



Scarpa, A. et al. (2017) Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*, 543(7643), pp. 65-71.  
(doi: [10.1038/nature21063](https://doi.org/10.1038/nature21063))

This is the author's final accepted version.

There may be differences between this version and the published version.  
You are advised to consult the publisher's version if you wish to cite from it.

<http://eprints.gla.ac.uk/137698/>

Deposited on: 12 December 2018

Enlighten – Research publications by members of the University of Glasgow  
<http://eprints.gla.ac.uk>

## Whole-genome landscape of pancreatic neuroendocrine tumours

Aldo Scarpa<sup>1,2,\*§</sup>, David K. Chang<sup>3,4, 7,29,36\*</sup>, Katia Nones<sup>5,6\*</sup>, Vincenzo Corbo<sup>1,2\*</sup>, Ann-Marie Patch<sup>5,6</sup>, Peter Bailey<sup>3,6</sup>, Rita T. Lawlor<sup>1,2</sup>, Amber L. Johns<sup>7</sup>, David K. Miller<sup>6</sup>, Andrea Mafficini<sup>1</sup>, Borislav Rusev<sup>1</sup>, Maria Scardoni<sup>2</sup>, Davide Antonello<sup>8</sup>, Stefano Barbi<sup>2</sup>, Katarzyna O. Sikora<sup>1</sup>, Sara Cingarlini<sup>9</sup>, Caterina Vicentini<sup>1</sup>, Skye McKay<sup>7</sup>, Michael C. J. Quinn<sup>5,6</sup>, Timothy J. C. Bruxner<sup>6</sup>, Angelika N. Christ<sup>6</sup>, Ivon Harliwong<sup>6</sup>, Sene Idrisoglu<sup>6</sup>, Suzanne McLean<sup>6</sup>, Craig Nourse<sup>3, 6</sup>, Ehsan Nourbakhsh<sup>6</sup>, Peter J. Wilson<sup>6</sup>, Matthew J. Anderson<sup>6</sup>, J. Lynn Fink<sup>6</sup>, Felicity Newell<sup>5,6</sup>, Nick Waddell<sup>6</sup>, Oliver Holmes<sup>5,6</sup>, Stephen H. Kazakoff<sup>5,6</sup>, Conrad Leonard<sup>5,6</sup>, Scott Wood<sup>5,6</sup>, Qinying Xu<sup>5,6</sup>, Shivashankar Hiriyur Nagaraj<sup>6</sup>, Eliana Amato<sup>1,2</sup>, Irene Dalai<sup>1,2</sup>, Samantha Bersani<sup>2</sup>, Ivana Cataldo<sup>1,2</sup>, Angelo P. Dei Tos<sup>10</sup>, Paola Capelli<sup>2</sup>, Maria Vittoria Davi<sup>11</sup>, Luca Landoni<sup>8</sup>, Anna Malpaga<sup>8</sup>, Marco Miotto<sup>8</sup>, Vicki L.J. Whitehall<sup>5,12,13</sup>, Barbara A. Leggett<sup>5,12,14</sup>, Janelle L. Harris<sup>5</sup>, Jonathan Harris<sup>15</sup>, Marc D. Jones<sup>3</sup>, Jeremy Humphris<sup>7</sup>, Lorraine A. Chantrill<sup>7</sup>, Venessa Chin<sup>7</sup>, Adnan M. Nagrial<sup>7</sup>, Marina Pajic<sup>7</sup>, Christopher J. Scarlett<sup>7,16</sup>, Andreia Pinho<sup>7</sup>, Ilse Rooman<sup>7†</sup>, Christopher Toon<sup>7</sup>, Jianmin Wu<sup>7,17</sup>, Mark Pinese<sup>7</sup>, Mark Cowley<sup>7</sup>, Andrew Barbour<sup>18</sup>, Amanda Mawson<sup>7†</sup>, Emily S. Humphrey<sup>7</sup>, Emily K. Colvin<sup>7</sup>, Angela Chou<sup>7,19</sup>, Jessica A. Lovell<sup>7</sup>, Nigel B. Jamieson<sup>3,4,20</sup>, Fraser Duthie<sup>3,21</sup>, Marie-Claude Gingras<sup>22,23</sup>, William E. Fisher<sup>23</sup>, Rebecca A. Dagg<sup>24</sup>, Loretta M.S. Lau<sup>24</sup>, Michael Lee<sup>25</sup>, Hilda A. Pickett<sup>25</sup>, Roger R. Reddel<sup>25</sup>, Jaswinder S. Samra<sup>26,27</sup>, James G. Kench<sup>7,27,28</sup>, Neil D. Merrett<sup>29,30</sup>, Krishna Epari<sup>31</sup>, Q. Nguyen<sup>32</sup>, Nikolajs Zeps<sup>33,34,35</sup>, Massimo Falconi<sup>8</sup>, Michele Simbolo<sup>1</sup>, Giovanni Butturini<sup>8</sup>, George Van Buren II<sup>23</sup>, Stefano Partelli<sup>8</sup>, Matteo Fassan<sup>1</sup>, Australian Pancreatic Cancer Genome Initiative<sup>‡</sup>, Kum Kum Khanna<sup>5</sup>, Anthony J. Gill<sup>7,27</sup>, David A. Wheeler<sup>22</sup>, Richard A. Gibbs<sup>22</sup>, Elizabeth A. Musgrove<sup>3</sup>, Claudio Bassi<sup>8</sup>, Giampaolo Tortora<sup>9</sup>, Paolo Pederzoli<sup>8</sup>, John V. Pearson<sup>5, 6</sup>, Nicola Waddell<sup>5,6§</sup>, Andrew V. Biankin Biankin<sup>3,4,7,29,36§</sup> & Sean M. Grimmond<sup>37§</sup>

<sup>1</sup>ARC-Net Centre for Applied Research on Cancer, University and Hospital Trust of Verona, Verona 37134, Italy.

<sup>2</sup>Department of Pathology and Diagnostics, University and Hospital Trust of Verona, Verona 37134, Italy.

<sup>3</sup>Wolfson Wohl Cancer Research Centre, Institute of Cancer Sciences, University of Glasgow, Garscube Estate, Switchback Road, Bearsden, Glasgow G61 1QH, UK.

<sup>4</sup>West of Scotland Pancreatic Unit, Glasgow Royal Infirmary, Glasgow G31 2ER, UK.

<sup>5</sup>QIMR Berghofer Medical Research Institute, Herston Road, Brisbane 4006, Australia.

<sup>6</sup>Queensland Centre for Medical Genomics, Institute for Molecular Bioscience, The University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia.

<sup>7</sup>The Kinghorn Cancer Centre, Cancer Division, Garvan Institute of Medical Research, University of New South Wales, 384 Victoria St, Darlinghurst, Sydney, New South Wales 2010, Australia.

<sup>8</sup>Department of Surgery, Pancreas Institute, University and Hospital Trust of Verona, Verona 37134, Italy.

<sup>9</sup>Medical Oncology, University and Hospital Trust of Verona, Verona, Italy.

- <sup>10</sup>Department of Pathology, General Hospital of Treviso, Department of Medicine, University of Padua, Italy.
- <sup>11</sup>Department of Medicine, Section of Endocrinology, University and Hospital Trust of Verona, Verona, Italy.
- <sup>12</sup>The University of Queensland, School of Medicine, Brisbane 4006, Australia.
- <sup>13</sup>Pathology Queensland, Brisbane 4006, Australia.
- <sup>14</sup>Royal Brisbane and Women's Hospital, Department of Gastroenterology and Hepatology, Brisbane 4006, Australia.
- <sup>15</sup>Institute of Health Biomedical Innovation, Queensland University of Technology, Brisbane, Australia.
- <sup>16</sup>School of Environmental & Life Sciences, University of Newcastle, Ourimbah, New South Wales 2258, Australia.
- <sup>17</sup>Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Centre for Cancer Bioinformatics, Peking University Cancer Hospital & Institute, Beijing 100142, China.
- <sup>18</sup>Department of Surgery, Princess Alexandra Hospital, Ipswich Rd, Woollongabba, Queensland 4102, Australia.
- <sup>19</sup>Department of Anatomical Pathology, St Vincent's Hospital, Sydney, New South Wales 2010, Australia.
- <sup>20</sup>Academic Unit of Surgery, School of Medicine, College of Medical, Veterinary and Life Sciences, University of Glasgow, Glasgow Royal Infirmary, Glasgow G4 OSF, UK.
- <sup>21</sup>Department of Pathology, Queen Elizabeth University Hospital, Greater Glasgow & Clyde NHS, Glasgow G51 4TF, UK.
- <sup>22</sup>Department of Molecular and Human Genetics, Human Genome Sequencing Center, Baylor College of Medicine, One Baylor Plaza, MS226, Houston, Texas 77030-3411, USA.
- <sup>23</sup>Michael E. DeBakey Department of Surgery and The Elkins Pancreas Center, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030-3411, USA.
- <sup>24</sup>Children's Hospital at Westmead, Westmead, New South Wales 2145, Australia.
- <sup>25</sup>Children's Medical Research Institute, The University of Sydney, Westmead, New South Wales 2145, Australia.
- <sup>26</sup>Department of Surgery, Royal North Shore Hospital, St Leonards, Sydney, New South Wales 2065, Australia.
- <sup>27</sup>University of Sydney, Sydney, New South Wales 2006, Australia.
- <sup>28</sup>Tissue Pathology and Diagnostic Oncology, Royal Prince Alfred Hospital, Camperdown, New South Wales 2050, Australia.
- <sup>29</sup>Department of Surgery, Bankstown Hospital, Eldridge Road, Bankstown, Sydney, New South Wales 2200, Australia.
- <sup>30</sup>School of Medicine, Western Sydney University, Penrith, New South Wales 2175, Australia.
- <sup>31</sup>Department of Surgery, Fremantle Hospital, Alma Street, Fremantle, Western Australia 6160, Australia.

