similar between groups compared. Genes showing copy number aberrations in less than 3% of samples were excluded from the analysis.

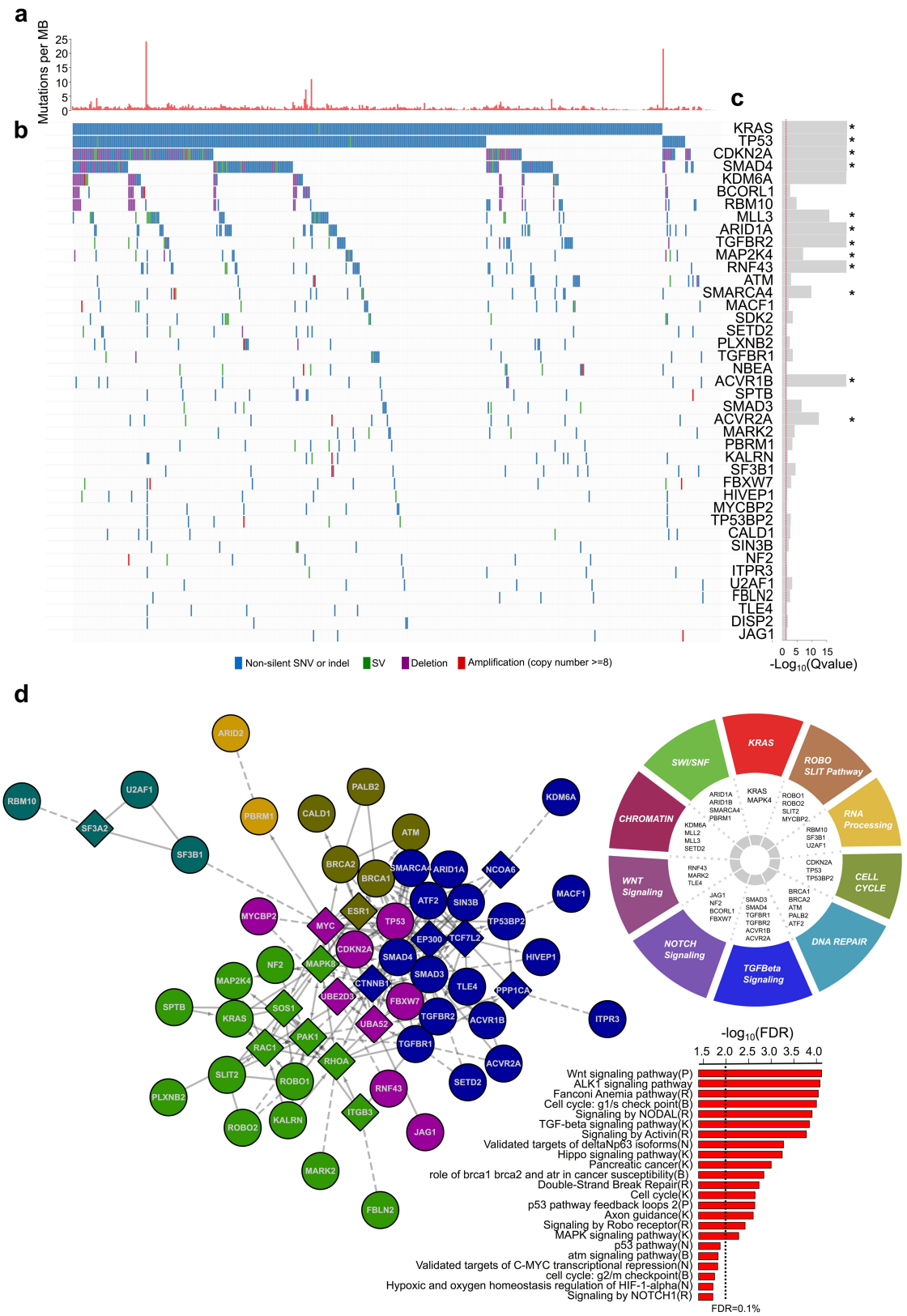
Survival analysis. The date of diagnosis and the date and cause of death were obtained from the Central Cancer Registry and treating clinicians. Median survival was estimated using the Kaplan–Meier method and the difference was tested using the log-rank test. *P* values of less than 0.05 were considered statistically significant. Clinicopathologic variables analysed with a *P* value <0.25 on log-rank test were entered into Cox proportional hazards multivariate analysis. Statistical analysis was performed using StatView 5.0 Software (Abacus Systems, Berkeley, CA, USA). Disease-specific survival was used as the primary endpoint.

Stromal cell and immune infiltrate quantification. To quantify stromal and immune cell tumour contributions we used the R package 'estimate'[50].

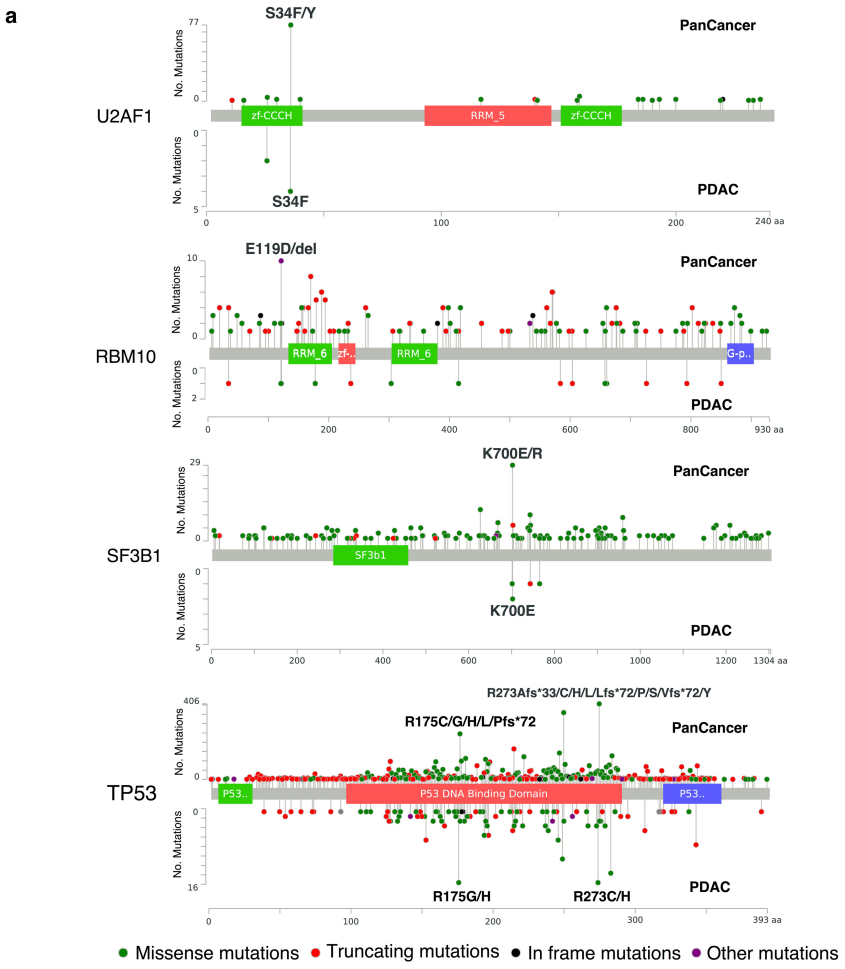
Statistical analysis. A Kruskal–Wallis test was applied to the indicated stratified scores to determine whether distributions were significantly different. For all PDAC class comparisons using RNA-seq data, the following sample sizes were compared: ADEX ($n=14$), immunogenic ($n=24$); squamous ($n=20$); and pancreatic progenitor ($n=25$). Fisher's exact tests were used to evaluate the association between dichotomous variables.

