bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Title:

Epigenetic landscape of pancreatic neuroendocrine tumours reveals distinct cells of origin and means of tumour progression

Authors: Annunziata Di Domenico[+, 1], Christodoulos P. Pipinikas[‡], Renaud Sylvain Maire[+], Konstantin Bräutigam[+], Cedric Simillion[†], Matthias S. Dettmer[+], Erik Vassella[+], Christina Thirlwell[‡], Aurel Perren[+] and Ilaria Marinoni[+]*

+ Institute of Pathology, University of Bern, Murtenstrasse 31, 3008 Bern Switzerland

1 Graduate School for Cellular and Biomedical Sciences, University of Bern, 3010 Bern

‡ UCL Cancer Institute, 72, Huntley Street, London. WC1E 6JD

† Bioinformatics and Computational Biology, University of Bern, Baltzerstrasse6, 3012 Bern

Keywords (5): PanNET, DNA methylation, α-cell, β-cell, origin.

Grant support: Marie Heim-Vögtlin SNF (PMPDP3_164484) and Tumour Forschung Bern to Ilaria Marinoni. Swiss Cancer League (KLS-4227-08-2017) to Aurel Perren. Cancer Research UK and Experimental Cancer Medicine Centre to Christina Thirlwell and Christodoulos Pipinikas.

Correspondence:

Ilaria Marinoni; Institute of Pathology, University of Bern, Murtenstrasse 31, 3008 Bern (CH), Tel +41316324991, Fax +41316324995, Email: ilaria.marinoni@pathology.unibe.ch

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ABSTRACT

Recent data suggest that Pancreatic Neuroendocrine Tumours (PanNETs) originate from α- or β-cells of the islets of Langerhans. The majority of PanNETs are non-functional and do not express cell-type specific hormones. We examined whether tumour DNA methylation (DNAme) profiling combined with genomic data could identify cell of origin and reveal pathways involved in PanNET progression. We analysed genome-wide DNAme data of 125 PanNETs and sorted α- and β-cells. To confirm cell identity, we investigated ARX and PDX1 expression. Based on epigenetic similarities, PanNETs clustered in α-like, β-like and intermediate tumours. The epigenetic similarity to α-cells progressively decreased in the intermediate tumours, which presented unclear differentiation. Specific transcription factor methylation and expression varied in the respective α/β-tumour groups. Depending on DNAme similarity to α/β-cells, PanNETs have different mutational spectra, stage of the disease and prognosis, indicating potential means of PanNET progression.

Keywords: cell of origin, alpha, beta, methylation, expression, epigenetic, neuroendocrine, progression, pancreatic, tumour

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Introduction

Pancreatic neuroendocrine tumours (PanNETs) are tumours of the islets of Langerhans. The cell of origin is unclear, and mechanisms associated with progression are largely unknown. Surgery is currently the only curative option; however, 5-year disease free survival is approximately 50% in patients following an R0 resection[1]. To date, there is no validated risk prediction tool to accurately guide follow up and select patients at high risk of recurrence who might benefit from adjuvant therapy[2]. PanNETs are clinically and genetically heterogeneous; ~40% of patients present with mutations in either *DAXX* or *ATRX* and *MEN1*, which encode for proteins involved in epigenetic regulation[3]. PanNET with mutations in *DAXX* or *ATRX* have a shorter disease free survival compared to wild type tumours[4].

The islets of Langerhans include five different cell types producing specific hormones: glucagon is produced by α-cells, insulin by β-cells, somatostatin by δ-cells, ghrelin by ε-cells and pancreatic polypeptide by PP-cells. Only a minority of PanNETs are functional, leading to clinical syndromes due to inadequate hormone secretion. The majority of functional PanNETs are insulinomas. Whether functional tumours and non-functional tumours originate from the same cell type remains uncertain.

Recent studies of gene expression and master regulator analysis alongside investigation of super-enhancer signatures have suggested both α- and β-cells as two possible cells of origin for non-functioning (NF)-PanNETs[5–8]. On the other hand, Sadanandam *et al*. reported that a group of aggressive PanNET, namely, "metastasis-like primary", have a phenotype characterised by "stemness" transcripts compared to well differentiated tumours, also suggesting a common progenitor cell origin[5]. Similarly, based on the identification of master regulator proteins, de-differentiation and acquisition of stem cell characteristics seem to be one of the pathways associated with tumour progression[6].

The cell of origin in cancer refers to the normal cell that acquires the initial cancer-promoting genetic hit(s). During development, cell lineage fate is determined by cell-type specific transcription factor (TF) expression, which in turn is dependent on the type of epigenetic markers that are located at the relative regulatory regions (e.g. super-enhancer activation)[9]. The five endocrine cell types derive from a common endocrine precursor, which has segregated from a ducto-endocrine bipotent cell population[10]. The TFs, *Pax4* (Paired Box 4) and *Arx* (Aristaless Related Homeobox), are required for β- and α-cell fates, respectively. Lineage decision is determined via cross-inhibitory interactions[11]. *Pdx1* (Pancreatic And Duodenal Homeobox 1) expression becomes restricted to cells at the stage of initiating insulin expression and, in the pancreatic islets of Langerhans, remains up-regulated exclusively in β-cells[12]. Integrative analysis of human epigenomes including histone modification patterns, DNA accessibility, DNA methylation and RNA expression has revealed that disease- and trait-associated genetic variants are enriched in tissue-specific epigenomic marks[13]. In the context of tumour biology, epigenetic states of cell lineages shape the vulnerability for specific genetic alterations and thereby reveal a distinct class of lineage-associated cancer genes[14–16]. Therefore, determining the cell of origin is crucial to understand tumour specific carcinogenesis and progression[17]. Cancer DNA methylation profiles have been utilised to determine the cell of origin of several tumour types[18,19]. In this study, we set out to determine the

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

putative cell of origin of PanNETs through DNA methylation analysis. We also identified genetic driver mutations specific to different cells of origin which are related to clinical outcome. Based on our findings we propose a new model of PanNET origin and progression.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Results

DNA methylation signature of PanNETs showed similarities to α- and β-cells

We analysed the methylomes of 125 PanNETs and of isolated normal α- and β-cells[20] to determine which of these is the cell of origin (cohort details are summarised in table 1 and reported in table S1). A flow chart of the analysis performed is provided in figure S1. The comparison between sorted normal α- (n=2) and β-cells (n=2) resulted in 2703 differentially methylated CpG sites (adj. p-value<0.001 and more than 20% difference between the mean β-values of each group - $|\Delta\beta|>0.2$, table S2), 2131 of them were retained, after filtering and normalization, in the tumour DNA methylation matrix. Phyloepigenetic analysis of the 2131 identified CpG sites using either normal β- or normal α-cells as roots, indicated clear hierarchical relationships between normal cells and tumour specimens (Fig. 1A and S2A). We defined PanNETs that grouped together with normal α- or β-cells as α-like or β-like PanNETs, respectively (Fig. 1A and table S1). We compared these phyloepigenetic groups to cell type similarity obtained from cell-type deconvolution analysis. We assigned each PanNET sample a degree of methylome similarity to each of the following cell types: α, β, acinar, ductal[20], fibroblastic pancreatic cells[21] and hematopoietic cells[22] (Fig. S2B). Notably, phyloepigenies that were close to normal α/β-cells showed similarity >65% and an average of 72% and 73% of similarity to α and β-cells, respectively (SD for α-like ±15% and for β-like ±9%, table S1). T-distributed stochastic neighbour embedding (*t*-SNE) analysis, using the identified 2131 CpGs and including normal cells and PanNET samples, showed a consistent segregation, according to the groups defined via similarity scores and phyloepigeny analysis (Fig. S2C). These results showed at least two groups of tumours with clear and high similarity to either α- or β-cells (Fig. 1A, S2A and S2C). The majority of tumours, however, revealed an intermediate differentiation profile. Intermediate tumours clustered between α- and β-cells and showed weak similarities to α- (as average 51 ±20 % of similarity) and very weak similarity to β-cells (as average 14 ±16 % of similarity), respectively.