<sup>32</sup>Department of Gastroenterology, Royal Adelaide Hospital, North Terrace, Adelaide, South Australia 5000, Australia.

<sup>33</sup>School of Surgery M507, University of Western Australia, 35 Stirling Highway, Nedlands, Western Australia 6009, Australia.

<sup>34</sup>St John of God Pathology, 12 Salvado Rd, Subiaco, Western Australia 6008, Australia.

<sup>35</sup>Bendat Family Comprehensive Cancer Centre, St John of God Subiaco Hospital, Subiaco, Western Australia 6008, Australia.

<sup>36</sup>South Western Sydney Clinical School, Faculty of Medicine, University of New South Wales, Liverpool, New South Wales 2170, Australia.

<sup>37</sup>University of Melbourne Centre for Cancer Research, University of Melbourne, Melbourne, 3010, Victoria, Australia.

†Present addresses: Oncology Research Centre, Vrije Universiteit Brussel, Brussels, Belgium (I.R.); Pancreatic Cancer Translational Research Group, Adult Cancer Program, Prince of Wales Clinical School, Lowy Cancer Research Centre, UNSW, Sydney, Australia (A.M.).

‡A list of participants and their affiliations is provided in the Supplementary Information.

\*These authors contributed equally to this work.

§These authors jointly supervised this work.

**The diagnosis of pancreatic neuroendocrine tumours (PanNETs) is increasing owing to more sensitive detection methods, and this increase is creating challenges for clinical management. We performed whole-genome sequencing of 102 primary PanNETs and defined the genomic events that characterize their pathogenesis. Here we describe the mutational signatures they harbour, including a deficiency in G:C>T:A base excision repair due to inactivation of *MUTYH*, which encodes a DNA glycosylase. Clinically sporadic PanNETs contain a larger-than-expected proportion of germline mutations, including previously unreported mutations in the DNA repair genes *MUTYH*, *CHEK2* and *BRCA2*. Together with mutations in *MEN1* and *VHL*, these mutations occur in 17% of patients. Somatic mutations, including point mutations and gene fusions, were commonly found in genes involved in four main pathways: chromatin remodelling, DNA damage repair, activation of mTOR signalling (including previously undescribed**

***EWSRI* gene fusions), and telomere maintenance. In addition, our gene expression analyses identified a subgroup of tumours associated with hypoxia and HIF signalling.**

Pancreatic neuroendocrine tumours (PanNETs) are the second most common epithelial neoplasm of the pancreas and have a mortality rate of 60%<sup>1</sup>. The World Health Organization (WHO) classification, which assesses the proliferative fraction of neoplastic cells, divides PanNETs into three groups: low grade (G1), intermediate grade (G2), and high grade (G3). While G3 tumours are invariably lethal, 90% of PanNETs are grade G1 or G2. These have an unpredictable clinical course that varies from indolent to highly malignant. Our current understanding of the molecular pathology of G1 and G2 PanNETs is insufficient for their clinical management, where the challenge is to predict the aggressiveness of individual tumours in order to identify patients who will benefit from early aggressive therapy and to minimize harm from the inadvertent overtreatment of patients with indolent disease.

PanNETs are usually sporadic but also occur as part of three hereditary syndromes: multiple endocrine neoplasia type 1 (MEN-1), von Hippel-Lindau syndrome (VHL), and occasionally tuberous sclerosis complex (TSC)<sup>1</sup>. Somatic mutations of *MEN1* occur in 35% of PanNETs<sup>1-3</sup>. Recent expression profiling and exome sequencing have highlighted the importance of activated mTOR signalling as a druggable mechanism in 14% of patients<sup>3,4</sup> and, although the mTOR inhibitor Everolimus is approved by the FDA for the treatment of advanced PanNET<sup>5</sup>, it is not yet possible to use molecular analysis to select patients who will benefit. In addition, the apoptotic regulator *DAXX* or the chromatin modifier *ATRX* are mutated in up to 40% of PanNETs<sup>3,6</sup>, where they promote alternative lengthening of telomeres (ALT) and chromosomal instability<sup>7,8</sup>.

Our comprehensive molecular analysis of 102 clinically sporadic PanNETs defines their molecular pathology and identifies several novel candidate mechanisms that activate mTOR signalling, including novel gene fusion events. We have uncovered an important role for germline *MUTYH* variants through a novel G:C>T:A mutational signature. In addition, we have identified a larger-than-anticipated germline contribution to clinically sporadic PanNETs, delineating future challenges in the clinical assessment of susceptibility.

### **Genomic landscape of PanNETs**

The study workflow is illustrated in Extended Data 1. Patients were recruited and consent for genomic sequencing obtained as part of the International Cancer Genome Consortium (<http://www.icgc.org>). All cases were classified according to WHO criteria<sup>1</sup>. The cohort

included 98 resected PanNETs (Supplementary Table 1 and Supplementary Table 2) with an average tumour content of 82%<sup>9</sup> with 36 tumours classified as G1, 57 as G2, and 5 as G3. Four additional PanNETs and one colon cancer from a patient with MUTYH-associated polyposis (MAP) were sequenced to validate mutational signatures (Supplementary Table 3). Matched pairs of tumour and normal DNA were used for whole-genome sequencing (WGS) (average 38× normal, 61× tumour) and high-density single nucleotide polymorphism (SNP) arrays (Supplementary Table 3), and orthogonal testing estimated that the accuracy of somatic calls exceeded 99% (Supplementary Tables 4, 5). Structural rearrangements were detected by integrating discordant read pairs, soft-clipping, split read and *de novo* assembly of non-mapping reads<sup>10,11</sup> (Supplementary Table 6). Copy number events were identified using genome alteration print (GAP) analysis of SNP arrays<sup>12</sup> and recurrent focal and arm length gains and losses using GISTIC<sup>13</sup> (Supplementary Table 7).

### Mutational mechanisms in PanNET

PanNETs have a lower mutation burden (0.82 per megabase, range 0.04–4.56) than their exocrine counterpart<sup>11</sup> (pancreatic ductal adenocarcinoma: mean 2.64, range 0.65–28.2), with the 98 PanNETs analysed here containing 258,678 high-confidence somatic point mutations and indels. Non-negative matrix factorization<sup>14</sup> defined five robust mutational signatures (Fig. 1a, Extended Data Fig. 2 and Supplementary Table 8), including the unknown aetiology ‘Cosmic signature 5’<sup>14</sup> reported in many tumour types, deamination, *APOBEC* (also known as *AID*), *BRCA* and a previously undescribed signature. The germline heterozygous *APOBEC3A–3B* deletion, which has been implicated in *APOBEC*-induced mutation(s) in breast cancers<sup>15</sup>, appears to play a minimal role in PanNETs, as only 2 out of 13 carriers exhibited a predominant *APOBEC* signature (Fig. 1a). One tumour with a highly prominent *BRCA* deficiency signature (more than 2 mutations per Mb), and the same widespread genomic instability pattern recently described in *BRCA*-deficient breast<sup>16</sup>, pancreatic ductal<sup>11</sup>, ovarian<sup>17</sup> and oesophageal carcinomas<sup>10</sup>, harboured a pathogenic germline *BRCA2* mutation (Fig. 1c).

A novel mutational signature, composed of G:C>T:A transversions, predominated in five PanNETs (range 0.2–4.2 mutations per Mb; Fig. 1a, Extended Data Fig. 2), which bore a known pathogenic or novel inactivating germline mutation in the base-excision-repair gene *MUTYH*, coupled with loss of heterozygosity (LOH; Fig. 1d). This mutational signature or pattern of biallelic *MUTYH* inactivation was not observed in 100 pancreatic ductal adenocarcinomas using the same analysis pipeline<sup>11</sup>. Germline biallelic inactivation of

*MUTYH* causes the autosomal recessive *MUTYH*-associated colorectal polyposis syndrome and is associated with somatic G:C>T:A transversions in the *APC* gene, the driver of colorectal polyps<sup>18</sup>.

To verify our findings, we used amplicon sequencing of *MUTYH* on 62 additional PanNETs and identified three tumours bearing pathogenic germline mutations coupled with LOH. Whole-genome analysis revealed that these three tumours displayed the same novel mutational signature (Fig. 1b, Extended Data Fig. 3). Interestingly, the two missense *MUTYH* mutations identified in the Italian cohort (c.536A>G, p.Y179C; c.1187G>A, p.G396D) are the most common MAP-linked variants in populations of European origin and have been shown to be founder mutations in a recent haplotype analysis of 80 families with MAP from Italy and Germany<sup>19</sup>. Finally, whole-genome sequencing of a colonic tumour from a patient with MAP confirmed that this signature indicates *MUTYH* deficiency (Fig. 1b, Extended Data Fig. 3). These data suggest that, in addition to predisposing to colonic, gastric and a variety of non-gastrointestinal cancers<sup>20</sup>, *MUTYH* deficiency plays a role in PanNET.