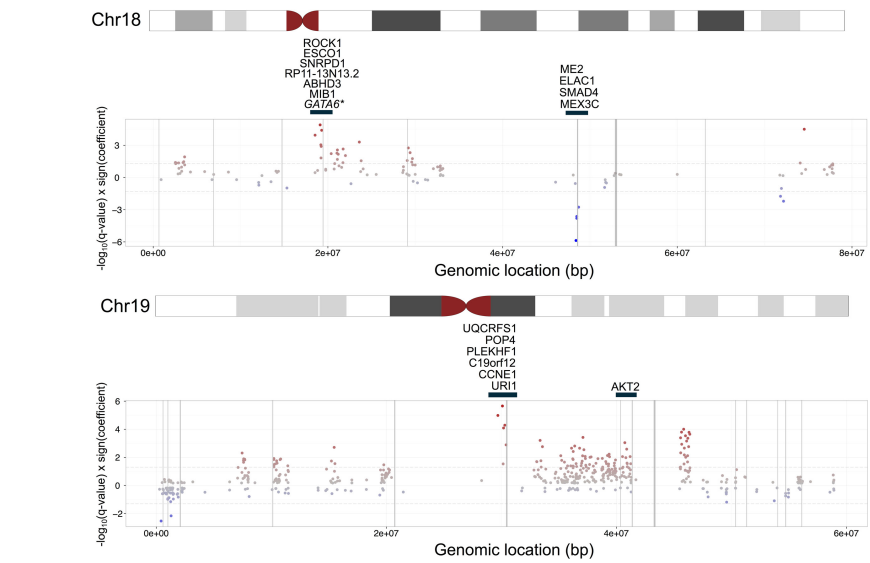
Data. Data presented in this study can be downloaded from the following repository https://dcc.icgc.org/repositories under the identifier PACA-AU.

29. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov.* **2,** 401–404 (2012).
30. Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499,** 214–218 (2013).
31. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nature Methods* **10,** 1081–1082 (2013).
32. Leiserson, M. D. *et al.* Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nature Genet.* **47,** 106–114 (2015).
33. Razick, S., Magklaras, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9,** 405 (2008).
34. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12,** 323 (2011).
35. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26,** 139–140 (2010).
36. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
37. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30,** 923–930 (2014).
38. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15,** 550 (2014).
39. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28,** 882–883 (2012).
40. Kuehn, H., Liberzon, A., Reich, M. & Mesirov, J. P. Using GenePattern for gene expression analysis. *Current Protoc. Bioinformatics* Chapter 7, Unit 7 12, (2008).
41. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26,** 1572–1573 (2010).
42. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15,** R29 (2014).
43. Hänzelmann, S., Castelo, R. & Guinney, J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* **14,** 7 (2013).
44. Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9,** 559 (2008).
45. Langfelder, P., Zhang, B. & Horvath, S. Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics* **24,** 719–720 (2008).
46. Fang, H. & Gough, J. The 'dnet' approach promotes emerging research on cancer patient survival. *Genome Med.* **6,** 64 (2014).
47. Wu, G. & Stein, L. A network module-based method for identifying cancer prognostic signatures. *Genome Biol.* **13,** R112 (2012).
48. Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24,** 1547–1548 (2008).
49. Tan, E. H. *et al.* Functions of TAp63 and p53 in restraining the development of metastatic cancer. *Oncogene* **25,** 3325–3333 (2014).
50. Yoshihara, K. *et al.* Inferring tumour purity and stromal and immune cell admixture from expression data. *Nature Commun.* **4,** 2612 (2013).