Regulatory sites of α- and β-cell specific transcription factors were differentially methylated

We used previously published data to select 10 α- and 10 β-cell-type associated TFs[23–25] (table 2). As methylation of enhancers and regulatory states are important elements of epigenetic guidance of cell states, these regions were selected based on integrated ChIP-seq, DNA methylation and ATAC-seq experiments of normal islets[26]. This analysis resulted in 49 CpGs for the α-cell TFs and 51 CpGs for the β-cell TFs. Consensus clustering of the 125 PanNETs according to each cell-specific TF signature (k-means = 2, reflecting one cell-type enriched cluster and one non-specific cluster) showed clear and consistent cluster formation. The β-specific cluster separated very clearly and included all the β-like

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

tumours. Interestingly, the α-specific cluster showed a broader inclusion of tumours. While the α-like (phyloepigenies with >65% of similarities to alpha cells) tumours clustered mainly close to each other on one extreme, intermediate PanNETs were included to this group as well but progressively distant from the α-like tumours (Fig. 1B).

Alpha-, intermediate and β- like PanNETs have distinct genetic aberrations

Within the β-like PanNETs 13 of 14 were *MEN1/DAXX/ATRX* wild type. Alpha-like PanNETs were enriched for *MEN1* only mutations (11/19), 8/19 α-like PanNETs were *MEN1/DAXX/ATRX* wild type and none harboured *MEN1/DAXX/ATRX* mutations (Fig. 1A and table S1). Sixty-seven percent (62/92) of intermediate PanNETs harboured mutations in *MEN1* and/or *DAXX/ATRX* genes (intermediate-ADM) while the rest were wild type (intermediate-WT) (Fig. 1A and table S1).

Inferred copy number aberrations (CNAs) from HM450 signal intensities stratified the tumours into 3 subtypes (Fig. S3E). Copy number aberration (CNA) group 1 had few copy number events. Group 2 included samples with gains and losses. CNA group 3 showed the largest number of CNAs, predominantly with loss of multiple chromosomes (Fig. S3E). Notably, α- and β-like tumours presented few copy number events while the intermediate PanNETs showed increasingly copy number events (Fig. 1A).

DNA methylation signatures associated with cell of origin correlated with expression of hormones and transcription factors

To confirm the methylation-based cell lineage similarity, we performed immunostaining for ARX, PDX1, insulin and glucagon on a subset of samples for which formalin fixed paraffin-embedded tissue was available (n=39, table S1). Nuclear positivity for ARX and PDX1 was present in the expected distribution in normal islets (Fig. S4). We identified ARX positivity in 28/39 PanNETs, 6 of them also produced glucagon (table S1 and Fig. 1C). As expected, all of the insulin producing PanNETs expressed PDX1, except the single malignant insulinoma, which showed nuclear positivity for ARX but not for PDX1 (table S1 and Fig. 1C). None of the 39 tumours were double positive and only 4 PanNETs were double negative (table S1 and Fig. 1C).

Immunohistochemistry confirmed the same group segregation as was determined by the methylation analysis: 1) α-like tumours expressed ARX, a subset produced glucagon (2/3, 67%), 2), β-like tumours expressed both PDX1 and insulin, with only 1 sample positive for ARX and PDX1 negative (Fig. 1C), 3) 24/29 (83%) intermediate PanNETs were positive for ARX and rarely produced glucagon (4/22, 18%). A minority of intermediate tumours (4/29, 14%) were negative for both PDX1 and ARX and only one intermediate PanNET expressed PDX1 and insulin (Fig. 1C, table S1).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

These results indicate that cell type assigned by DNA methylation, corresponded to the cell type assigned by TF expression.

PDX1 and ARX immunopositivity correlates with mutation spectrum of PanNETs in an independent cohort

To validate the correlation of putative cell of origin to mutational status, we performed immunostaining for ARX, PDX1, insulin, glucagon and DAXX/ATRX on an independent cohort of 65 G1/G2 PanNETs (cohort details are summarised in table 3 and reported in table S3). We scored nuclear positivity for ARX in 34 samples (52%), for PDX1 in 21 samples (32%), 3 tumours (5%) showed strong double positivity and 13 (20%) were negative for both TFs (Fig2A and 2B). Notably, all the DAXX/ATRX negative samples expressed ARX, confirming the α-like tumour susceptibility for these mutations (Fig. 1A and 2C). While only a subset of ARX^+ tumours secreted glucagon (n=7, 20% of all the ARX^+ tumours, Fig. 2C), almost all the $PDX1^+$ tumours produced insulin (n=18, 95% of all $PDX1^+$ tumours, Fig. 2C).

We observed that advanced stage PanNETs either expressed ARX or none of the TFs and often showed DAXX/ATRX loss (Fig. 2C, S5A and table S3). Tumours positive for PDX1 had the lowest risk of relapse (Fig. S5B and C). However, DAXX/ATRX status did not improve stratification of ARX^+ samples for risk of relapse (Fig. S5B).

Epigenetic differentiation status determines clinical outcome of PanNETs

In order to determine clinical utility, we reviewed clinical outcome and prognosis for α-like, β-like and intermediate tumours. Fourteen of 19 α-like PanNETs were G1 and 14/18 (1 without data) were of low stage (T1 or T2, Fig. 1A and table S1). Similarly, 8/14 β-like PanNETs were G1 and 9/11 (3 without data) were T1 or T2 (Fig. 1A and table S1). Only one patient with an α-like tumour had relapse and none of the patients with a β-like tumour (median follow-up time 37.2 months, table S1 and Fig. 3A). Tumours that belonged to the intermediate phyloepigenies, were instead enriched for high stage (63/88 were T3 or T4, 4 without data) and higher grade (65/92 are G2) PanNETs had increased relapse risk (37/70, 22 had not available data - Fig. 1A, 3A and table S1). Disease free survival in intermediate tumours was significantly shorter than in α-like and β-like tumours ($p<0.001$, Fig. 3A). To assess if epigenetic profile can stratify patients better than TF and DAXX/ATRX expression we performed disease free survival analysis on all samples where DNA methylation and TF expression data were available. We determined time to relapse, using either expression of ARX, PDX1 and DAXX/ATRX or the epigenetic status as discriminant (Fig. 3B and C). Epigenetic profiles could predict more accurately the risk of relapse as shown in figures 3B and C.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Further, to identify a tumour dependent signature, suitable for stratification in clinical practice, we have compared the three cell-type specific clusters, α-like, β-like and intermediate. The comparison revealed 6364 unique differentially methylated sites (tables S4-S6; adj. p-value<0.001 and $|\Delta\beta|>0.2$). Consensus clustering of the 6364 CpG sites showed the maximal stability for k=4 (Fig. S6A). The existence of α-like and β-like clusters was confirmed (Fig. 3D and tableS1). The intermediate subgroup was divided into 2 further groups, one enriched for *DAXX/ATRX/MEN1* mutations, intermediate-ADM, (61/76 had mutations in at least 1 of the 3 genes; Fig. 3D and tableS1) and the other enriched for *DAXX/ATRX/MEN1* wild type samples, intermediate-WT (15/16, only 1 *DAXX/ATRX* mutated;, Fig. 3D and tableS1). PanNETs included in the intermediate-ADM and intermediate-WT showed comparable risk to relapse (Fig. S6B). To prove the stability of the signature, we repeated the analysis including 32 new G1/G2 PanNETs, for which DNA methylation data were public available[7] (table S7). After filtering and normalization processes, 6359/6364 of the previously identified differentially methylated sites were used in the consensus clustering algorithm. The results confirmed the 4 PanNET subgroups with risks to relapse similar to previous analysis (Fig. S6 C-E).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Discussion

This study identified at least two cells of origin for PanNETs (α-like / β-like) and demonstrated that DNA methylation analysis can discriminate α-like, β-like and intermediate (-ADM and -WT) PanNETs. Additionally, this study identified that relapses and metastases occurred most commonly in the intermediate (-ADM and -WT) PanNET groups.