Clusters of breakpoints were identified (Extended Data Fig. 4a) in nine tumours (9%), which displayed structural variations and copy-number changes consistent with chromothripsis<sup>21</sup> (Supplementary Table 9). Four of these nine tumours had recurrent catastrophic rearrangements on chromosome 11q (Extended Data Fig. 4b), all involving 11q13, and two of these rearrangements led to loss of *MEN1*. Notably, despite *TP53* being considered a hallmark of chromothripsis<sup>22</sup>, no *TP53* mutations were present in tumours with genomic catastrophes in PanNETs.

### Germline mutations

The discovery of germline deleterious mutations in *MUTYH* and *BRCA2* prompted us to screen the germlines of all patients for mutations in DNA damage repair genes or genes associated with hereditary syndromes (Supplementary Tables 10, 11). In the case of known neuroendocrine predisposition genes, six patients carried either known pathogenic or novel deleterious germline *MEN1* mutations (four frameshifts, one splice site mutation, and one copy-number loss). A single novel truncating *CDKN1B* germline mutation was identified (Q163X). *CDKN1B* germline mutations cause multiple endocrine neoplasia type 4<sup>23</sup> and somatic mutations occur in intestinal neuroendocrine tumours<sup>24</sup>. A single, novel, pathogenic germline mutation was identified within *VHL*, a negative regulator of hypoxia signalling that

promotes neuroendocrine proliferation and PanNETs<sup>25</sup>. In every case, germline alterations were coupled with somatic LOH.

The DNA-damage repair gene *CHEK2*, a known tumour suppressor in breast cancer and other cancer types, had predicted damaging germline variants in four individuals (4%): a nonsense mutation (c.58C>T, Q20X), a 15-base pair in-frame deletion (c.246–260del, p.D77–E82del), a missense mutation in exon 2 implicated in prostate cancer predisposition (rs121908702, c.844G>A, p.E282K)<sup>26</sup> and a missense variant in exon 4 (c.529G>C, p.D177H). Mutation modelling predicted these variants to be damaging (Extended Data Fig. 5a, b). To investigate the functional consequences of *CHEK2* mutations, we generated a panel of *Flag*–*CHEK2* constructs encoding the wild-type, P85L, D177H, E282K and  $\Delta 77$ –82 variants. The conservative mutant P85L showed expression and kinase activity equivalent to the wild type, but cells expressing the predicted pathogenic mutants D177H, E282K and  $\Delta 77$ –82 showed low kinase activity due to a combination of reduced protein expression and kinase inactivity (both autophosphorylation and phosphorylation of the substrate CDC25C) (Extended Data Fig. 5c–j).

### Somatic driver mutations

A total of 15,751 somatic coding mutations (7,703 non-silent) were detected in 2,787 genes (Supplementary Tables 4, 5). Sixteen significantly and recurrently mutated genes were defined using IntOGen<sup>27</sup> analysis ( $Q < 0.1$ ) (Extended Data Fig. 6a, Supplementary Table 12). Consistent with previous reports, *MEN1* was the most frequently mutated of these genes, present in 37% of tumours (Extended Data Fig. 6b). Mutually exclusive inactivating mutations of *DAXX* and *ATRX* were found in 22 and 11 samples, respectively, including a structural rearrangement of *ATRX*. The mechanistic target of rapamycin (mTOR) pathway genes *PTEN* ( $n = 7$ ) and *DEPDC5* ( $n = 2$ ) were also significantly mutated. *PTEN* mutations were mutually exclusive with mutations in *TSC1* ( $n = 2$ ) and *TSC2* ( $n = 2$ ), which encode other negative regulators of mTOR signalling (Supplementary Table 5). As mutations of the tumour suppressor gene *DEPDC5* have not been previously described in PanNETs, we surveyed an additional 62 cases (Supplementary Table 14) and identified a further 2 tumours that harboured *DEPDC5* mutations, which were again mutually exclusive to *PTEN* ( $n = 3$ ) and *TSC2* mutations ( $n = 3$ ) (Supplementary Table 15). Consistent with the literature, deleterious *TP53* mutations were uncommon ( $n = 3$ ). Among the genes not meeting the significance threshold by IntOGen, the histone modifier *SETD2* was mutated in five samples (Extended Data Fig. 6a, Supplementary Table 5). Similar to previous observations in renal



cell carcinoma<sup>28</sup>, we observed multiple independent *SETD2* mutations in presumed subclones of one tumour (a nonsense at 3%, a missense at 14% and a frameshift at 11% allelic frequency), suggesting strong selection for *SETD2* inactivation in that particular tumour.

### Copy number changes

Copy number analysis revealed four discrete groups of patients based on arm length copy number patterns (Extended Data Fig. 7a). These were classified into: 1) recurrent pattern of whole chromosomal loss (RPCL); 2) limited copy number events, many of which were losses affecting chromosome 11; 3) polyploidy; and 4) aneuploidy (Extended Data Fig. 7a, b). Notably, the RPCL subtype consistently presented loss of specific chromosomes (Extended Data Fig. 7) and was significantly enriched in G2 PanNETs ( $P = 0.0247$ ,  $\chi^2$  test). The polyploid group had the highest somatic mutation rate ( $P \leq 0.002$ , Mann–Whitney test) with an average of 1.98 mutations per Mb (Extended Data Fig. 7c).

Recurrent regions of gain and loss (Supplementary Table 7, Extended Data Fig. 7d) included broad regions of loss containing the known neuroendocrine tumour suppressors *MEN1* (chromosome 11q13.1) and *CDKN2A* (chromosome 9q21.3), whereas focal losses highlighted potential tumour suppressor roles for *EYAI* (chromosome 8q13.3; a known target of *MEN1*<sup>29</sup>), *FMBT1* (chromosome 3p21.1; encoding a key component of histone modification machinery implicated in cancer<sup>30</sup>) and *RABGAP1L* (chromosome 1q25.1; frequently deleted in neurofibromas<sup>31</sup>). Significant, recurrently amplified regions included *PSPN* (chromosome 19p13.3), a member of the glial cell line-derived neurotrophic factor family that activates phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha (PIK3CA) signalling via RET and is upregulated in thyroid medullary cancers<sup>32,33</sup>; and *ULK1* (chromosome 12q24.33), a serine-threonine kinase that is involved in mTOR-regulated autophagy<sup>34</sup>.

### Chromosomal rearrangements

Structural rearrangements are less common in PanNETs (mean, 29 events per tumour; range 3–216) (Supplementary Table 6) than in pancreatic ductal adenocarcinoma (119 per tumour; range 15–558)<sup>11</sup>. Inactivation of tumour suppressor genes through rearrangement occurred in *MTAP* ( $n = 4$ ), *ARID2* ( $n = 5$ ), *SMARCA4* ( $n = 3$ ), *MLL3* ( $n = 3$ ), *CDKN2A* ( $n = 1$ ), and *SETD2* ( $n = 1$ ). Rearrangements can also create oncogenic drivers through in-frame gene fusions. We identified 66 somatic fusions capable of expressing in-frame chimaeric genes (Supplementary Table 6). The *EWSR1* gene was involved in fusion events in three PanNETs.

Two tumours possessed in-frame *EWSR1–BEND2* fusion genes, which were expressed as mRNA (Fig. 2a, b), defining *BEND2* as a novel *EWSR1* fusion partner. Of note, *BEND2* was recently reported to be a fusion partner of *MNI*<sup>35</sup>. One tumour contained an *EWSR1* exon 7–*FLII* exon 6 gene fusion confirmed by RNA sequencing (RNA-seq) (Fig. 2c). *EWSR1–FLII* fusions occur in sarcomas<sup>36,37</sup>, where *EWSR1* exon 7–*FLII* exon 6 is the most common (Type I)<sup>38</sup>, but sarcomas rarely arise in the pancreas or gastrointestinal tract<sup>39</sup>. Unlike PanNET, sarcomas typically exhibit a distinctive, high-grade, undifferentiated round cell morphology along with strong membrane expression of the glycoprotein CD99. The three tumours that contained *EWSR1* fusions had morphological and immunophenotypical features typical of PanNETs, had absent or weak staining for CD99, and lacked any clinicopathological features of Ewing sarcoma. Inactivation of tumour suppressor genes (*STAG2*, *TP53*, and *CDKN2A*) and specific chromosomal copy-number alterations (gains of chromosome 1 and 8q), which are common in Ewing sarcomas<sup>40–42</sup>, were absent in these three cases. The three fusion events involved the region of *EWSR1* that is most susceptible to breakage or translocation in a variety of soft tissue tumours and Ewing sarcoma, and can be detected by fluorescent *in situ* hybridization (FISH) using break-apart probes for *EWSR1* (Fig. 2a–c). By applying FISH to the 62 cases of the validation cohort, we identified an additional PanNET with positive split signals due to an *EWSR1* exon 7–*FLII* exon 5 fusion (Type II) confirmed at the mRNA level (Fig. 2d). This latter tumour had strong immunostaining for CD99 and mutations in *MEN1*, *ATRX* and *TSC2*.