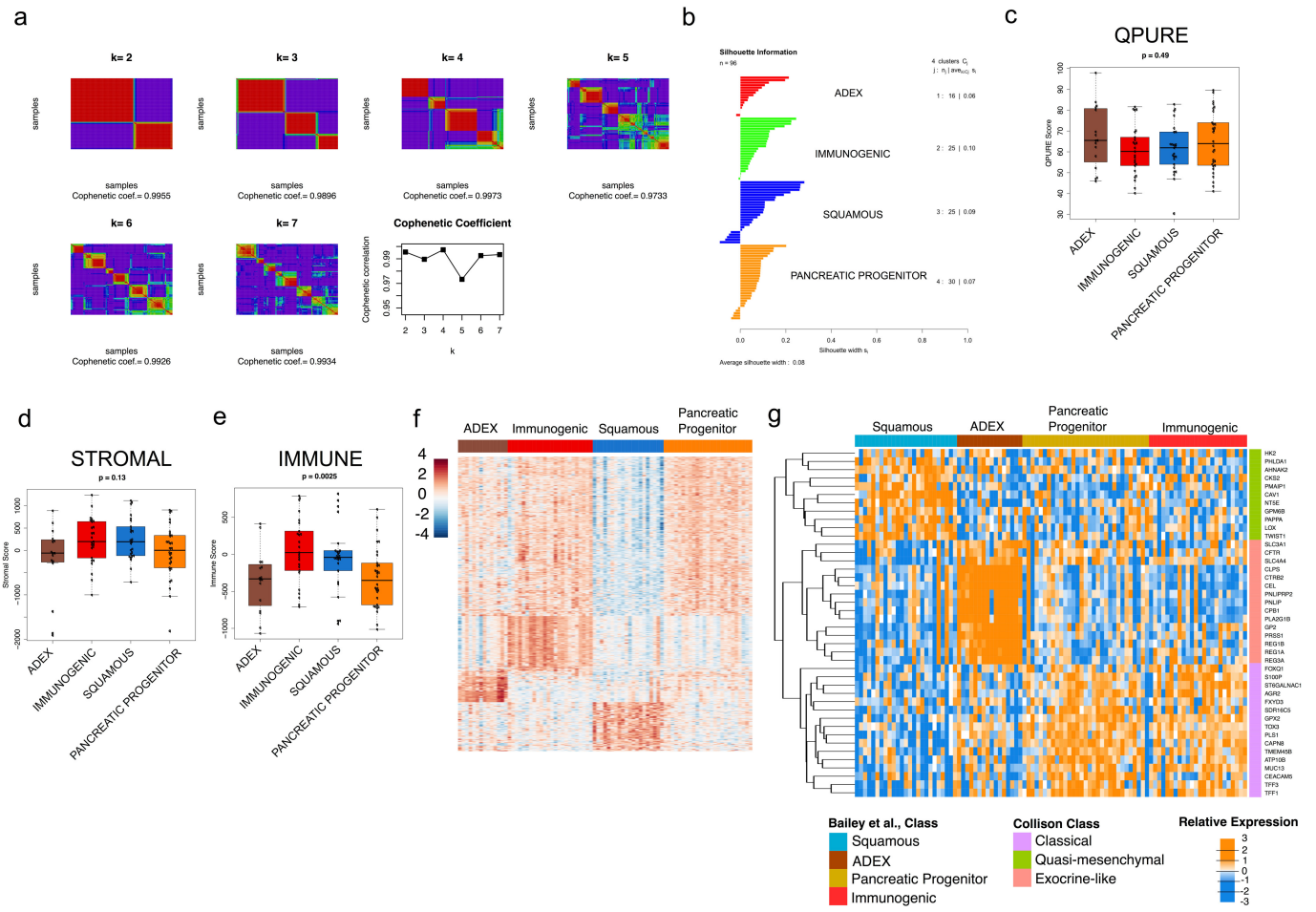


Extended Data Figure 1 | Mutational landscape of PC. a, Barplot representing the somatic mutation rate for each of the 456 samples included in this analysis. **b**, Non-silent mutations (blue), amplifications (≥8 copies, red), deletions (purple) and structural variants (SV, green) ranked in order of exclusivity. **c**, Significantly mutated genes identified by OncodriverFM. An asterisk denotes a significantly mutated gene identified by both MutSigCV and OncodriverFM. **d**, PC mutation functional interaction (FI) sub-network identified by the ReactomeFI cytoscape plugin. Mutated genes are indicated as coloured circles and linker genes (that is, genes not significantly mutated but highly connected to mutated genes in the network) indicated as coloured diamonds. Different node colours indicate different network clusters or closely interconnected genes. *P* values represent FDR < 0.05. Pathways significantly enriched in the identified FI sub-network are shown in the accompanying bar graph. Linker genes were not included in the enrichment analysis. Pie chart representing significantly altered genes and pathways in PC.