Phyloepigenetic analysis of PanNETs, according to differentially methylated CpG sites between normal α- and β-cells, showed two clusters around normal α- and β- cells, which we named α-like and β-like. PanNET sample segregation was consistent in the *t*-SNE analysis. Using a similar approach, in a recent study, methylation-based subclasses of colorectal cancer were explained as clonal amplifications of one specific epigenotype, confirmed by enrichment of different mutations in the peculiar subtypes and by the investigation of methylation signatures in serial tumour xenografts and derived spheroids[18]. Similarly, in medulloblastoma subtypes with distinct developmental origin, DNA methylation signatures were able to stratify the tumours according to the specific subtype and partially resemble the cell of origin as well as acquired epigenetic changes during tumour progression[14,27,28]. Even if in the aforementioned studies tumours were not directly compared to the putative normal cells of origin, the epigenetic and mutational landscape of the malignancies reflected the tumour classes and cells of origin defined via *in vivo* studies. Altogether the data confirm the valuable use of DNA methylation profiles for the identification of tumour cells of origin.

Cell-type deconvolution analysis of methylome data from ductal, acinar, inflammatory and pancreatic stromal cells together with α- and β- cells found that α-like and β-like PanNETs are largely similar (at ~73%) to α- and β- cells, respectively. This figure is comparable to other studies: Houseman and Ince[29] demonstrated the application of their algorithm to estimate normal cell proportions in breast cancer heterogeneous tissues. Similarly, cell type deconvolution analysis based on DNA methylation for mantel cell lymphoma (MCL), showed a minimum of similarity between MCL and B cell samples of 40%[30]. When a similar algorithm (L1-regularized logistic regression) is used to classify cancers of unknown origin (CUPs) based on DNA methylation, a probability >30% was used to ascribe CUPs to a specific tumour type[31]. The deconvolution of the methylation estimates for PanNETs identified two important aspects of the tumour methylomes: that the results were not influenced by the composition of non-tumoral cells within the PanNET samples (Fig. S2B) and that the composition of early stage α-like and β-like PanNETs is abundant for either α- or β-cells (Fig. 1A).

Analysis of methylation status of specific TFs regulating α and β differentiation might help to better identify the intermediate groups of PanNET. While β-like samples separated very clearly according to β-cell TF methylation sites, α-cell similarity decreased gradually among the remaining samples. Seventy of seventy-six intermediate-ADM PanNETs clustered together with the α-like tumours (table S1). Twelve of sixteen intermediate-WT PanNETs did cluster neither with α-like nor with β-like tumours

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(table S1). Additionally, intermediate-ADM PanNETs were on average more similar to α- (53±17%) than β-cells (12±13%). Ninety-two percent of the intermediate-ADM tumours were positive for ARX (23/25) and none for PDX1. This might indicate that intermediate-ADM PanNETs are more related to α-cells rather than β-cells. The PanNET methylation sub-groups that we have identified reflect the groups A (tumours that most likely originate form α-cells), B (tumours that most likely originate form β-cells) and C (tumours with intermediate phenotype) recently described by Cejas *et al.*, via the analysis of super-enhancer signatures[8]. Comparable to our study, the analysis of TF expression via IHC performed by Cejas *et al.* revealed mutually exclusive expression of either ARX or PDX1 in the majority of the cases. Additionally, while PDX1 expression appears to be specific for benign and low stage tumours, ARX expression is retained at early and at advanced stages and only DAXX/ATRX status provide more information about stage and risk of disease progression[8].

We identified a clear correlation between driver mutation status and epigenetic profiles across all PanNETs. Alpha-like PanNETs and intermediate-ADM PanNETs harboured *MEN1* mutations, this is supported by the results of Chan *et al.* and Cejas *et al.*[7,8]. Fifty-eight percent of the clinically indolent α-like PanNETs were characterized by *MEN1* mutations only. *MEN1* inactivation, is an early event in PanNET progression[32,33], which enhances endocrine cell proliferation[34], hence it might be a tumour initiating event for the α-like PanNETs and for the intermediate-ADM. In turn Intermediate-ADM PanNETs might progress upon acquisition of DAXX/ATRX inactivation. In addition, CNA increased from α-like and β-like PanNETs to intermediate-ADM tumours. All these data together suggest a potential progression from α-like to intermediate-ADM PanNET. We cannot exclude, however, that the two intermediate PanNET clusters might originate from putative endocrine precursor cells.

During islet cell development, ARX is already expressed in endocrine precursors[11], furthermore different endocrine cell types, as α- and γ-cells share ARX expression[23,24]. Conversely PDX1 expression is restricted to differentiated β-cells, within the endocrine lineages[12]. Of the intermediate-WT tumours for which IHC data were available (n=4), one was positive for PDX1, one for ARX and 2 were negative for both TFs. Additionally, similarly to the β-like tumours, intermediate-WT PanNETs showed only few copy number events (group 1, table S1). Currently our data only weakly support progression from β-like to intermediate-WT PanNETs, nevertheless this possibility cannot be excluded.

While the vast majority of β-like PanNETs expressed PDX1 and insulin, the two malignant (N1 and/or M1) insulinomas of our cohort, showed one intermediate-ADM (*DAXX/ATRX* mutated) and one α-like (*DAXX/ATRX* wild type) methylation signatures (table S1). Only for the intermediate-ADM insulinoma, data on PDX1 and ARX were available; it resulted positive for ARX, and negative for PDX1 expression. Accordingly, the only malignant insulinoma of the Chan *et al.* cohort was mutated for *ATRX* and clustered with the intermediate-ADM tumours. In line with this observation, a recent study including 37 sporadic insulinomas (35 primary and 2 liver metastases) showed that all the five insulinomas which

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

metastasized were ARX positive and 4/5 had ALT activation (3 primary and 2 liver metastases)[33]. These data suggest that malignant insulinomas may arise from α-cells or stem-cells rather than β-cells. Under certain conditions and stimuli α-cells are able to trans-differentiate into β-cells[35–39]. Additionally, α-cell specific *Men1* knockout in mice leads to the development of glucagonomas which evolve into mixed glucagonoma/insulinoma to ultimately become insulinomas, possibly via trans-differentiation of the *Men1*-deleted α-cells[40,41]. Beta-like and/or PDX1 positive tumours are strongly enriched for benign insulinomas in both the first and second cohort. They generally show no mutations in any of the most commonly mutated genes for PanNETs (*DAXX*, *ATRX* and *MEN1*). These data confirmed the genetic difference between non-functioning tumours and insulinomas[5,42–44]. Of the thirteen wild type PanNETs obtained from Chan *et al.*[7], five were included in the β-like cluster (2 insulinomas and 3 NF-PanNETs), 6 in the intermediate-WT cluster and only 2 in the intermediate-ADM cluster. All the ADM-mutant PanNETs were included in the intermediate-ADM cluster (table S7). These data and the few β-like NF-PanNETs included in the cohort 1, demonstrate that occasionally also NF-PanNETs might originate from β-cells.

Our methylation data supports the possibility of two evolutionary pathways for PanNET development, originating from α- and β-cells (Fig. 3D). Beta-like PanNETs usually manifest as insulinomas. The α-like PanNETs are susceptible to *MEN1* mutations in early tumorigenesis. Tumour progression occurs upon *DAXX/ATRX* mutations, coupled with ALT activation and a characteristic CNA profile[4,45–47]. Progressive loss of differentiation might further predispose the tumours to enhanced proliferation and higher cell plasticity (Fig. 3D). Alternatively or additionally to this dedifferentiation model, potential endocrine progenitor cells might be the cell of origin of intermediate tumours (Fig. 3D).

We acknowledge that our study is based on static observations taken at one timepoint in each tumour's development. Sequential sampling or *in vivo* experiments would be able to determine real time tumour evolution and would be able to address the unanswered questions we propose here.

Clinically α-like, β-like and intermediate tumours have different outcomes. While α-like and β-like tumours are indolent, intermediate tumours (-ADM and -WT) have high risk of relapse. Interestingly, DAXX/ATRX status alone is not sufficient to discriminate between ARX^+ with high and low risk of relapse. Indeed, in the intermediate-WT group we observed the presence of ARX^+ tumours, with high risk of relapse. Only the DNA methylation profile is able to separate ARX^+ PanNET with low risk from high risk of relapse. The intermediate groups of PanNET are characterized by high risk of relapse but are molecularly different (ADM and WT) (Fig. 3D and S6D). Further work is needed to assess whether intermediate-ADM and intermediate-WT aggressive PanNETs have different origins or respond differently to therapies.