### Telomere integrity and PanNET molecular subtypes

Telomere repeat content was quantified using whole-genome sequencing data, and ALT was assessed using C-tailing qPCR in 86 cases (Extended Data Fig. 8a). In total, 22 of 26 *ATRX* or *DAXX* mutant tumours displayed ALT, and in *DAXX* mutations were more frequent (19/22) than *ATRX* mutations (3/22), in contrast to *in vitro* studies in which *ATRX* alterations are more prevalent<sup>43</sup>. Biallelic inactivation of *ATRX* or *DAXX* through LOH was strongly associated with an increase in telomere length ( $P < 0.0001$ , Mann–Whitney test; Extended Data Fig. 8b, c). *MEN1* somatic mutations were also associated with increased telomere length ( $P < 0.0001$ , Mann–Whitney test) (Extended Data Fig. 8d, e), suggesting that *MEN1* has a role in chromosome maintenance.

To better understand the consequences of ALT, we compared somatic telomere content with copy-number and structural variation patterns (Fig. 3). Genomic catastrophes and *EWSR1* fusions were associated with short telomeres, which is consistent with

observations that telomere exhaustion plays a role in chromothripsis and bridge–fusion–break events in solid cancers. Surprisingly, tumours with ALT were strongly associated with the PanNET RPCL phenotype, with 16 out of 21 tumours with RPCL displaying ALT. Previously, ALT tumours have been reported to undergo recurrent regions of gene copy gain and loss in a panel of human cancer cell lines *in vitro*<sup>43</sup>; in contrast, whole-chromosome loss of specific chromosomes predominates in PanNETs.

### **Integrated analysis of PanNET cancer pathways**

RNA sequencing of 30 cases revealed 3 groups of PanNET tumours (Extended Data Fig. 9a), which were similar to three previously described expression subtypes termed insulinoma, intermediate and metastasis-like (MLP)<sup>44</sup> (Extended Data Fig. 9b). One group was similar to the intermediate subtype. The group most similar to the MLP subtype contained differential expression of genes associated with hypoxia and HIF signalling (Extended Data Fig. 9c–e).

Four pathways were commonly altered by mutation in PanNETs. Perturbations in these pathways may potentially define clinically relevant subtypes that could be used to direct stratified therapeutic approaches (Fig. 4): i) DNA damage repair: germline-damaging variants of the base-excision-repair *MUTYH* gene and the homologous recombination genes *CHEK2* and *BRCA2* were present in 11% of patients. ii) Chromatin remodelling: *MEN1*, *SETD2*, *ARID1A* and *MLL3* were recurrently inactivated, and these mutations are likely to drive widespread transcriptional dysregulation. iii) Telomere maintenance: upregulation of *TERT* and telomere lengthening is a well-established pro-survival mechanism in solid tumours. *MEN1* binds the *TERT* promoter and influences the machinery that controls telomere integrity<sup>45</sup>. Consistent with previous reports, inactivating mutations in *DAXX* or *ATRX* were present in one third of PanNETs and correlated strongly with somatic telomere repeat content and telomere length (Extended Data Fig. 8b, c). Tumours harbouring *DAXX* or *ATRX* mutations were associated with a poor prognosis in the G2 subgroup (Extended Data Fig. 10a, b), and correlated significantly ( $P=0.0214$ ) with mutations in mTOR regulators (Supplementary Table 13). Tumours with unaltered telomere length had a better outcome (Extended Data Fig. 10c). iv) mTOR signalling activation: the role of mTOR in PanNETs has been well established<sup>3,4</sup>. Inactivating mutations in negative regulators of mTOR signalling (*PTEN*, *TSC2*, and *TSC1* and *DEPDC5* reported here) were present in 12% of patients, and were associated with a poor prognosis in the G2 subgroup of patients (hazard ratio = 6.85, 95% confidence interval = 1.14–41.7;  $P = 0.0353$ , Supplementary Table 1 and Extended Data Fig. 10d). These mutations may represent putative biomarkers for the selection of patients for

mTOR inhibitor therapy<sup>5</sup>. We uncovered three potential novel mTOR pathway activation mechanisms: inactivating mutations of the tumour suppressor *DEPDC5*, which encodes a subunit of the GATOR1 complex, a suppressor of mTOR signalling<sup>46</sup>; a putative mTOR activation mechanism involving *EWSRI* fusion genes; and amplification of the *RET* receptor ligand *PSPN*.

Moreover, inactivation of *MEN1*, which has a broad range of functions, directly influences all these four key processes<sup>45,47-49</sup> (Fig. 4). *MEN1* encodes the histone modifier Menin and its inactivation drives a variety of phenotypes including widespread transcriptional dysregulation via histone modification<sup>49</sup>, activation of mTOR through AKT expression<sup>48</sup>, suppression of homologous recombination DNA damage response genes<sup>47</sup> and dysregulation of TERT<sup>45</sup>.

## Conclusion

We have described the mutational landscape of PanNETs and the mutational signatures that underlie their pathogenesis, including a previously undescribed mutational mechanism involving *MUTYH* inactivation. We uncovered previously undescribed mTOR pathway activation mechanisms including *DEPDC5* inactivation and *EWSRI* fusion events. In addition, we identified subtypes of PanNET on the basis of global copy number profiles and gene mutations that have potential clinical utility. There are three key clinical considerations that these data bring to the fore. 1) The discovery of a larger-than-anticipated germline mutation contribution to PanNET development, particularly in patients without a family history, has implications for individuals that carry these mutations and have an increased but unquantifiable risk of disease. 2) The mutational status of *DAXX*, *ATRX* and mTOR pathway genes could be used to stratify the prognosis of intermediate grade (G2) PanNETs, the subgroup with the least predictable clinical behaviour. This calls for exploration of the clinical utility of this approach in prospective clinical trials. 3) The identification of previously undescribed mechanisms that activate mTOR signalling may lead to the development of biomarkers that could be used to predict therapeutic responsiveness to mTOR inhibitors such as everolimus, which are currently poorly defined.

Received 10 February; accepted 15 December 2016.

Published online XX 2017.

1. Bosman, F. T., Carneiro, F., Hruban, R. H. & Theise, N. D. *WHO Classification of Tumours of the Digestive System* 4th edn (International Agency for Research on Cancer, 2010).
2. Corbo, V. *et al.* MEN1 in pancreatic endocrine tumors: analysis of gene and protein status in 169 sporadic neoplasms reveals alterations in the vast majority of cases. *Endocr. Relat. Cancer* **17**, 771–783 (2010).
3. Jiao, Y. *et al.* DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**, 1199–1203 (2011).
4. Missiaglia, E. *et al.* Pancreatic endocrine tumors: expression profiling evidences a role for AKT-mTOR pathway. *J. Clin. Oncol.* **28**, 245–255 (2010).
5. Neychev, V. *et al.* Mutation-targeted therapy with sunitinib or everolimus in patients with advanced low-grade or intermediate-grade neuroendocrine tumours of the gastrointestinal tract and pancreas with or without cytoreductive surgery: protocol for a phase II clinical trial. *BMJ Open* **5**, e008248 (2015).
6. Elsässer, S. J., Allis, C. D. & Lewis, P. W. Cancer. New epigenetic drivers of cancers. *Science* **331**, 1145–1146 (2011).
7. Heaphy, C. M. *et al.* Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**, 425 (2011).
8. Marinoni, I. *et al.* Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology* **146**, 453–460 (2014).
9. Song, S. *et al.* qpure: A tool to estimate tumor cellularity from genome-wide single-nucleotide polymorphism profiles. *PLoS One* **7**, e45835 (2012).
10. Nones, K. *et al.* Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis. *Nat. Commun.* **5**, 5224 (2014).
11. Waddell, N. *et al.* Whole genomes redefine the mutational landscape of pancreatic cancer. *Nature* **518**, 495–501 (2015).
12. Popova, T. *et al.* Genome Alteration Print (GAP): a tool to visualize and mine complex cancer genomic profiles obtained by SNP arrays. *Genome Biol.* **10**, R128 (2009).

13. Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
14. Alexandrov, L. B. *et al.* Signatures of mutational processes in human cancer. *Nature* **500**, 415–421 (2013).
15. Nik-Zainal, S. *et al.* Mutational processes molding the genomes of 21 breast cancers. *Cell* **149**, 979–993 (2012).
16. Nik-Zainal, S. *et al.* Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* **534**, 47–54 (2016).
17. Patch, A. M. *et al.* Whole-genome characterization of chemoresistant ovarian cancer. *Nature* **521**, 489–494 (2015).
18. Al-Tassan, N. *et al.* Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat. Genet.* **30**, 227–232 (2002).
19. Aretz, S. *et al.* MUTYH-associated polyposis (MAP): evidence for the origin of the common European mutations p.Tyr179Cys and p.Gly396Asp by founder events. *Eur. J. Hum. Genet.* **22**, 923–929 (2014).
20. Vogt, S. *et al.* Expanded extracolonic tumor spectrum in MUTYH-associated polyposis. *Gastroenterology* **137**, 1976–1985 (2009).
21. Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
22. Rausch, T. *et al.* Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations. *Cell* **148**, 59–71 (2012).
23. Georgitsi, M. *et al.* Germline CDKN1B/p27Kip1 mutation in multiple endocrine neoplasia. *J. Clin. Endocrinol. Metab.* **92**, 3321–3325 (2007).
24. Francis, J. M. *et al.* Somatic mutation of CDKN1B in small intestine neuroendocrine tumors. *Nat. Genet.* **45**, 1483–1486 (2013).
25. Lubensky, I. A. *et al.* Multiple neuroendocrine tumors of the pancreas in von Hippel-Lindau disease patients: histopathological and molecular genetic analysis. *Am. J. Pathol.* **153**, 223–231 (1998).