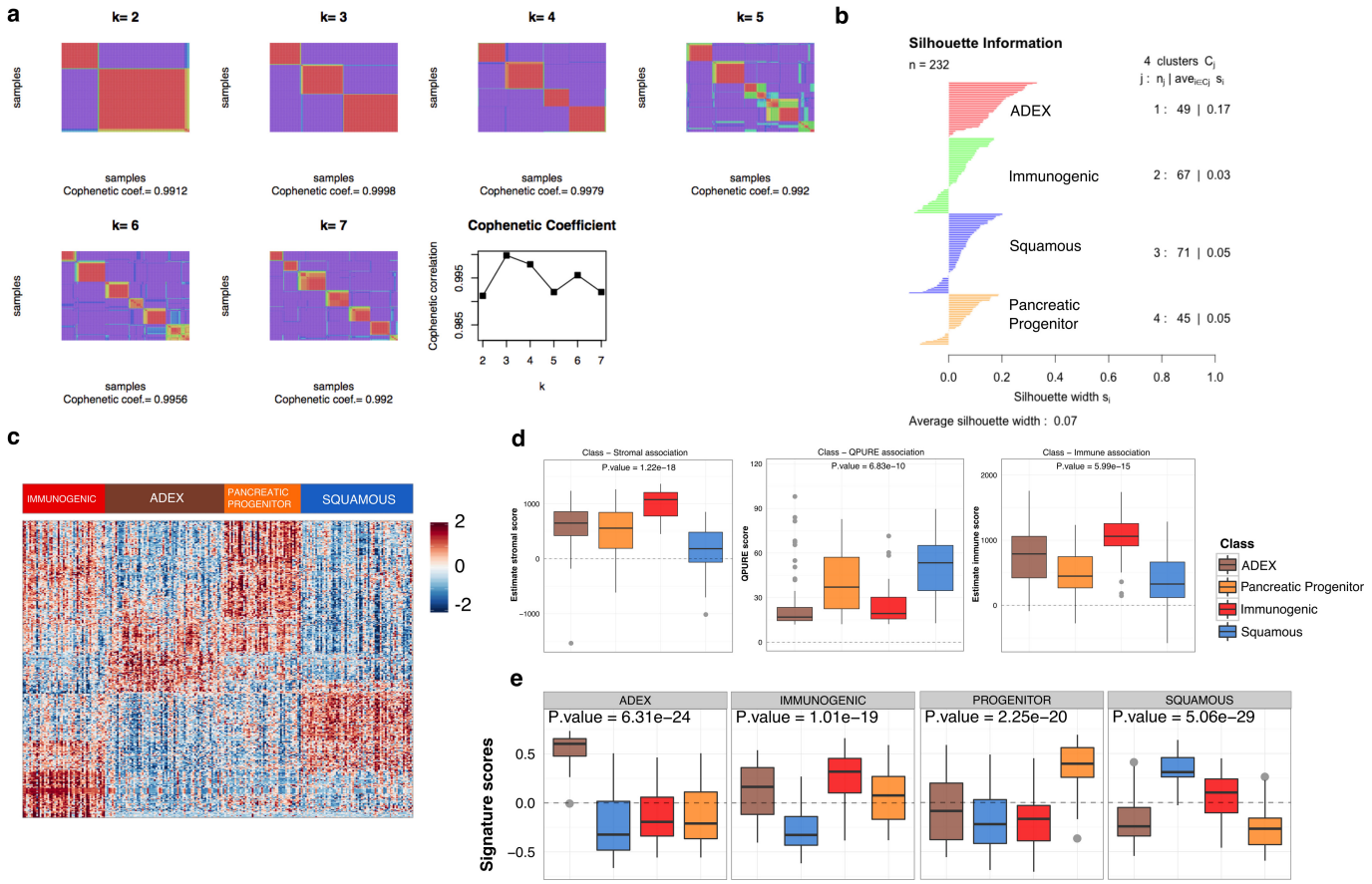




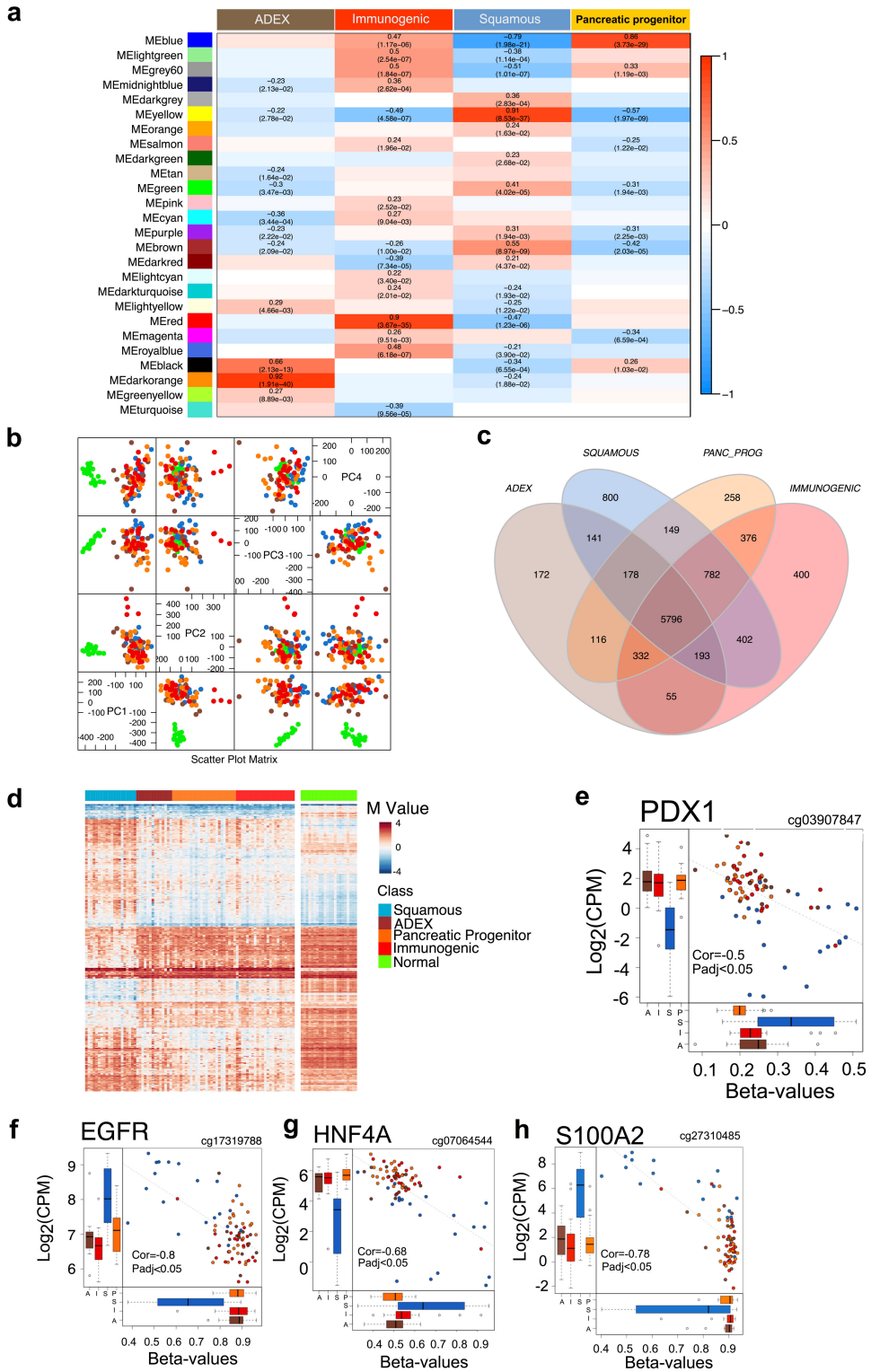
Extended Data Figure 2 | Selected genomic events in PC. a, Lollipop plots showing the type and location of mutations in the RNA processing genes *RBM10*, *SF3B1* and *U2AF1* and the tumour suppressor *TP53*. In each plot, mutations observed across multiple cancers (top plot; PanCancer) are compared with those observed in the current study (bottom plot; PDAC). Significant recurrent mutations are labelled above the relevant lollipop. **b**, Regions of copy number alteration showing concordant gene expression changes. For each of the indicated chromosomes, significant GISTIC peaks are shown at their respective genomic locations (*x* axis) as grey bars. Each gene is represented by a dot at its specific chromosomal coordinate, with blue representing concordant copy number loss and gene downregulation and red representing concordant copy number amplification (copy number ≥ 8) and gene upregulation. Significance of concordant copy number/expression change is measured as a value of $-\log_{10}$ (*q*-value) times the sign of the direction of change. Dotted lines represent a significance threshold of $-\log_{10}$ (*q*-value = 0.05) times the sign of the direction of change. Genes showing concordant copy number/expression changes and overlapping GISTIC peaks are listed above the plot. Asterisk denotes known PC oncogenes showing amplification but non-significant concordant copy number/expression change.