DNA methylation analysis clearly bears advantages over ChIPseq assays, it is easily performed on diagnostic routine FFPE specimens. We foresee a potential clinical use of epigenetic profiling for PanNETs similar to CNS tumor classification, able to define clinically relevant groups[28,48,49]. In addition,

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

299 DNA methylation remains stable in circulating tumor cells (CTCs) and cell-free DNA (cfDNA)
300 (reviewed in [50]), hence might be applied to monitor progression in individual patients in a non-invasive
301 liquid biopsy.

302 In conclusion DNA methylation analysis could be easily implemented in clinical practice to identify
303 patients with high relapse risk and those which might benefit from adjuvant therapy.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Methods

Patient cohorts

A cohort of primary PanNETs was assembled from two international centres; 19 samples previously analyzed[51] from UCL Cancer Institute (London, UK) and 26 from Institute of Pathology (Bern, Switzerland). All cases were classified according to WHO 2017 criteria[52]. TNM staging is based on the 8th edition UICC/AJCC. Inclusion criteria were histopathologic diagnosis of well-differentiated G1/G2 PanNETs, availability of tissue material and sufficient tumour purity (>70%). Seventeen of 45 tumour samples (38%) were classified as G1, 28/45 (62%) as G2 (table 1). Two samples derived from MEN1-patients. Median follow-up time is 51.5 months. All analyses were performed on Formalin-Fixed Paraffin-Embedded (FFPE) tissue specimens obtained from routine pathological work-up. Additional clinico-pathological characteristics are reported in table S1. The study on this cohort was approved by the local ethics committees (Bern: number 105/2015; London: number 09/H0722/27).

Clinical and molecular data of 80 PanNETs were provided by the International Cancer Genome Consortium (ICGC, https://icgc.org/, projects: PAEN-AU and PAEN-IT). Cohort features are summarised in table 1 and table S1. Genomic and molecular analysis of the samples were performed within Scarpa *et al.* study[45].

As immunohistochemical validation cohort, 65 PanNET samples were assembled at the Institute of Pathology (Bern, Switzerland). All cases were reclassified according to WHO 2017 criteria[52]. TNM staging is based on the 8th edition UICC/AJCC[53]. As for the previous cohorts, only well-differentiated G1/G2 PanNETs were included. Cohort features are summarised in table 3 and extensively reported in table S3.

A flow chart of the analysis performed on the 2 cohorts is provided in figure S1.

Immunohistochemistry (IHC) and Telomeric-Fluorescence In Situ Hybridization (Telo-FISH)

All samples were assessed for DAXX and ATRX expression via IHC and ALT activation via Telo-FISH on 2.5 µm sections prepared from a tissue microarray (TMA - Bern cohort and validation cohort) or whole block sections (London cohort) as previously described[4].

Similarly, 2.5 µm sections from TMAs or whole blocks were used for ARX (1:1500, R&D Systems, sheep, AF7068), PDX1 (1:100, R&D Systems, mouse, MAB2419), insulin (1:12000, Sigma Aldrich, mouse, I-2018) and glucagon (1:20000, Sigma Aldrich, mouse, G-2654) immunostainings. The immunostaining for all antigens was performed on an automated staining system (Leica Bond RX; Leica Biosystems, Nunningen, Switzerland). Antigen retrieval was performed by heating Tris30 buffer at 95° for 30 minutes. The primary antibody was incubated for 30 minutes at the specified dilutions.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Visualization was performed using Bond Polymer Refine Detection kit, using DAB as chromogen (3,3'-Diaminobenzidine).

Samples showing single cell positivity of any of the TFs were finally scored as negative, most of the cases were indeed strong positive for one of the two TFs with only few exceptions (single cell positivity is reported for each sample in tables S1 and S3). Only strong positivity for insulin and glucagon was considered for classification of the tumours as hormone producing PanNETs. Nuclear positivity for ARX and PDX1 was first assessed on normal islets, proving the selectivity for endocrine α- and β-cells (Fig. S4).

DNA methylation analysis

We extracted DNA from FFPE tissues according to manufacturer recommendations (QIAamp DNA minikit, Qiagen). Serial sections were cut and macrodissected using a razor blade upon histological evaluation (5 x 6µm), to make sure to achieve >70% tumour purity. We assessed DNA quality using the Illumina FFPE QC Kit. Ligation of FFPE DNA and bisulphite conversion were performed as described[54]. Efficiency of bisulphite conversion was confirmed by quantitative PCR as previously shown[54]. Converted DNA was processed on the HumMeth450 BeadChip (Illumina HM450). We analysed all the DNA methylation data included in this study using the ChAMP pipeline (v2.12.4, minfi method was used for raw data loading)[55–57]. Filtering was performed as implemented in the ChAMP pipeline[58]. Batch correction was performed using the ComBat algorithm as part of the ChAMP pipeline[59,60].

We identified differentially methylated sites between normal α– and β–cells according to the ChAMP pipeline[61,62]. To build phyloepigenetic trees, distances between samples were calculated according to pearson correlation and the neighbour-joining tree estimation was used within the ape (v5.3) R package[63]. The tSNE approach was performed as implemented in the R package tsne using the following parameters: perplexity=50, max_iter=5,000.

We used the RnBeads pipeline[64] for calculating cell type contribution according to the Houseman *et al.* method[65]. Sorted normal hematopoietic cell type data (granulocytes, CD4+ and CD8+ T cells, CD14+ monocytes, CD19+ B cells, CD56+ natural killer cells, neutrophils and eosinophils) were downloaded from Reinius *et al.*[22] (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE35069). Sorted acinar, duct, alpha and beta pancreatic cell DNA methylation profiles were downloaded from Jäkel *et al.*[20] (European Genome-Phenome Archive, https://ega-archive.org/, accession number: EGAS00001002533). DNA methylation data for sorted normal pancreatic fibroblast cells were downloaded from Xiao *et al.*[21] (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE80369). Marker selection was performed by

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

371 screening 10 000 CpG positions and the final number of selected cell type markers was set to 500 (as
372 for default).

373 Consensus clustering was performed following the ConsensusClusterPlus pipeline[66], samples were
374 clustered according to the hierarchical clustering algorithm, ward.D2 method was used for inner linkage
375 and average method was used for the final linkage.

376 Based on genomic position, the 450K probes were annotated with the chromatin state, as assigned to
377 normal pancreatic islets (obtained from the integration of ATAC-Seq, DNAme and ChIP-seq data)[26].
378 Probes associated with the cell specific transcription factors (TFs) and associated to the epigenetic states
379 "closed weak enhancer", "lowly-methylated weak enhancer", "open weak enhancer", "closed strong
380 enhancer", "open strong enhancer", "genic enhancer", "insulator" and "polycomb repressed states" were
381 included for looking at the specific TF methylation. Enhancer regions were associated to the nearest
382 gene in the genome using the GenomicRanges R package[67].

383

384 **Next Generation Sequencing**

385 We sequenced *MEN1, DAXX* and *ATRX* genes by semi-conductive sequencing using two Ion Torrent
386 AmpliSeq NGS custom made panels (Life Technologies), one for *ATRX* and one for *MEN1* and *DAXX,*
387 covering whole protein coding exons. Protein-coding exons were amplified by multiplex polymerase
388 chain reaction using 2 pools designed by the Ion AmpliSeq Designer and the Ion AmpliSeq Library kit
389 2.0 according to the manufacturer's recommendations (Life Technologies). Template preparation was
390 performed using the Ion Chef System. Sequencing was performed using the Ion S5. The Torrent Suite
391 5.10 platform was used for sequence alignment with the hg19 human genome reference. Variant calling
392 was performed with the variant caller and the IonReporter 5.10 software (Life Technologies). The
393 coverage depth was sufficient (quality criteria for a sample to be analysed was minimum 500 reads).
394 The 80 ICGC samples were sequenced by whole-genome sequencing (WGS) within the Scarpa *et al.*
395 study[45].

396

397 **Copy-number aberration analysis**

398 Genome-wide CNAs were inferred from HM450 signal intensities using the Conumee R package[68].
399 Chromosome bins and segments were measured (bin size: 50,000 to 5,000,000 bp; minimum number of
400 probes per bin: 15). CNA profile zero-threshold was manually adjusted according to FISH results for
401 the UB-UCL cohort (FISH for chr4 and chr17 was performed for detecting gains, FISH for MEN1 and
402 chr11 was used for detecting losses; see supplementary materials and Fig. S3A-D). Zero-threshold of
403 the CNA profiles obtained from the ICGC cohort was manually adjusted according to the results
404 obtained from Scarpa *et al.*[45]. Copy number (CN) of each chromosomal arm for each tumour was

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

obtained calculating the median of the relative copy number bins. Arm level copy number data were clustered using Ward's method, Euclidian distance.