26. Dong, X. *et al.* Mutations in CHEK2 associated with prostate cancer risk. *Am. J. Hum. Genet.* **72**, 270–280 (2003).
27. Gonzalez-Perez, A. *et al.* IntOGen-mutations identifies cancer drivers across tumor types. *Nat. Methods* **10**, 1081–1082 (2013).
28. Gerlinger, M. *et al.* Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* **366**, 883–892 (2012).
29. Li, B.E. *et al.* Distinct pathways regulated by menin and by MLL1 in hematopoietic stem cells and developing B cells. *Blood.* **122**, 2039–2046 (2013).
30. Tang, M. *et al.* The malignant brain tumor (MBT) domain protein SFMBT1 is an integral histone reader subunit of the LSD1 demethylase complex for chromatin association and epithelial-to-mesenchymal transition. *J. Biol. Chem.* **288**, 27680–27691 (2013).
31. Asai, A. *et al.* High-resolution 400K oligonucleotide array comparative genomic hybridization analysis of neurofibromatosis type 1-associated cutaneous neurofibromas. *Gene* **558**, 220–226 (2015).
32. Baba, T. *et al.* Persephin: A potential key component in human oral cancer progression through the RET receptor tyrosine kinase-mitogen-activated protein kinase signaling pathway. *Mol. Carcinog.* **54**, 608–617 (2015).
33. Lindahl, M. *et al.* Human glial cell line-derived neurotrophic factor receptor alpha 4 is the receptor for persephin and is predominantly expressed in normal and malignant thyroid medullary cells. *J. Biol. Chem.* **276**, 9344–9351 (2001).
34. Alers, S., Löffler, A. S., Wesselborg, S. & Stork, B. Role of AMPK-mTOR-Ulk1/2 in the regulation of autophagy: cross talk, shortcuts, and feedbacks. *Mol. Cell. Biol.* **32**, 2–11 (2012).
35. Sturm, D. *et al.* New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cell* **164**, 1060–1072 (2016).
36. Delattre, O. *et al.* Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. *Nature* **359**, 162–165 (1992).
37. May, W. A. *et al.* Ewing sarcoma 11;22 translocation produces a chimeric transcription factor that requires the DNA-binding domain encoded by FLI1 for transformation. *Proc. Natl Acad. Sci. USA* **90**, 5752–5756 (1993).

38. Sankar, S. & Lessnick, S. L. Promiscuous partnerships in Ewing's sarcoma. *Cancer Genet.* **204**, 351–365 (2011).
39. Stockman, D. L. *et al.* Malignant gastrointestinal neuroectodermal tumor: clinicopathologic, immunohistochemical, ultrastructural, and molecular analysis of 16 cases with a reappraisal of clear cell sarcoma-like tumors of the gastrointestinal tract. *Am. J. Surg. Pathol.* **36**, 857–868 (2012).
40. Brohl, A. S. *et al.* The genomic landscape of the Ewing Sarcoma family of tumors reveals recurrent STAG2 mutation. *PLoS Genet.* **10**, e1004475 (2014).
41. Crompton, B. D. *et al.* The genomic landscape of pediatric Ewing sarcoma. *Cancer Discov.* **4**, 1326–1341 (2014).
42. Tirode, F. *et al.* Genomic landscape of Ewing sarcoma defines an aggressive subtype with co-association of STAG2 and TP53 mutations. *Cancer Discov.* **4**, 1342–1353 (2014).
43. Lovejoy, C. A. *et al.* Loss of ATRX, genome instability, and an altered DNA damage response are hallmarks of the alternative lengthening of telomeres pathway. *PLoS Genet.* **8**, e1002772 (2012).
44. Sadanandam, A. *et al.* A cross-species analysis in pancreatic neuroendocrine tumors reveals molecular subtypes with distinctive clinical, metastatic, developmental, and metabolic characteristics. *Cancer Discov.* **5**, 1296–1313 (2015).
45. Lin, S. Y. & Elledge, S. J. Multiple tumor suppressor pathways negatively regulate telomerase. *Cell* **113**, 881–889 (2003).
46. Bar-Peled, L. *et al.* A Tumor suppressor complex with GAP activity for the Rag GTPases that signal amino acid sufficiency to mTORC1. *Science* **340**, 1100–1106 (2013).
47. Fang, M. *et al.* MEN1 is a melanoma tumor suppressor that preserves genomic integrity by stimulating transcription of genes that promote homologous recombination-directed DNA repair. *Mol. Cell. Biol.* **33**, 2635–2647 (2013).
48. Wang, Y. *et al.* The tumor suppressor protein menin inhibits AKT activation by regulating its cellular localization. *Cancer Res.* **71**, 371–382 (2011).



49. Matkar, S., Thiel, A. & Hua, X. Menin: a scaffold protein that controls gene expression and cell signaling. *Trends Biochem. Sci.* **38**, 394–402 (2013).

50. Landrum M. J *et al.* ClinVar: public archive of interpretations of clinically relevant variants. *Nucleic Acids Res.* 44(D1):D862-8 (2016). doi: 10.1093/nar/gkv1222

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

**Acknowledgements** We thank E. Missiaglia, S. Beghelli, N. Sperandio, G. Bonizzato, S. Grimaldi, F. Pisani, C. Cantù, G. Zamboni and P. Merlini for assistance at the ARC-Net Research Centre and Verona University; 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; D. Gwynne and D. Stetner for support at the Queensland Centre for Medical Genomics; and The Kinghorn Centre for Clinical Genomics for genome sequencing of validation samples. Funding support was from: Italian Ministry of Research (Cancer Genome Project FIRB RBAP10AHJB); Associazione Italiana Ricerca Cancro (AIRC n. 12182); Fondazione Italiana Malattie Pancreas – Ministero Salute (CUP\_J33G13000210001); National Health and Medical Research Council of Australia (NHMRC; 631701, 535903, CDF 1112113, PRF 1025427, SRF 455857, 535903); The Queensland State Government Smart State National and International Research Alliances Program (NIRAP); Institute for Molecular Bioscience/University of Queensland; The Royal Australasian College of Physicians, Sidney Catalyst, NHMRC, Pancare Australia; 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; Avner Nahmani Pancreatic Cancer Research Foundation; R.T. Hall Trust; Petre Foundation; Philip Hemstritch Foundation; Gastroenterological Society of Australia (GESA Senior Research Fellowship); Royal Australasian College of Surgeons (RACS); Royal Australasian College of Physicians (RACP); Royal College of Pathologists of Australasia (RCPA); QIMR Berghofer Medical Research; The Keith Boden Fellowship (K.N.); NHGRI U54 HG003273; CPRIT grant RP101353-P7; Wellcome Trust Senior Investigator Award (103721/Z/14/Z); CRUK Programme (C29717/A17263 and C29717/A18484); CRUK Glasgow Centre (C596/A18076); CRUK Clinical Training Award (C596/A20921); Pancreatic Cancer UK Future Research Leaders Fund; The Howat Foundation; and the University of Glasgow.

**Author Contributions** Biospecimens were collected at affiliated hospitals and processed at each biospecimen core resource centre. Investigator contributions are as follows: A.S., D.K.C., Nicola.W., A.V.B., S.M.G. (concept and design); A.S., D.K.C., Nicola.W., A.V.B., S.M.G. (project leaders); A.S., D.K.C., K.N., V.Co., Nicola.W., A.V.B., S.M.G. (writing team); K.N., A.-M.P., P.B., R.T.L., A.L.J., B.R., S.C., M.C.J.Q, P.J.W., S.H.N., I.D., A.P.D.T, M.V.D., L.L., A.Mal., M.M., M.D.J., J.Hu., L.A.C., V.Ch., A.M.N., M.Pa., M.Pi., C.J.S.,

A.P., I.R., C.T., V.Ch, A.Maw., E.S.H., E.K.C., A.C., J.A.L., N.B.J., F.D., M.C.G., J.S.S., N.D.M., K.E., N.Q.N., N.Z., M.Fal., M.Fas., G.B., S.P., W.E.F., A. Malp., A. Maw., G.V.B., D.A.W., R.A.G., E.A.M., A.B., C.B., G.T., P.P., A.V.B. (sample collection, processing, quality control & clinical annotation); A.S., D.K.C., J.G.K., A.J.G, A.V.B. (clinico-pathological analyses and interpretation); V.L.J.W., B.A.L. (colon sample collection and clinical annotation); A.S., B.R., I.C., P.C., J.G.K., M.Fas., A.J.G (pathology assessment); V.Co., D.K.M., M.Sc., M.Si., D.A., C.V., T.J.C.B., A.N.C., I.H., S.I., S.McL., C.N., E.N., E.A., S.Be., M.Si. (sequencing); O.H., R.A.D., L.M.S.L., M.L., H.A.P., R.R.R., J.V.P. (telomere analysis); K.N., A.M.P., P.B., R.T.L., A.L.J., A.Maf., S.Ba., K.O.S., S.S., M.C.J.Q., P.J.W., M.J.A., J.L.F., F.N., Nick.W., O.H., S.H.K., C.L., S.W., Q.X., J.W., M.Pi., M.C., J.V.P., Nicola.W., S.M.G. (bioinformatics); K.K.K, J.Ha. (protein modelling); J.L.H., K.K.K (functional validation of CHEK2 variants); A.P.D.T. (revision of fusion cases and FISH analysis); A.S., D.K.C., K.N., V.Co., P.B., P.P., N.B.J., F.D., Nicola.W., S.M.G., A.V.B. (data interpretation). All authors have read and approved the final manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.S. ([aldo.scarpa@univr.it](mailto:aldo.scarpa@univr.it)), A.V.B. ([andrew.biankin@glasgow.ac.uk](mailto:andrew.biankin@glasgow.ac.uk)) and S.M.G. ([sean.grimmond@unimelb.edu.au](mailto:sean.grimmond@unimelb.edu.au)).