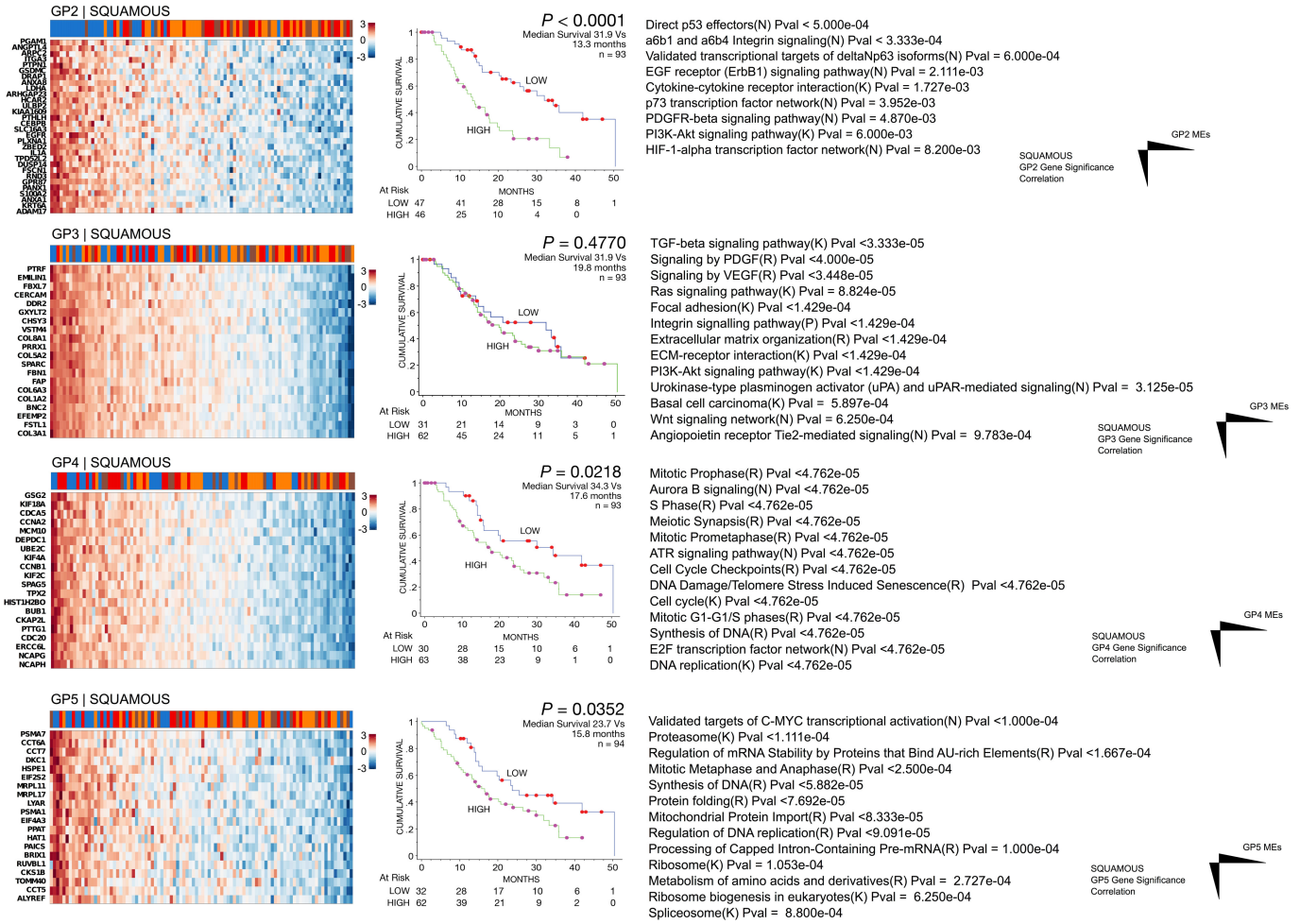
Extended Data Figure 3 | Classification of PC into 4 classes. **a**, Unsupervised classification of PC RNAseq using NMF. Solutions are shown for $k = 2$ to $k = 7$ classes. A peak cophenetic correlation is observed for $k = 4$ classes. **b**, Silhouette information for $k = 4$ classes. **c**–**e**, Boxplots representing QPURE, stromal signature scores and immune signature scores stratified by class. Boxplots are annotated by a Kruskall–Wallis *P* value. For comparisons the following sample sizes were used: ADEX ($n = 16$); immunogenic ($n = 25$); squamous ($n = 25$); and pancreatic progenitor ($n = 30$). **f**, Heatmap showing differential gene expression between classes. Samples with positive silhouette widths were retained for 'sam' analysis. **g**, Heatmap showing overlap of the 4 classes identified in the current study and Collisson *et al.* classification[27].



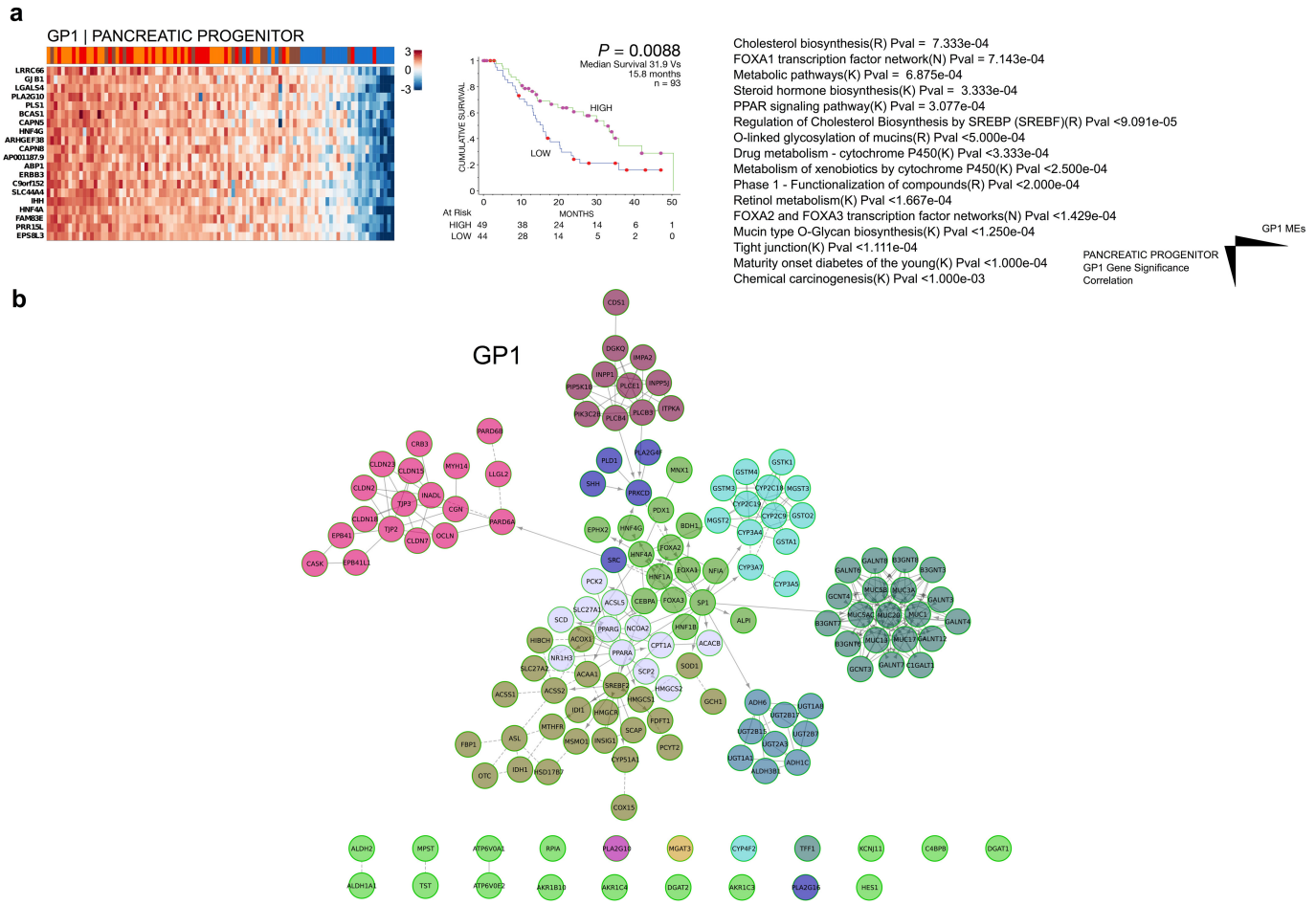
Extended Data Figure 4 | Identification of 4 robust PC classes in 232 PCs with mixed low and high cellularity. a, Unsupervised classification of PC expression array data representing 232 samples using NMF. Solutions are shown for $k = 2$ to $k = 7$ classes. **b**, Silhouette information for $k = 4$ classes. **c**, Heatmap showing differential gene expression between classes. **d**, Boxplots representing QPURE, stromal signature scores and immune signature scores stratified by class. **e**, Boxplots representing ADEX, pancreatic progenitor, squamous and immunogenic signature scores defined using the RNA-seq PC set stratified by class. Boxplots in **d** and **e** are annotated by a Kruskall–Wallis *P* value. For comparisons the following sample sizes were used: ADEX ($n = 49$); immunogenic ($n = 67$); squamous ($n = 71$); and pancreatic progenitor ($n = 45$).