Statistical analysis and graphic representation

Statistical analysis and graphical representations were performed within the R environment (v. 3.5.0)[69]. Specific packages used in the study and parameters selected are mentioned in the relative method chapters. For disease free survival analysis, the "survival" and "survminer" packages were used[70,71].

Data availability

The authors declare that all data supporting the findings of this study are available within the article and its Supplementary data and figures. The datasets generated during the current study (UB-UCL cohort) are available in the ArrayExpress repository (EMBL-EBI, https://www.ebi.ac.uk/arrayexpress/, accession number: E-MTAB-7924). The datasets analysed during the current study (ICGC cohort) are available in the ICGC repository (ICGC, https://icgc.org/, projects: PAEN-AU and PAEN-IT). The datasets of sorted normal hematopoietic cells are available in the GEO repository (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE35069). The datasets of sorted acinar, duct, alpha and beta pancreatic cells are available in the EGA repository (European Genome-Phenome Archive, https://ega-archive.org/, accession number: EGAS00001002533). The datasets of sorted normal pancreatic fibroblastic cells are available in the GEO repository (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE80369). Chan *et al.*[7] dataset is available in the GEO repository (Gene Expression Omnibus, http://www.ncbi.nlm.nih.gov/geo/, accession number: GSE117852).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

427 **Table 1: Patient cohort used for methylome analysis**

	UniBern cohort	UCL cohort	ICGC cohort	Total
Total number of patients	26	19	80	125
Sex				
Females	10	12	31	53
Males	16	7	49	72
Grade (WHO 2010)				
G1	7	10	32	49
G2	19	9	48	76
Tumour Stage (AJCC 8th ed.)				
T1	3	3	10	16
T2	8	6	18	32
T3	15	2	26	43
T4	0	0	26	26
No data	0	8	0	8
N stage				
N0	10	5	42	57
N1	13	3	36	52
No data	3	11	2	16
M stage				
M0	12	1	67	80
M1	10	1	13	24
No data	4	17	0	21
DAXX/ATRX				
Wild Type	8	11	53	72
Mutated	18	7	27	52
Not applicable	0	1	0	1
ALT				
Negative	8	11	42	61
Positive	16	3	29	48
Not Applicable	2	5	9	16
***MEN1* (somatic mutations)**				
Wild Type	12	13	49	74
Mutated	14	6	31	51
Hormone functionality				
F-PanNETs	8	2	12	22
NF-PanNETs	18	0	68	86
Not Applicable	0	17	0	17
Syndromic –MEN1	0	2	3	5

428

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

429 **Table 2: Specific cell-type transcription factors**

Alpha TFs	Beta TFs
ARX	*MA*
FEV	*PDX1*
HMGB3	*SMAD9*
IRX2	*CDKN1C*
LDB2	*TFCP2L1*
MAFB	*SIX3*
PGR	*SIX2*
PTGER3	*MNX1*
RFX6	*BMP5*
SMARCA1	*PIR*

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Table 3: Validation cohort

	Patient cohort
Total number of patients	65
Sex	
Females	28
Males	31
No data	6
Grade (WHO 2017)	
G1	34
G2	25
No data	6
Tumour Stage (AJCC 8th ed.)	
T1	29
T2	11
T3	15
T4	1
No data	9
N stage	
N0	30
N1	15
No data	20
M stage	
M0	40
M1	11
No data	14
DAXX/ATRX	
Positive	51
Negative	12
Not applicable	2
Hormone functionality	
F-PanNETs	20
NF-PanNETs	44
Syndromic –MEN1	1
ARX	
Positive	34
Negative	31
PDX1	
Positive	21
Negative	44

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

436 **Figure legends**

437

438 **Figure 1. PanNET methylomes resemble distinct endocrine cell lineages. A.** Phyloepigenetic
439 analysis of PanNET and normal α- and β-cell samples. A rooted tree was created with an arbitrary chosen
440 β-sample as the root and selecting the differentially methylated CpGs between sorted normal α- and β-
441 cell samples (n=2131, adj. p-value<0.001 and |Δβ|>0.2). Blue and orange scales show the degree of
442 similarity of each PanNET sample to α- and β-cells, respectively. The degree of similarity was
443 calculated via DNA methylation cell-type deconvolution comparing tumour methylomes to sorted
444 normal α- and β-cell methylomes. Each line represents a patient. Clinical and molecular features for
445 each sample are indicated. **B.** Consensus clustering of the 125 PanNETs according to the α- and β-cell
446 specific TF methylation sites (49 CpGs for β-cells and 51 CpGs for α-cells). The k mean value was set
447 to 2. Consensus cluster correlation is indicated according to the blue scale as depicted. Fractions of
448 similarity to α- and β-cells for each sample are reported at the bottom of each matrix as well as the cell-
449 type specific tumour groups. **C.** Schematic overview of the PanNET subtypes. Doughnut chart for all
450 samples showing cell specific PanNET subtypes, ARX and PDX1 expression, hormone production,
451 *MEN1* mutations and DAXX/ATRX loss.

452

453 **Figure 2. Expression of ARX, PDX1, insulin and glucagon can be used as surrogate for defining**
454 **tumour cell-type specificity. A.** Immunostaining for the different subtypes of PanNETs. From the top
455 to the bottom: tumour positive for ARX and glucagon; tumour positive for ARX; double positive tumour
456 for PDX1 and ARX; double negative tumour for ARX and PDX1; β-like tumour positive for PDX1 and
457 insulin. For the TFs only nuclear staining was considered for scoring. **B.** Venn diagram displaying the
458 total number of samples positive for ARX and/or PDX1or negative for both the transcription factors. **C.**
459 Schematic overview of the PanNET subtypes in cohort 2. Doughnut chart for all samples showing
460 glucagon, insulin, PDX1 and ARX positivity as well as loss of DAXX/ATRX.

461

462 **Figure 3. Epigenetic differentiation status defines clinically different PanNETs and draws two**
463 **possible ways of PanNET evolution. A.** Kaplan-Meier disease free survival of 98 patients (cohort 1)
464 stratified according to cell type specific methylation groups. Intermediate tumours have higher risk of
465 relapse compare to α- and β-like tumours (p-value = 0.00013). In **B** and **C** disease free survival of a
466 subset of patients of cohort 1 (n=35) with available data for both DNA methylation profile and PDX1,
467 ARX and DAXX/ATRX IHC. In **B** patient stratification according to cell type specific methylation
468 groups and in **C** according to PDX1, ARX and DAXX/ATRX IHC **D.** Consensus clustering of the 125
469 PanNETs according to the 6364 differentially methylated sites between α-like, β-like and intermediate
470 tumours (adj. p-value<0.001 and |Δβ|>0.2). Cluster stability was reached for *k*=4 (see Fig. S6 A).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Consensus cluster correlation is indicated according to the blue scale as depicted. Each column represents one PanNET sample. Tumour mutations and cell type subgroups are indicated according to the reported colours. **E.** Progression model hypothesis based on epigenetic and genetic evolution: α-like tumours originate from α-cells upon *MEN1* inactivation, the progression is enhanced by secondary events, including loss of DAXX/ATRX and chromosomal instability with recurrent LOH and activation of ALT. These events are associated to a gradual loss of differentiation. On the other hand, intermediate tumours may originate directly from a common unknown progenitor cell. Beta-like tumours originate from β-cells upon different genetic event and they are mainly insulinoma, usually indolent.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