**Reviewer Information** *Nature* thanks S. Chanock and the other anonymous reviewer(s) for their contribution to the peer review of this work.

**Figure 1 | Mutational signatures in pancreatic neuroendocrine tumours. a**, Five signatures (A–E) were identified in 98 PanNET samples. Signature A is previously undescribed and has been named *MUTYH* because all tumours dominated by this signature carried a germline inactivating mutation in *MUTYH* with concurrent loss of the wild-type allele; signatures B to E have been previously described and are reported with their given names. Tumours with a high *MUTYH*, *APOBEC*, *BRCA* or Age signature showed a higher number of mutations per megabase. **b**, Validation of *MUTYH* signature in four additional PanNETs and one colon tumour. Three PanNETs and the colon tumour contained a dominant *MUTYH* signature and harboured a germline-damaging *MUTYH* mutation with concurrent loss of the wild-type allele, and one PanNET with a benign *MUTYH* variant did not contain the signature. **c**, The tumour with the *BRCA* signature contained a *BRCA2* germline variant (R3052W) and genomic instability as seen in pancreatic ductal adenocarcinoma. **d**, Variants displayed above the *MUTYH* protein were associated with the *MUTYH* signature and contained somatic biallelic inactivation of the *MUTYH* gene (highlighted in bold); those below were not and are either benign or showed no loss of the wild-type allele. Asterisks show variants reported as pathogenic by ref.<sup>50</sup> (IDs #5294 & #5293). Variants predicted by

SIFT as deleterious and by Polyphen as probably damaging are underlined. All mutations shown are for the transcript ENST00000372098.3 and protein ENSP00000361170.3.

**Figure 2 | *EWSR1* gene fusions in pancreatic neuroendocrine tumours.** **a, b,** Somatic fusion events between *EWSR1* and *BEND2* confirmed in RNA by capillary electrophoresis of RT–PCR products. Representative sections showing typical PanNET morphology (stained with haematoxylin and eosin, H&E) and immunoreactivity for the neuroendocrine marker chromogranin A (inset), lack of immunostaining for CD99, and positive *EWSR1* split signals (arrowheads) detected with FISH. **c,** Somatic fusion event between *EWSR1* and *FLII* confirmed by RNA-seq. Representative sections showing typical PanNET morphology (H&E) and immunoreactivity for chromogranin A (inset), faint immunoreactivity for CD99, and positive *EWSR1* split signals (arrowheads) at FISH. **d,** Somatic fusion event between *EWSR1* and *FLII* confirmed by capillary electrophoresis of RT–PCR products. Representative sections showing typical PanNET morphology (H&E) and immunoreactivity for chromogranin A (inset), strong immunoreactivity for CD99, and positive *EWSR1* split signals (arrowheads). Scale bar, 100  $\mu$ m. Insets, 600 $\times$  magnification.

**Figure 3 | Mutational processes in pancreatic neuroendocrine tumours.** **a,** Telomere length was estimated using whole-genome sequencing in 98 PanNETs. The relative telomere length in each tumour compared to the matched normal is shown as  $\log_2$ . Twenty-four tumours contained telomeres that were 1.5 $\times$  longer than matched normal DNA and 36 contained telomeres that were 1.5 $\times$  shorter than matched normal DNA. **b,** Most of the tumours with long telomeres had ALT and 13 contained genomes with large amounts of whole chromosome arm losses. Somatic mutations in *DAXX* or *ATRX* were strongly associated with increased telomere length ( $P < 0.0001$ , Mann–Whitney test). Tumours with short telomeres contained fewer mutations in *DAXX* and *ATRX* and more chromothripsis events or *EWSR1* gene fusions.

**Figure 4 | Core pathways in PanNETs.** The frequency of somatic mutations and copy number change are shown for key genes of the mTOR signalling, histone modification, altered telomere length and DNA damage repair pathways. *MEN1*, which is frequently mutated in PanNETs, is involved in all of these processes. Activating changes (red) include a copy number change of 6 or more and inactivating changes (blue) include copy number 1 or copy number two with LOH. Asterisks show genes that also contain germline variants predicted as pathogenic. *ATRX* is located on chromosome X and showed point mutations in six males and four females, and in three of the females the inactivation was biallelic by LOH.

## **METHODS**

The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

### **Human research ethical approval**

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, 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. APCI: 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. Baylor College of Medicine: Institutional Review Board protocol numbers H-29198 (Baylor College of Medicine tissue resource), H-21332 (Genomes and Genetics at the BCM-HGSC), and H-32711 (Cancer Specimen Biobanking and Genomics).

### **PanNET patient and tissue cohort**

Patients were recruited and consent obtained for genomic sequencing through the ARC-Net Research Centre at Verona University, Australian Pancreatic Cancer Genome Initiative (APGI), and Baylor College of Medicine as part of the ICGC ([www.icgc.org](http://www.icgc.org)). A patient criterion for admission to the study was that they were clinically sporadic. This information was acquired through direct interviews with participants and a questionnaire regarding their personal history and that of relatives with regard to pancreas cancers and any other cancers during anamnesis. Clinical records were also used to clarify familial history based on patient indications.

Samples were prospectively and consecutively acquired through institutions affiliated with the Australian Pancreatic Cancer Genome Initiative. Samples from the ARC-Net biobank are the result of consecutive collections from a single centre.

All tissue samples were processed as previously described<sup>50</sup>. Representative sections were reviewed independently by at least one additional pathologist with specific expertise in pancreatic diseases. Samples either had full face frozen sectioning performed in optimal cutting temperature (OCT) medium, or the ends excised and processed in formalin to verify the presence of tumour in the sample to be sequenced and to estimate the percentage of neoplastic cells in the sample relative to stromal cells. Macrodissection was performed if required to excise areas that did not contain neoplastic epithelium. Tumour cellularity was determined using SNP arrays (Illumina) and the qpure tool<sup>9</sup>.

### **Sample size**

PanNET is a rare tumour type and the samples were collected via an international network. We estimate that with 98 unique patients in the discovery cohort, we will achieve 90% power for 90% of genes to detect mutations that occur at a frequency of ~10% above the background rate for PanNET (assuming a somatic mutation frequency of more than 2 per Mb)<sup>51</sup>.

### **Colon sample acquisition**

Cancer and matched normal colonic mucosa were collected at the time of surgical resection from the Royal Brisbane and Women's Hospital and snap frozen in liquid nitrogen. A biallelic germline mutation in the *MUTYH* gene was detected by restriction fragment length polymorphism analysis and confirmed by automated sequencing to be the G382D mutation (or ENST00000450313.5 G396D, ClinVar#5294) in both alleles<sup>52</sup>.

### **Immunohistochemistry**

The primary antibodies used for immunohistochemical staining were: cytokeratin 8/18 (5D3, Novocastra), chromogranin A (DAK-A3, Dako), and CD99 (O13, Biolegend). Antibodies and staining conditions have been described elsewhere<sup>39</sup>.

### **Sequencing and mutation analysis**

Whole-genome sequencing with 100-bp paired reads was performed with a HiSEQ2000 (Illumina). Sequence data were mapped to a GRCh37 using BWA and BAM files are available in the EGA (accession number: EGAS00001001732). Somatic mutations and germline variants were detected using a previously described consensus mutation calling

strategy<sup>11</sup>. Mutations were annotated with gene consequence using SNPeff. The pathogenicity of germline variants was predicted using cancer-specific and locus-specific genetic databases, medical literature, computational predictions with ENSEMBL Variant Effect Predictor (VEP) annotation, and second hits identified in the tumour genome. Intogen<sup>27</sup> was used to find somatic genes that were significantly mutated. Somatic structural variants were identified using the qSV tool as previously described<sup>10,11,17</sup>. Coding mutations are included in supplementary tables and all mutations have been uploaded to the International Cancer Genome Consortium Data Coordination Center.

### **Mutational signatures**

Mutational signatures were predicted using a published framework<sup>14</sup>. Essentially, the 96-substitution classification was determined for each sample. The signatures were compared to other validated signatures and the prevalence of each signature per megabase was determined.

### **Copy number analysis**

Somatic copy number was estimated using high density SNP arrays and the GAP tool<sup>12</sup>. Arm level copy number data were clustered using Ward's method, Euclidian distance. GISTIC<sup>13</sup> was used to identify recurrent regions of copy number change.