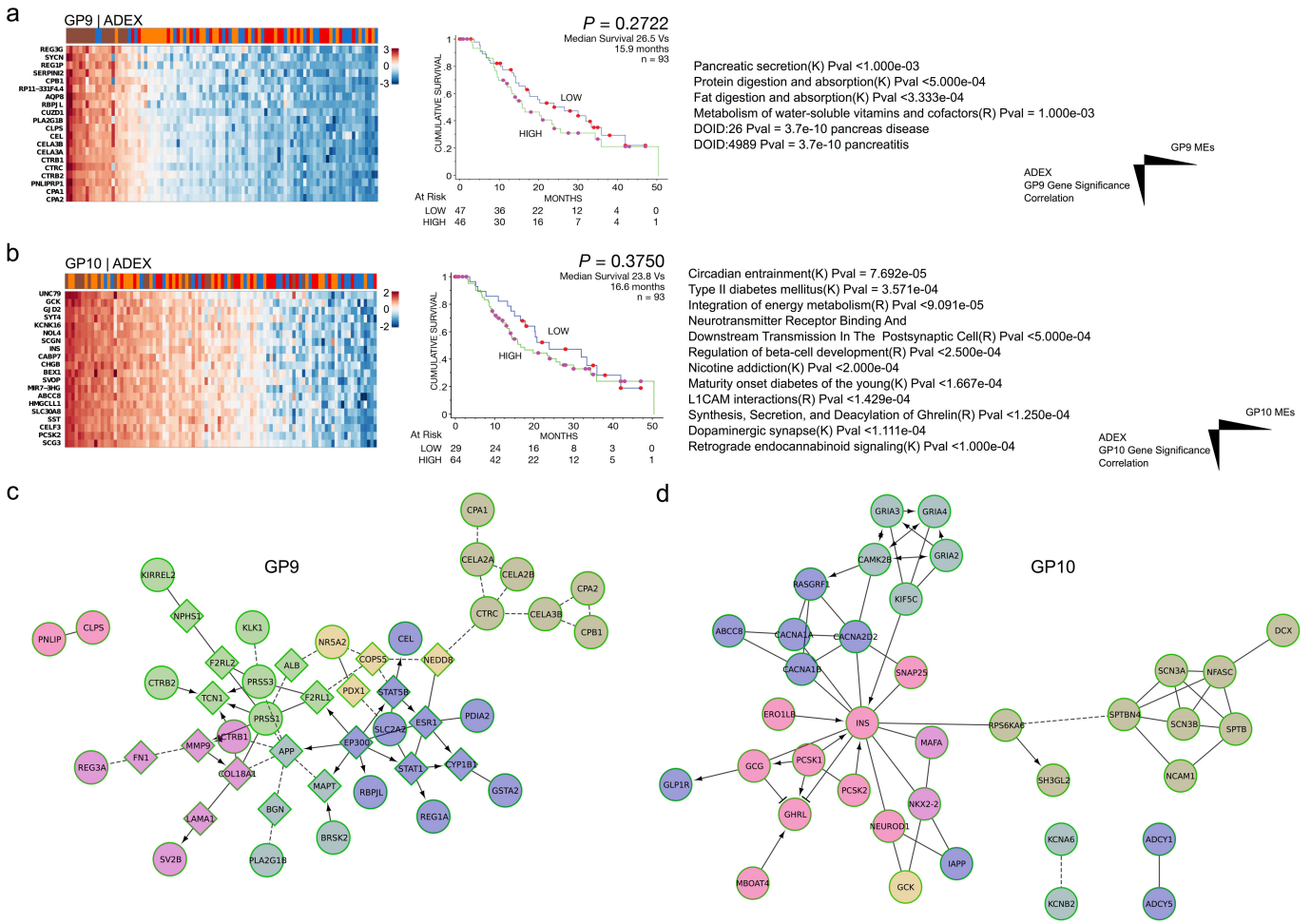
Extended Data Figure 5 | Characterization of PC subtypes. a, Heatmap showing the statistical significance of correlations observed between the expressions of genes significantly expressed in each PC class and gene programmes identified by WGCNA. Pearson correlations and Student's asymptotic *P* values are provided in each cell. **b**, Principal component analysis (PCA) using methylation data. Plot showing pairwise comparisons of samples distributed along the identified principle components (PC). Adjacent non-tumorous pancreatic samples represented as green points cluster as a distinct group. PC samples represented by points coloured brown (ADEX), blue (squamous), orange (pancreatic progenitor) and red (immunogenic) cluster together. **c**, Venn diagram showing the number of common and unique genes differentially methylated in the indicated PC subtypes when compared to adjacent non-tumorous pancreas. It is observed that distinct subsets of genes are differentially methylated in the 4 PC subtypes. **d**, Heatmap showing genes that are significantly methylated between tumours comprising the squamous class and all other classes. Methylation values for the same genes in adjacent non-tumorous pancreas are also shown. **e**–**h**, Plots showing regulation of gene expression by methylation. Hyper- or hypomethylation of the indicated probe is associated with either the concordant downregulation or upregulation of the indicated gene. Pearson correlation and adjusted *P* values are provided for each gene methylation comparison. Boxplot colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). Single letter designations representing the first letter of each class are provided under the relevant boxes in each plot.