References

1. Boninsegna, L. *et al.* Malignant pancreatic neuroendocrine tumour: Lymph node ratio and Ki67 are predictors of recurrence after curative resections. *European Journal of Cancer* **48**, 1608–1615 (2012).
2. Falconi, M. *et al.* ENETS Consensus Guidelines Update for the Management of Patients with Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine Tumors. *Neuroendocrinology* **103**, 153–171 (2016).
3. Di Domenico, A., Wiedmer, T., Marinoni, I. & Perren, A. Genetic and epigenetic drivers of neuroendocrine tumours (NET). *Endocr. Relat. Cancer* **24**, R315–R334 (2017).
4. 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.e5 (2014).
5. 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).
6. Alvarez, M. J. *et al.* A precision oncology approach to the pharmacological targeting of mechanistic dependencies in neuroendocrine tumors. *Nat Genet* **50**, 979–989 (2018).
7. Chan, C. S. *et al.* ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. *Nat Commun* **9**, 4158 (2018).
8. Cejas, P. *et al.* Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. *Nat. Med.* **25**, 1260–1265 (2019).
9. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
10. Jennings, R. E. *et al.* Development of the human pancreas from foregut to endocrine commitment. *Diabetes* **62**, 3514–3522 (2013).
11. Napolitano, T. *et al.* Pax4 acts as a key player in pancreas development and plasticity. *Seminars in Cell & Developmental Biology* **44**, 107–114 (2015).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

12. Guz, Y. *et al.* Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during ontogeny. *Development* **121**, 11–18 (1995).
13. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).
14. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**, 1095–1099 (2010).
15. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**, 311–317 (2017).
16. Azzarelli, R., Simons, B. D. & Philpott, A. The developmental origin of brain tumours: a cellular and molecular framework. *Development* **145**, (2018).
17. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
18. Bormann, F. *et al.* Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal Carcinogenesis. *Cell Reports* **23**, 3407–3418 (2018).
19. Gaiti, F. *et al.* Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia. *Nature* **569**, 576–580 (2019).
20. Jäkel, C. *et al.* Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas reveal aberrations in genome stability. *Nat Commun* **8**, 1–10 (2017).
21. Xiao, Q. *et al.* Cancer-Associated Fibroblasts in Pancreatic Cancer Are Reprogrammed by Tumor-Induced Alterations in Genomic DNA Methylation. *Cancer Res.* **76**, 5395–5404 (2016).
22. Reinius, L. E. *et al.* Differential DNA Methylation in Purified Human Blood Cells: Implications for Cell Lineage and Studies on Disease Susceptibility. *PLOS ONE* **7**, e41361 (2012).
23. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Systems* **3**, 385-394.e3 (2016).
24. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metabolism* **24**, 593–607 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

25. Ramond, C. *et al.* Reconstructing human pancreatic differentiation by mapping specific cell populations during development. *eLife* **6**, e27564 (2017).
26. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**, e31977 (2018).
27. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).
28. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
29. Houseman, E. A. & Ince, T. A. Normal Cell-Type Epigenetics and Breast Cancer Classification: A Case Study of Cell Mixture–Adjusted Analysis of DNA Methylation Data from Tumors. *Cancer Inform* **13**, 53–64 (2014).
30. Queirós, A. C. *et al.* Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the Entire B Cell Lineage. *Cancer Cell* **30**, 806–821 (2016).
31. Fernandez, A. F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**, 407–419 (2012).
32. Perren, A. *et al.* Multiple Endocrine Neoplasia Type 1 (MEN1): Loss of One MEN1 Allele in Tumors and Monohormonal Endocrine Cell Clusters But Not in Islet Hyperplasia of the Pancreas. *J Clin Endocrinol Metab* **92**, 1118–1128 (2007).
33. Hackeng, W. M. *et al.* Aberrant Menin expression is an early event in pancreatic neuroendocrine tumorigenesis. *Hum. Pathol.* **56**, 93–100 (2016).
34. Karnik, S. K. *et al.* Menin regulates pancreatic islet growth by promoting histone methylation and expression of genes encoding p27Kip1 and p18INK4c. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14659–14664 (2005).
35. Thorel, F. *et al.* Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell loss. *Nature* **464**, 1149–1154 (2010).
36. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic α to β cell reprogramming. *J Clin Invest* **123**, 1275–1284 (2013).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

559 37. Chakravarthy, H. *et al.* Converting adult pancreatic islet α-cells into β-cells by targeting both
560 Dnmt1 and Arx. *Cell Metab* **25**, 622–634 (2017).
561 38. Xiao, X. *et al.* Endogenous Reprogramming of Alpha Cells into Beta Cells Induced by Viral Gene
562 Therapy Reverses Autoimmune Diabetes. *Cell Stem Cell* **22**, 78-90.e4 (2018).
563 39. Zhang, Z. *et al.* A New Way for Beta Cell Neogenesis: Transdifferentiation from Alpha Cells
564 Induced by Glucagon-Like Peptide 1. *Journal of Diabetes Research*
565 https://www.hindawi.com/journals/jdr/2019/2583047/ (2019)
566 doi:https://doi.org/10.1155/2019/2583047.
567 40. Lu, J. *et al.* Alpha cell-specific Men1 ablation triggers the transdifferentiation of glucagon-
568 expressing cells and insulinoma development. *Gastroenterology* **138**, 1954–1965 (2010).
569 41. Shen, H.-C. J. *et al.* Multiple endocrine neoplasia type 1 deletion in pancreatic alpha-cells leads to
570 development of insulinomas in mice. *Endocrinology* **151**, 4024–4030 (2010).
571 42. Cao, Y. *et al.* Whole exome sequencing of insulinoma reveals recurrent T372R mutations in YY1.
572 *Nature Communications* **4**, 1–6 (2013).
573 43. Lichtenauer, U. D. *et al.* Frequency and clinical correlates of somatic Ying Yang 1 mutations in
574 sporadic insulinomas. *J. Clin. Endocrinol. Metab.* **100**, E776-782 (2015).
575 44. Hong, X. *et al.* Whole-genome sequencing reveals distinct genetic bases for insulinomas and non-
576 functional pancreatic neuroendocrine tumours: leading to a new classification system. *Gut*
577 (2019) doi:10.1136/gutjnl-2018-317233.
578 45. Scarpa, A. *et al.* Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature* **543**,
579 65–71 (2017).
580 46. Lawrence, B. *et al.* Recurrent loss of heterozygosity correlates with clinical outcome in pancreatic
581 neuroendocrine cancer. *NPJ Genom Med* **3**, 18 (2018).
582 47. Pea, A. *et al.* Genetic Analysis of Small Well-differentiated Pancreatic Neuroendocrine Tumors
583 Identifies Subgroups With Differing Risks of Liver Metastases. *Ann. Surg.* (2018)
584 doi:10.1097/SLA.0000000000003022.

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

48. Sturm, D. *et al.* Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological Subgroups of Glioblastoma. *Cancer Cell* **22**, 425–437 (2012).

49. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of Progression in Diffuse Glioma. *Cell* **164**, 550–563 (2016).

50. Pixberg, C. F., Schulz, W. A., Stoecklein, N. H. & Neves, R. P. L. Characterization of DNA Methylation in Circulating Tumor Cells. *Genes (Basel)* **6**, 1053–1075 (2015).

51. Pipinikas, C. P. *et al.* Epigenetic dysregulation and poorer prognosis in DAXX-deficient pancreatic neuroendocrine tumours. *Endocr. Relat. Cancer* **22**, L13-18 (2015).

52. RV, L., RY, O., G, K. & J, R. *WHO Classification of Tumours of Endocrine Organs*.

53. *AJCC Cancer Staging Manual*. (Springer International Publishing, 2017).

54. Thirlwell, C. *et al.* Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods* **52**, 248–254 (2010).

55. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430 (2014).

56. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).

57. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile Within Array Normalization for Illumina Infinium HumanMethylation450 BeadChips. *Genome Biology* **13**, R44 (2012).

58. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22 (2017).

59. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).

60. 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).