### **Telomere length**

The whole genome sequence data was used to determine the length of the telomeres in each sample using the qMotif tool. Essentially, qMotif determines telomeric DNA content by calculating the number of reads that harbour the telomere motif (TTAGG), and then estimates the relative length of telomeres in the tumour compared to the normal. qMotif is available online (<http://sourceforge.net/projects/adamajava>). Telomere length was validated by qPCR as previously described<sup>53</sup>.

### **RNA-seq and analysis**

RNASeq library preparation and sequencing were performed as previously described<sup>54</sup>. Essentially, sequencing reads were mapped to transcripts corresponding to ensemble 70 annotations using RSEM. 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<sub>2</sub> transformed. Genes without at least 1 c.p.m. in 20% of the sample were excluded from further analysis<sup>54</sup>. Unsupervised class discovery was performed using consensus clustering as implemented in

the ConsensusClusterPlus R package<sup>55</sup>. The top 2,000 most variable genes were used as input. Differential gene expression analysis between representative samples was performed using the R package ‘edgeR’<sup>56</sup>. Ontology and pathway enrichment analysis was performed using the R package ‘dnet’<sup>57</sup>. PanNET class enrichment using published gene signatures<sup>44</sup> was performed using Gene Set Variation Analysis (GSVA) as described previously<sup>54</sup>.

### **Validation of fusion transcripts**

Two strategies were used to verify fusion transcripts. For verification of *EWSR1–BEND2* fusions, cDNAs were synthesized using the SuperScript VILO cDNA synthesis kit (Thermofisher) with 1 µg purified total RNA. For each fusion sequence, three samples were used: the PanNET sample containing the fusion, the PanNET sample without that fusion, and a non-neoplastic pancreatic sample. The RT–PCR product were evaluated on the Agilent 2100 Bioanalyzer (Agilent Technologies) and verified by sequencing using the 3130XL Genetic Analyzer (Life Technologies). Primers specific for *EWSR1–BEND2* fusion genes are available upon request. To identify the *EWSR1* fusion partner in the case ITNET\_2045, a real-time RT–PCR translocation panel for detecting specific Ewing sarcoma fusion transcripts was applied as described<sup>58</sup>. Following identification of the fusion partner, PCR amplicons were subjected to sequencing using the 3130XL Genetic Analyzer.

### **Fluorescent *in situ* hybridization analysis**

*EWSR1* rearrangements were assayed on paraffin-embedded tissue sections using a commercial split-signal probe (Vysis LSI EWSR1 (22q12) Dual Colour, Break Apart Rearrangement FISH Probe Kit) that consists of a mixture of two FISH DNA probes. One probe (~500 kb) is labelled in SpectrumOrange and flanks the 5' side of the *EWSR1* gene, extending through intron 4, and the second probe (~1,100 kb) is labelled in SpectrumGreen and flanks the 3' side of the *EWSR1* gene, with a 7-kb gap between the two probes. With this setting, the assay enables the detection of rearrangements with breakpoints spanning introns 7–10 of the *EWSR1* gene. Hybridization was performed according to the manufacturer's instructions and scoring of tissue sections was assessed as described elsewhere<sup>59</sup>, counting at least 100 nuclei per slide.

### **Targeted sequencing**

Recurrently mutated genes identified by whole-genome sequencing were independently evaluated in a series of 62 PaNETs from the ARC-Net Research Centre, University of Verona. Four Ion Ampliseq Custom panels (Thermofisher) were designed to target the entire

coding regions and flanking intron–exon junctions of the following genes: *MEN1*, *DAXX*, *ATRX*, *PTEN* and *TSC2* (panel 1); *DEPDC5*, *TSC1* and *SETD2* (panel 2); *ARID1A* and *MTOR* (panel 3); *CHEK2* and *MUTYH* (panel 4). Twenty nanograms of DNA were used per multiplex PCR amplification. The quality of the obtained libraries was evaluated by the Agilent 2100 Bioanalyzer on chip electrophoresis. Emulsion PCR was performed with the OneTouch system (Thermofisher). Sequencing was run on the Ion Torrent Personal Genome Machine (PGM, Thermofisher) loaded with 316 or 318 chips. Data analysis, including alignment to the hg19 human reference genome and variant calling, was done using Torrent Suite Software v4.0 (Thermofisher). Filtered variants were annotated using a custom pipeline based on the Variant Effector Predictor (VEP) software. Alignments were visually verified with the Integrative Genomics Viewer: IGV v2.3 (Broad Institute).

#### **Modelling of CHEK2 and MUTYH germline variants *in silico***

There is no contiguous structure available for CHEK2, so we produced a model of isoform C using PDBid 3i6w<sup>60</sup> as a template for predicting the structure of sequence O96017. Modelling was carried out within the YASARA suite of programs<sup>61</sup> and consisted of an initial BLAST search for suitable templates followed by alignment, building of loops not present in selected template structure and energy minimization in explicit solvent. Modelling was carried out in the absence of a phosphopeptide ligand, which was added on completion by aligning the model with structure 1GXC and merging the ligand contained therein with the model structure. Similarly, MUTYH is represented by discontinuous structures and so this too was modelled using PDBids 3N5N and 4YPR as templates together with sequence NP\_036354.1. Having constructed both models, amino acid substitutions were carried out to make the wild-type sequences conform to the variants described above. Each substitution was carried out independently and the resulting variant structures were subject to simulated annealing energy minimization using the AMBER force field. The resulting energy-minimized structures formed the basis of the predictions.

#### **Functional analysis of CHEK2 germline variants *in vitro***

*CHEK2* site mutants were generated by site-directed mutagenesis of wild-type pCMV–FLAG CHEK2 (primer sequences in Supplementary Table 16). Proteins were expressed in HEK293T, a highly transfectable derivative of HEK293 cells that were retrieved from the cell culture bank at the QIMR Berghofer medical research institute. Cells were authenticated by STR profiling and were negative for mycoplasma. Transfected cells were lysed in NP-40 modified RIPA with protease and phosphatase inhibitors. Protein expression levels were



analysed by western blotting with anti-FLAG antibodies and imaging HRP luminescent signal on a CCD camera (Fuji) and quantifying in MultiGauge software (Fuji). Kinase assays were performed using recombinant GST–CDC25C (amino acids 200–256) as substrate, essentially as described<sup>62</sup>. Kinase assay quantification was performed by scintillation counting of excised gel bands in OptiPhase scintillant (Perkin Elmer) using a Tri-Carb 2100TR beta counter (Packard). Counts for each reaction set were expressed as a fraction of the wild type. All experiments were performed at least three times.

### Clinical correlations

The date of diagnosis and the date and cause of death for each patient 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. The hazard ratio and its 95% confidence interval were estimated using Cox proportional hazard regression modelling. The correlation between *DAXX* or *ATRX* mutational status and other clinico-pathological variables was calculated using the  $\chi^2$  test. Statistical analysis was performed using StatView 5.0 Software (Abacus Systems). Disease-specific survival was used as the primary endpoint.

### Data availability

Genome sequencing data presented in this study have been submitted to the European Genome-Phenome Archive under accession number EGAS00001001732

(<https://www.ebi.ac.uk/ega/search/site/EGAS00001001732>).

50. Biankin, A. V. *et al.* Pancreatic cancer genomes reveal aberrations in axon guidance pathway genes. *Nature* **491**, 399–405 (2012).
51. Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
52. Kambara, T. *et al.* Role of inherited defects of MYH in the development of sporadic colorectal cancer. *Genes Chromosom. Cancer* **40**, 1–9 (2004).
53. O’Callaghan, N. J. & Fenech, M. A quantitative PCR method for measuring absolute telomere length. *Biol. Proced. Online* **13**, 3 (2011).
54. Bailey, P. *et al.* Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47–52 (2016).

55. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
56. 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).
57. Fang, H. & Gough, J. The ‘dnet’ approach promotes emerging research on cancer patient survival. *Genome Med.* **6**, 64 (2014).
58. Lewis, T. B., Coffin, C. M. & Bernard, P. S. Differentiating Ewing’s sarcoma from other round blue cell tumors using a RT-PCR translocation panel on formalin-fixed paraffin-embedded tissues. *Mod. Pathol.* **20**, 397–404 (2007).
59. Rossi, S. *et al.* EWSR1-CREB1 and EWSR1-ATF1 fusion genes in angiomatoid fibrous histiocytoma. *Clin. Cancer Res.* **13**, 7322–7328 (2007).
60. Cai, Z., Chehab, N. H. & Pavletich, N. P. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. *Mol. Cell* **35**, 818–829 (2009).
61. Krieger, E., Koraimann, G. & Vriend, G. Increasing the precision of comparative models with YASARA NOVA--a self-parameterizing force field. *Proteins* **47**, 393–402 (2002).
62. Bell, D. W. *et al.* Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. *Int. J. Cancer* **121**, 2661–2667 (2007).

### **Extended Data Figure 1 | Flow chart of the experiments performed on 160 PanNETs.**

The chart shows the workflow of analyses conducted on the discovery set of 98 PanNETs and on the validation set of an additional 62 PanNETs and 1 colorectal cancer. CNA, copy-number analysis.