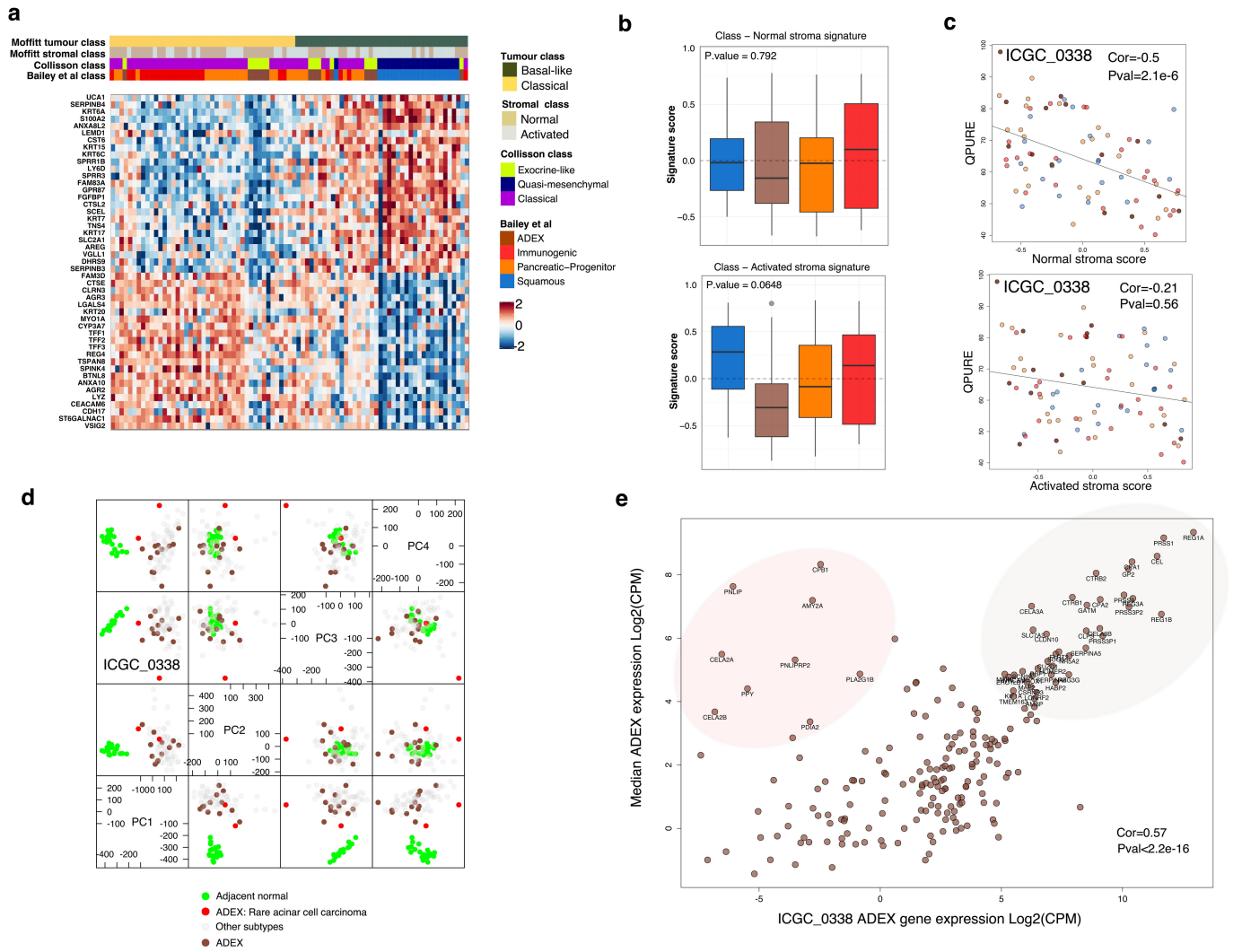
Extended Data Figure 6 | Core gene programmes (GP) defining the squamous class. Each panel shows from left to right: (i) a heatmap representing the genes in the specified gene programme most correlated with the indicated PC class with tumours ranked according to their gene programme module eigengene values (MEs) (PC classes are designated by colour as follows: ADEX (brown); pancreatic progenitor (orange); immunogenic (red); and squamous (blue)); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low gene programme MEs; (iii) pathways significantly enriched in a given GP functional interaction (FI) sub-network defined by the ReactomeFI Cytoscape plugin. *P* values represent FDR < 0.05.