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

611 61. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in
612 microarray experiments. *Stat Appl Genet Mol Biol* **3**, Article3 (2004).
613 62. Peters, T. J. *et al.* De novo identification of differentially methylated regions in the human
614 genome. *Epigenetics Chromatin* **8**, 6 (2015).
615 63. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary
616 analyses in R. *Bioinformatics* **35**, 526–528 (2019).
617 64. Assenov, Y. *et al.* Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods*
618 **11**, 1138–1140 (2014).
619 65. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture
620 distribution. *BMC Bioinformatics* **13**, 86 (2012).
621 66. Wilkerson, M. D. & Hayes, D. N. ConsensusClusterPlus: a class discovery tool with confidence
622 assessments and item tracking. *Bioinformatics* **26**, 1572–1573 (2010).
623 67. Lawrence, M. *et al.* Software for Computing and Annotating Genomic Ranges. *PLOS*
624 *Computational Biology* **9**, e1003118 (2013).
625 68. Hovestadt, V. & Zapatka, M. conumee: Enhanced copy-number variation analysis using Illumina
626 DNA methylation arrays. (2017) doi:10.18129/B9.bioc.conumee.
627 69. Team RC. R: The R Project for Statistical Computing. https://www.r-project.org/.
628 70. Kassambara, A., Kosinski, M., Biecek, P. & Fabian, S. *survminer: Drawing Survival Curves using*
629 *'ggplot2'*. (2019).
630 71. Therneau, T. M., until 2009), T. L. (original S.->R port and R. maintainer, Elizabeth, A. & Cynthia,
631 C. *survival: Survival Analysis*. (2020).
632
633

bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Acknowledgements

We thank Cornelia Schlup and Maja Neuenschwander for technical assistance, Heidi Erika Lisa Tschanz-Lischer, for scientific and bioinformatics support, Tissue Biobank Bern (TBB) for support, Ruth Pidsley and Magali Humbert for helpful discussions.

Author Contributions

A.D.D. was involved in the study design, was responsible for generation, assembly and analysis of the data and writing of the manuscript. C.P.P. contributed data acquisition and supported in DNA methylation analysis. M.R.S. gave technical support. K.B., M.D. and A.P. revised pathological data. C.S. provided bioinformatic support. E.V. performed NGS analysis. C.T. contributed acquisition of clinical and pathological data, provided PanNET methylation data from UCL cases and supported in DNA methylation analysis. C.P.P., K.B., C.S., M.D. and C.T. contributed critical revision of the manuscript for important intellectual content. A.P. and I.M. contributed concept and design of the study, generation, assembly and analysis of the data, writing of the manuscript. C.P.P., C.T, A.P. and I.M. obtained funding.

All the authors have read and approved the final manuscript.

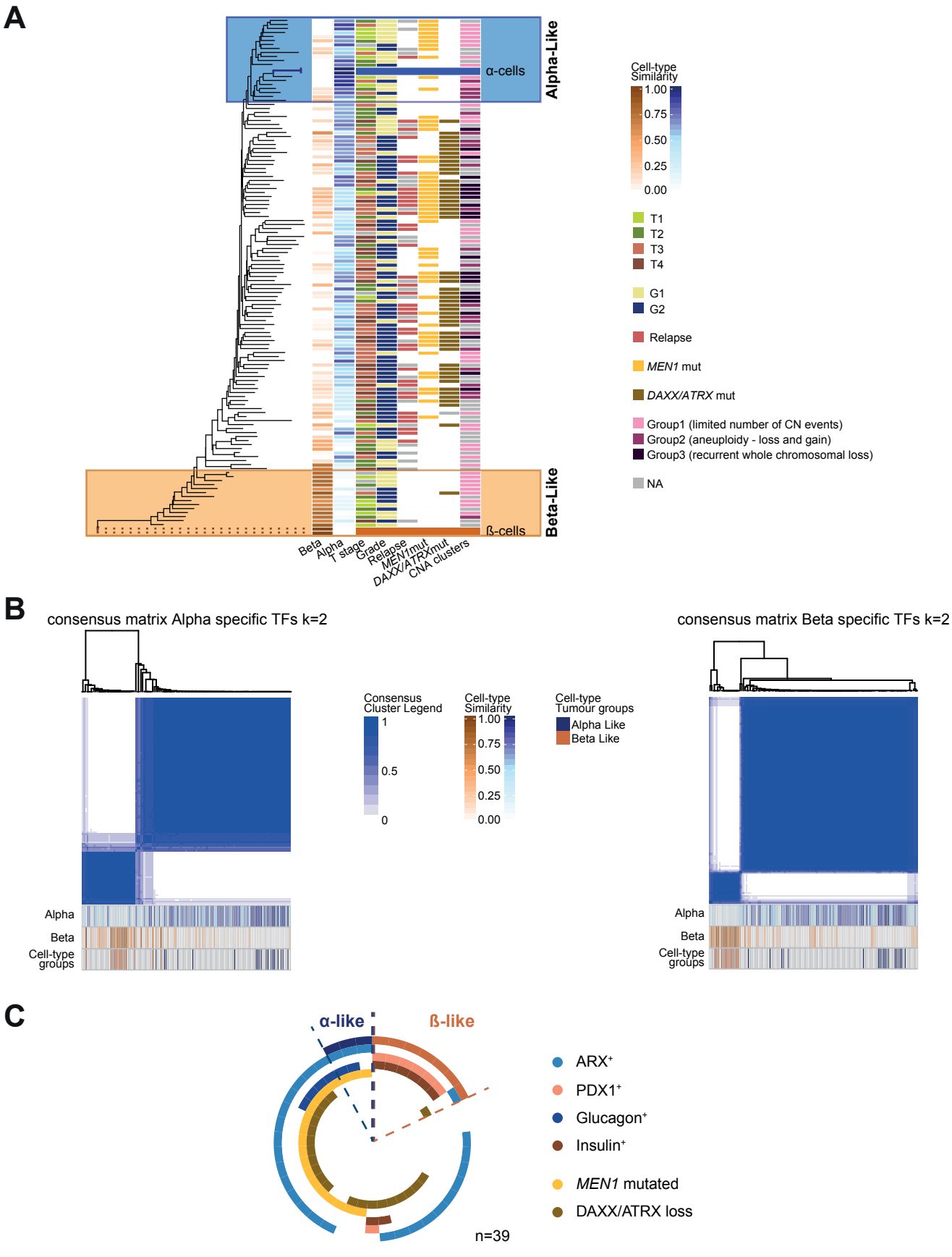
The authors declare no potential conflict of interest.

Financial Disclosures: the authors have nothing to disclose.

Writing Assistance: none

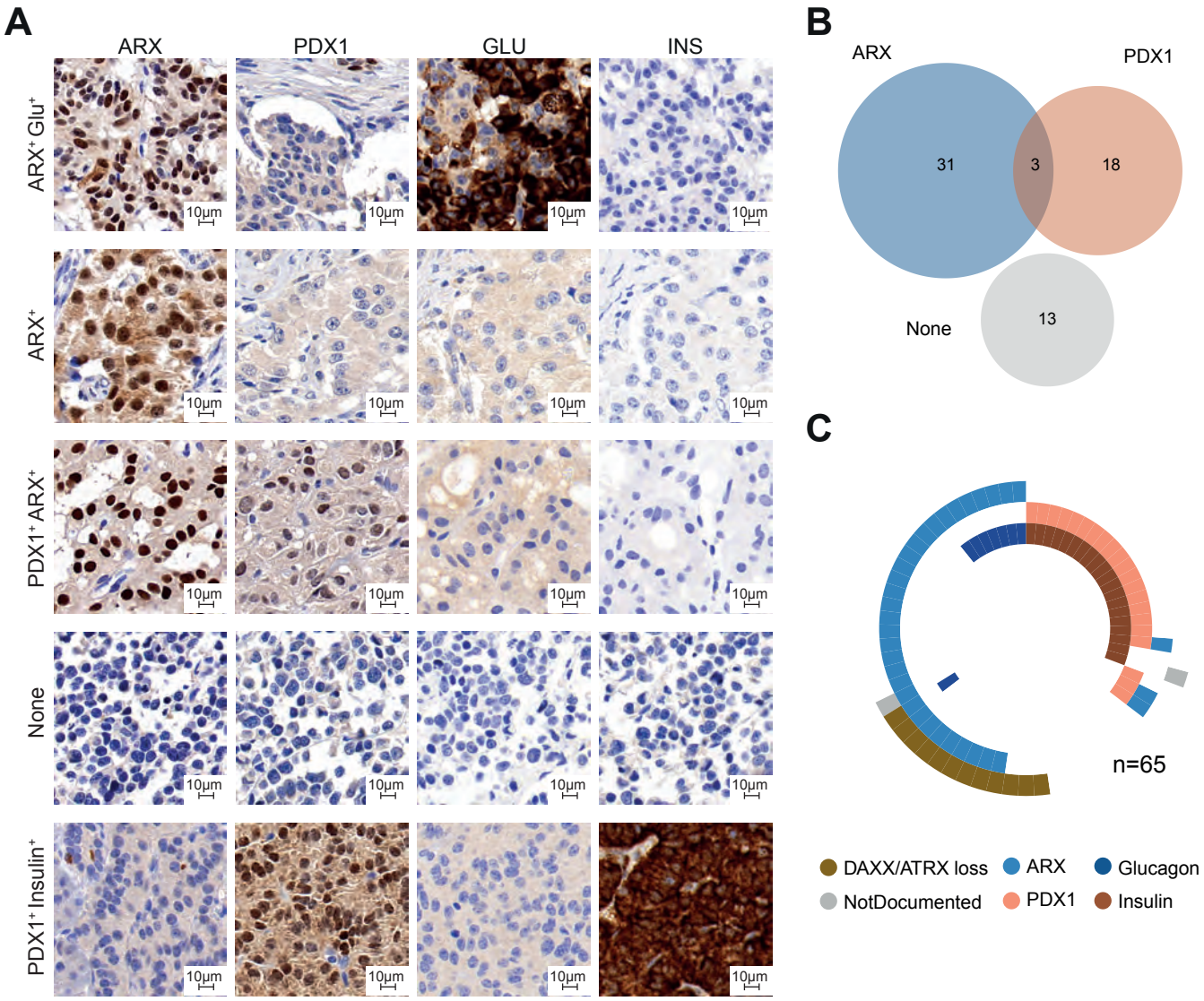
bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 1



bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Figure 2



bioRxiv preprint doi: https://doi.org/10.1101/2020.04.08.029785; this version posted April 8, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

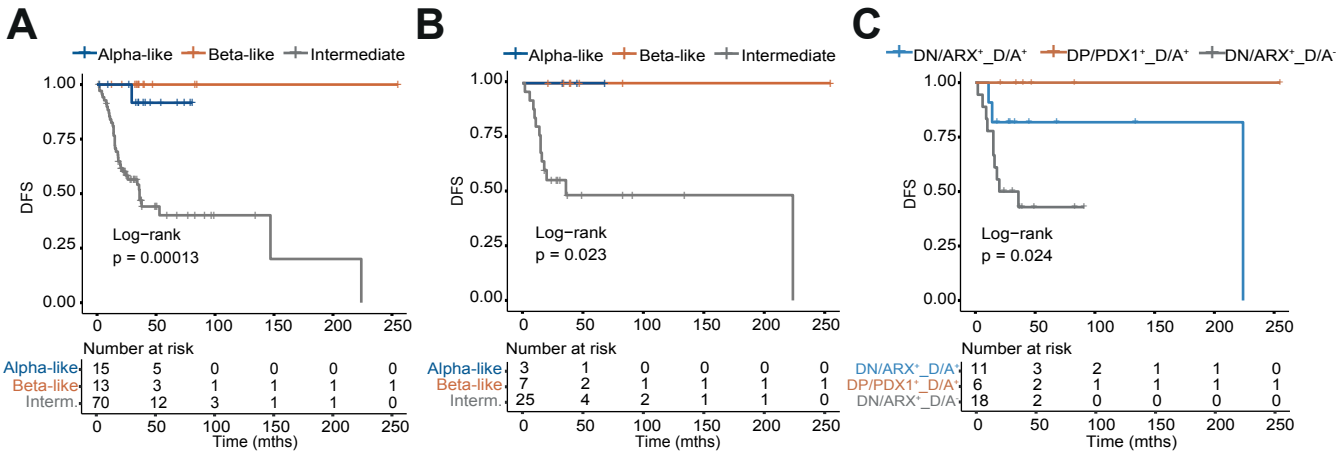
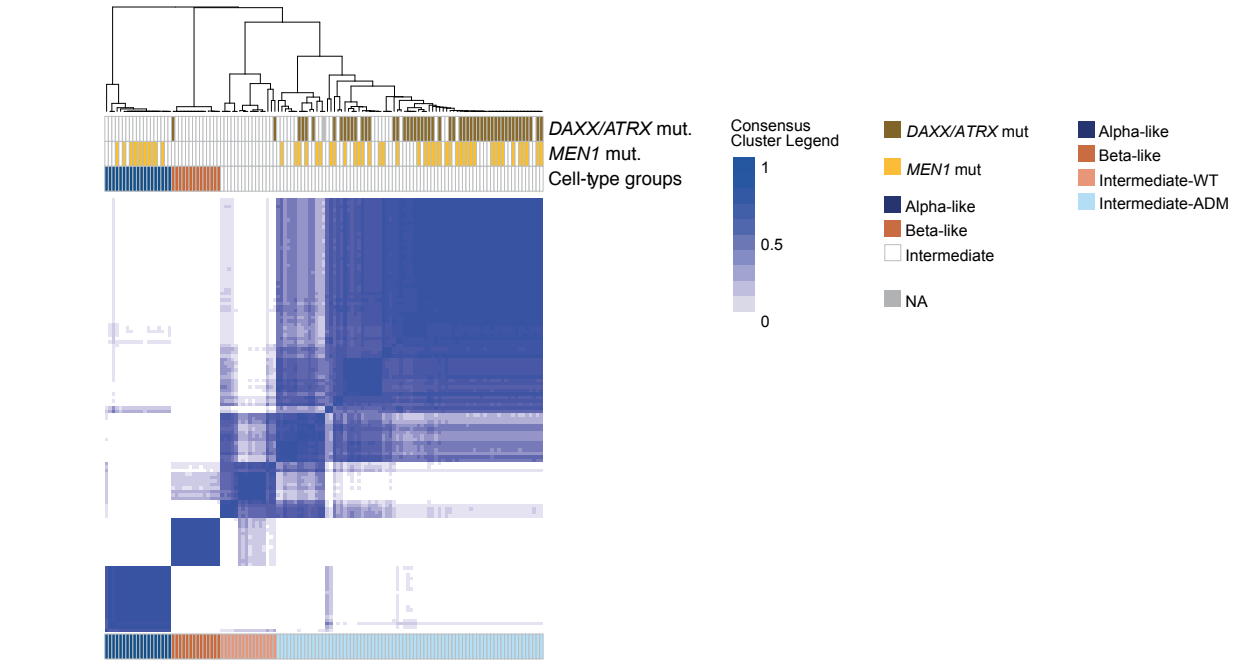
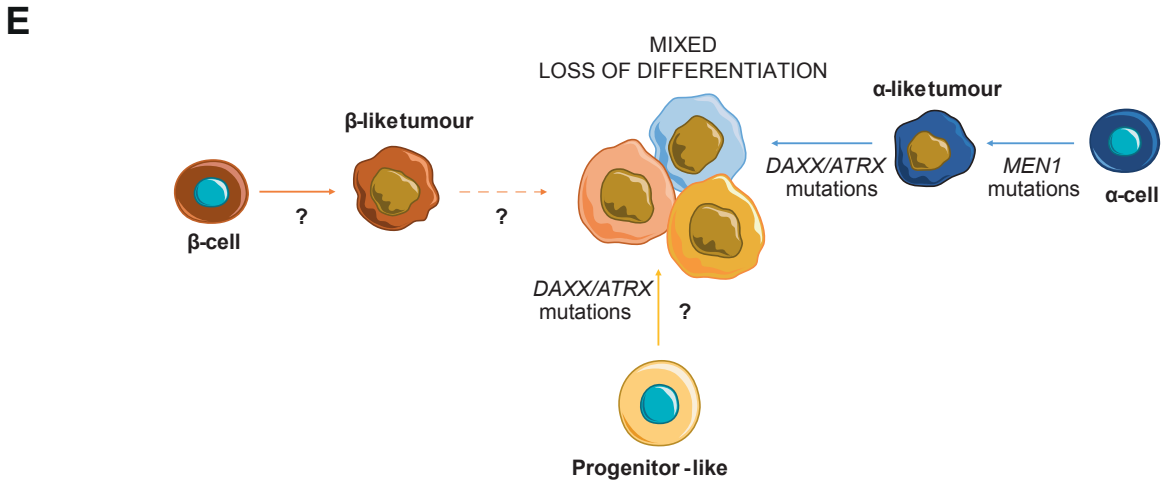



Figure 3