### **Extended Data Figure 2 | Five mutation signatures in pancreatic neuroendocrine tumours.**

**a**, Stability plot indicates there are five mutation signatures (>0.9). **b**, The profile of the five mutational signatures (A–E) and what function has been assigned to these signatures (MUTYH, APOBEC, BRCA, Age and ‘Signature 5’).

### **Extended Data Figure 3 | Validation of the novel signature in additional MUTYH carriers.**

**a**, Four PanNet samples, three of which harboured a pathogenic *MUTYH* germline variant, and a colon tumour with a pathogenic *MUTYH* mutation underwent WGS to validate

the association of *MUTYH* biallelic inactivation with the *MUTYH* mutation signature. **b**, Family pedigree of the patient with colon cancer. The 64-year-old male patient with colon cancer was identified as a candidate for *MUTYH* mutation analysis owing to the presence of two synchronous cancers in the proximal colon, each arising in a contiguous tubulovillous adenoma, as well as approximately 50 adenomatous polyps predominantly in the caecum and ascending colon. The index patient's brother presented with colorectal cancer at 45 years of age and his sister presented with colorectal cancer at 64 years of age and with breast cancer at 59 years of age. The index patient's son had polyps removed at 36 years of age. Mutation signature analysis was performed using the 98 discovery PanNET samples and the colon and 4 PanNET validation samples. **c**, Stability plot showing the solution for the five mutational signatures ( $>0.75$ ). **d**, The profile of the five mutational signatures (A–E) and what function has been assigned to these signatures (*MUTYH*, *APOBEC*, *BRCA*, Age and 'Signature 5'). **e** The contribution of each signature (mutations per Mb) and proportion of the signatures in each tumour are shown.

**Extended Data Figure 4 | Structural rearrangements in pancreatic neuroendocrine tumours.** **a**, Top, the number and type of somatic structural rearrangements in each tumour. Bottom, tumours with more events tended to have longer telomeres. **b**, Two methods were used to determine clusters of somatic structural rearrangement breakpoints. Orange squares, chromosomes with a significant cluster of events as determined by a goodness-of-fit test against the expected distribution ( $P < 0.0001$ , Kolmogorov–Smirnov test). Blue squares, chromosomes deemed to harbour a high number of breakpoints because they had a chromosomal breakpoint per Mb rate that exceeded the 75th percentile of the chromosomal breakpoint per Mb rate for the cohort by five times the interquartile range. Red squares, chromosomes for which both of these criteria were met. Clusters of events were reviewed and nine tumours were found to harbour regions of chromothripsis. **c**, Recurrent chromothripsis for chromosome 11 was detected in four tumours. The chromothripsis event caused loss of the *MEN1* gene locus in two of these samples.

**Extended Data Figure 5 | Functional analysis of *CHEK2* variants.** **a**, *CHEK2* structure indicating the positions of the germline variants. Mutations are highlighted by rendering as magenta sticks with protein domains coloured as indicated in the adjacent keys. The model includes a superimposed phosphopeptide (red). **b**, A summary of the *CHEK2* variants and their predicted impact on protein structure. To functionally test the *CHEK2* variants, a panel of FLAG–*CHEK2* constructs encoding P85L,  $\Delta 77$ –82, D177H and E282K was generated. **c**,

**d**, FLAG western blot of transfected HEK293T whole cell lysates (**c**) or anti-FLAG immunoprecipitates (**d**) showed that, compared to the wild type, there was normal expression of P85L but reduced expression of  $\Delta 77-82$ , D177H and E282K. **e**, Assessment of kinase activity of *CHEK2* variants. Immunoprecipitated proteins were incubated either with GST alone (–) or with GST–CDC25C amino acids 200–256 (+) in the presence of  $\gamma$ -P32 ATP. Input and kinase activity were assessed by film radiography (top) and coomassie staining (bottom). Immunoprecipitates of  $\Delta 77-82$ , D177H and E282K had significantly reduced kinase activity in terms of both autophosphorylation and phosphorylation of CDC25C whereas the activity of P85L was normal. **f**, Quantification of expression levels by western blotting expressed as a fraction of wild type. Data points represent independent experiments. Error bars are mean  $\pm$  s.e.m. **g**, **h**, Quantification of kinase activity. P32 counts for CDC25C (**g**) and CHEK2 (**h**) bands were scintillation counted. Corresponding bands from untransfected controls were used for background subtraction. Background-corrected P32 counts per minute were then standardized to wild type for each experiment. Data points represent independent experiments. Error bars are mean  $\pm$  s.e.m. **i**, **j**, Quantification of kinase activity relative to protein expression. Kinase activity (from **i** and **j**) was standardized to protein expression level (from **f**). D177H was not analysed in this manner owing to its very low expression level. Error bars are mean  $\pm$  s.e.m. Once the low expression level of  $\Delta 77-82$  is taken into account, it is evident that the expressed protein retains normal kinase activity. On the other hand, E282K is kinase defective even after adjusting for its reduced expression. D177H expression is so low that it is not possible to reliably correct kinase activity for relative expression level, so it is unclear whether D177H is kinase dead as well as unstable. Data are summarized in Supplementary Table 16.

**Extended Data Figure 6 | Recurrently mutated genes in pancreatic neuroendocrine tumours.** **a**, The number of SNVs and indels within the genome of each patient ( $n = 98$ ) is shown in the histogram. The driver plot displays the somatic mutations in key genes or those identified as significantly mutated (Intogen  $Q < 0.1$ ). *SETD2* is also reported, although its  $Q$  value was 0.15, as it was recurrently inactivated in six samples and multiple independent deleterious *SETD2* mutations were observed in one tumour (a nonsense present at 3%, a missense at 14%, and a frameshift at 11%; only the nonsense is shown but the case is highlighted with a black arrow), suggesting strong selection for *SETD2* inactivation in that tumour. **b**, Somatic mutations in *MEN1* are predominantly nonsense mutations or insertions—

deletions causing frame shifts and premature protein termination, and occur throughout the protein.

**Extended Data Figure 7 | Genome characteristics of PanNETs.** Copy number was determined using Illumina SNP arrays in a cohort of 98 PanNETs. **a**, Copy number events were mainly comprised of whole chromosome arm loss or gain. Cluster analysis of the chromosome arm level copy number state stratified the tumours into four subtypes. Group 1: recurrent pattern of whole chromosomal loss, affecting specific chromosomes (1, 2, 3, 6, 8, 10, 11, 15, 16 and 22); group 2: samples with a limited number of events, many with loss affecting chromosome 11; group 3: polyploid tumours, with gain of all chromosomes; and group 4: aneuploid tumours, containing predominantly whole chromosome gains affecting multiple chromosomes). **b**, The proportion of bases within the genome affected by copy number change. **c**, The mutations per Mb (SNPs and small insertion deletions). **d**, GISTIC analysis showing recurrent gains (red) and losses (blue) of the entire cohort.

**Extended Data Figure 8 | Telomere length is associated with somatic mutations.** Whole genome sequence data were used to estimate telomere length in PanNETs relative to the matched normal sample. **a**, Telomere length estimated by whole-genome sequencing correlated with the telomere length calculated from qPCR ( $R^2 = 0.8091$ ). Values are plotted on a  $\log_{10}$  scale. **b–e**, Boxplots were used to show the association of relative telomere length and *DAXX* or *ATRX* and *MEN1* mutation status. Mann–Whitney tests were used to determine significant associations ( $P < 0.05$ ). **b**, **c**, Tumours harbouring *DAXX* or *ATRX* mutations contain longer telomeres. **d**, Tumours harbouring *MEN1* mutations contain longer telomeres. **e**, Telomere length is shown in relation to *DAXX* or *ATRX* and *MEN1* somatic mutations.

**Extended Data Figure 9 | RNA-seq of PanNET tumours.** Unsupervised clustering, network and gene enrichment analysis for available RNA-seq data identify PanNET subgroups associated with hypoxia and metabolic reprogramming. **a**, Unsupervised clustering identified three distinct PanNET subgroups (1–3). **b**, A gene signature defining three expression groups previously described in PanNETs showed enrichment of expression of the intermediate-group genes<sup>43</sup> in Group 1 and the metastasis-like PanNET (MLP) genes<sup>43</sup> in Group 3. **c**, Network analysis identified a significant sub-network of genes differentially expressed between Group 3 and other groups (Group 1 and Group 2). Red nodes represent genes upregulated in Group 3 and green nodes represent genes upregulated in other groups. Shaded areas represent network communities. **d**, Gene enrichment analysis for genes belonging to the sub-network shown in **b**. **e**, Heatmap showing the differential expression of

genes belong to the identified sub-network. Somatic mutations in some of the recurrently mutated genes are shown (*MEN1*, *DAXX*, *ATRX* and members of the mTOR pathway: *DEPDC5*, *MTOR*, *PTEN*, *TSC1* and *TSC2*).

**Extended Data Figure 10 | Genomic events associated with outcome.** Kaplan–Meier survival curves. **a, b**, Tumours harbouring *DAXX* or *ATRX* mutations had a poor prognosis in the whole cohort (**a**) and in the G2 cohort (**b**). **c**, Tumours with telomere lengths that were neither short or long had a better prognosis. **d**, Tumours harbouring mutations in genes that activate the mTOR pathway had a poor prognosis in the G2 cohort ( $\chi^2$  test was used in all instances).