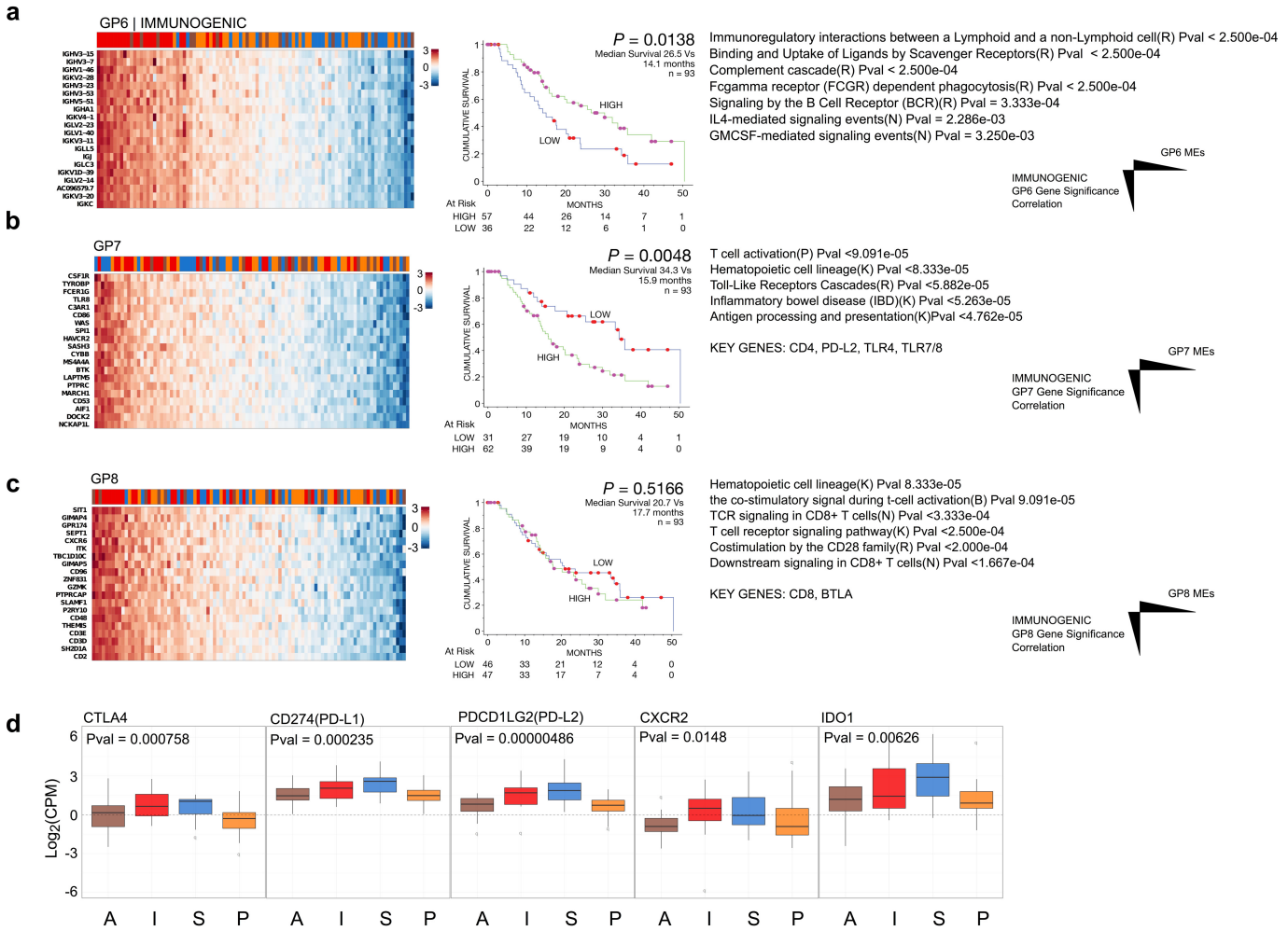
Extended Data Figure 7 | Gene programme defining the pancreatic progenitor class. a, Panel showing from left to right: (i) a heatmap representing the genes in GP1 most correlated with the pancreatic progenitor class with tumours ranked according to their GP1 module eigengene values (MEs); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low GP1 MEs; (iii) pathways significantly enriched in a GP1 FI sub-network defined by the ReactomeFI Cytoscape plugin. *P* values represent FDR <0.05. **b**, Network diagram depicting pathways significantly enriched in GP1 (FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes.



Extended Data Figure 8 | Gene programmes defining the ADEX class. **a**, **b**, Panel showing from left to right: (i) a heatmap representing the genes in the specified GP most correlated with the ADEX class with tumours ranked according to their GP module eigengene values (MEs); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low GP MEs; (iii) pathways significantly enriched in a GP FI sub-network defined by the ReactomeFI Cytoscape plugin. *P* values represent FDR <0.05. **c**, Network diagram depicting pathways significantly enriched in GP9 (FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes. Genes comprising GP9 are indicated as coloured circles, whereas linker genes (genes not comprising GP9 but forming multiple connections in the network) are indicated as coloured diamonds. **d**, Network diagram depicting pathways significantly enriched in GP10 (FDR <0.0001). Different node colours indicate different network clusters or closely interconnected genes.



Extended Data Figure 9 | Stratification of PC RNASeq data according to Moffitt *et al*. **a**, Heatmap showing the stratification of the PC cohort of the current study using the tumour subtype classifier published in Moffitt *et al.*[28]. PCs were classified by consensus clustering using the top 50 weighted genes associated with the basal-like or classical subtypes. **b**, Boxplots showing the distribution of normal and activated stroma signature scores between the 4 PC classes identified in the current study. Boxplots are annotated by a Kruskall–Wallis *P* value. A significant difference in activated stroma signature scores was observed between squamous and ADEX tumours P value < 0.01 (t-test). Boxplot colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). **c**, Plots showing correlation between tumour cellularity, presented as a QPURE score, and either activated or normal stroma signature scores. Plots are annotated with Pearson correlation scores and significance values, with a linear fit represented by a solid line. Sample ICGC_0338, a rare acinar cell carcinoma is highlighted. This sample exhibits near 100% cellularity and has low activated or normal stroma signature scores. **d**, Principal component analysis (PCA) using methylation data. Plot showing pairwise comparisons of samples distributed along the identified principle components (PC). Adjacent non-tumorous pancreatic samples represented as green points cluster as a distinct group relative to ADEX samples (brown and red points). Rare acinar cell carcinomas (red) cluster with other ADEX samples (brown). All other PC samples are shown as grey points. **e**, Plot showing the correlation of expression of representative genes expressed in acinar cell carcinoma sample ICGC_0338 compared to the median expression of the same genes across all other ADEX samples. A red shaded region encompasses genes showing high median expression in all other ADEX but low expression in ICGC_0338. A brown shaded region encompasses genes showing high median expression in all other ADEX and correlatively high expression in ICGC_0338. Pearson's correlation and significance are indicated.



Extended Data Figure 10 | Gene programmes defining the immunogenic class. a–c, Each panel shows from left to right: (i) a heatmap representing the genes in the specified gene programme most correlated with the indicated PC class with tumours ranked according to their gene programme module eigengene values (MEs). PC classes are designated by colour as follows: ADEX (brown); pancreatic progenitor (orange); immunogenic (red); and squamous (blue); (ii) Kaplan–Meier analysis comparing survival of patients having either high or low gene programme MEs; (iii) pathways significantly enriched in a given GP functional interaction (FI) sub-network defined by the ReactomeFI Cytoscape plugin. Corresponding Cytoscape files comprising GP ReactomeFI subnetworks are provided. **d**, Boxplot of immune gene expression stratified by class. Boxplots are annotated by a Kruskall–Wallis *P* value and box colours designate class: ADEX (brown); immunogenic (red); squamous (blue); and pancreatic progenitor (orange). Single letter designations representing the first letter of each class are provided under the relevant boxes in each plot.