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

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

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

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

77 **Results**

78

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

80 We analysed the methylomes of 125 PanNETs and of isolated normal α- and β-cells[20] to determine
81 which of these is the cell of origin (cohort details are summarised in table 1 and reported in table S1). A
82 flow chart of the analysis performed is provided in figure S1. The comparison between sorted normal
83 α- (n=2) and β-cells (n=2) resulted in 2703 differentially methylated CpG sites (adj. p-value<0.001 and
84 more than 20% difference between the mean β-values of each group - |Δβ|>0.2, table S2), 2131 of them
85 were retained, after filtering and normalization, in the tumour DNA methylation matrix. Phyloepigenetic
86 analysis of the 2131 identified CpG sites using either normal β- or normal α-cells as roots, indicated
87 clear hierarchical relationships between normal cells and tumour specimens (Fig. 1A and S2A). We
88 defined PanNETs that grouped together with normal α- or β-cells as α-like or β-like PanNETs,
89 respectively (Fig. 1A and table S1). We compared these phyloepigenetic groups to cell type similarity
90 obtained from cell-type deconvolution analysis. We assigned each PanNET sample a degree of
91 methylome similarity to each of the following cell types: α, β, acinar, ductal[20], fibroblastic pancreatic
92 cells[21] and hematopoietic cells[22] (Fig. S2B). Notably, phyloepigenies that were close to normal α/β-cells
93 showed similarity >65% and an average of 72% and 73% of similarity to α and β-cells, respectively
94 (SD for α-like ±15% and for β-like ±9%, table S1). T-distributed stochastic neighbour embedding (*t*-
95 SNE) analysis, using the identified 2131 CpGs and including normal cells and PanNET samples, showed
96 a consistent segregation, according to the groups defined via similarity scores and phyloepigeny analysis
97 (Fig. S2C). These results showed at least two groups of tumours with clear and high similarity to either
98 α- or β-cells (Fig. 1A, S2A and S2C). The majority of tumours, however, revealed an intermediate
99 differentiation profile. Intermediate tumours clustered between α- and β-cells and showed weak
100 similarities to α- (as average 51 ±20 % of similarity) and very weak similarity to β-cells (as average 14
101 ±16 % of similarity), respectively.

102

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

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

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

115

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

117 Within the β-like PanNETs 13 of 14 were *MEN1/DAXX/ATRX* wild type. Alpha-like PanNETs
118 were enriched for *MEN1* only mutations (11/19), 8/19 α-like PanNETs were *MEN1/DAXX/ATRX* wild
119 type and none harboured *MEN1/DAXX/ATRX* mutations (Fig. 1A and table S1). Sixty-seven percent
120 (62/92) of intermediate PanNETs harboured mutations in *MEN1* and/or *DAXX/ATRX* genes
121 (intermediate-ADM) while the rest were wild type (intermediate-WT) (Fig. 1A and table S1).

122 Inferred copy number aberrations (CNAs) from HM450 signal intensities stratified the tumours into
123 3 subtypes (Fig. S3E). Copy number aberration (CNA) group 1 had few copy number events. Group 2
124 included samples with gains and losses. CNA group 3 showed the largest number of CNAs,
125 predominantly with loss of multiple chromosomes (Fig. S3E). Notably, α- and β-like tumours presented
126 few copy number events while the intermediate PanNETs showed increasingly copy number events (Fig.
127 1A).

128

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

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

139 Immunohistochemistry confirmed the same group segregation as was determined by the methylation
140 analysis: 1) α-like tumours expressed ARX, a subset produced glucagon (2/3, 67%), 2), β-like tumours
141 expressed both PDX1 and insulin, with only 1 sample positive for ARX and PDX1 negative (Fig. 1C),
142 3) 24/29 (83%) intermediate PanNETs were positive for ARX and rarely produced glucagon (4/22,
143 18%). A minority of intermediate tumours (4/29, 14%) were negative for both PDX1 and ARX and only
144 one intermediate PanNET expressed PDX1 and insulin (Fig. 1C, table S1).

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

147

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

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

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

163

164 **Epigenetic differentiation status determines clinical outcome of PanNETs**

165 In order to determine clinical utility, we reviewed clinical outcome and prognosis for α-like, β-like
166 and intermediate tumours. Fourteen of 19 α-like PanNETs were G1 and 14/18 (1 without data) were of
167 low stage (T1 or T2, Fig. 1A and table S1). Similarly, 8/14 β-like PanNETs were G1 and 9/11 (3 without
168 data) were T1 or T2 (Fig. 1A and table S1). Only one patient with an α-like tumour had relapse and
169 none of the patients with a β-like tumour (median follow-up time 37.2 months, table S1 and Fig. 3A).
170 Tumours that belonged to the intermediate phyloepigenies, were instead enriched for high stage (63/88
171 were T3 or T4, 4 without data) and higher grade (65/92 are G2) PanNETs had increased relapse risk
172 (37/70, 22 had not available data - Fig. 1A, 3A and table S1). Disease free survival in intermediate
173 tumours was significantly shorter than in α-like and β-like tumours ($p<0.001$, Fig. 3A). To assess if
174 epigenetic profile can stratify patients better than TF and DAXX/ATRX expression we performed
175 disease free survival analysis on all samples where DNA methylation and TF expression data were
176 available. We determined time to relapse, using either expression of ARX, PDX1 and DAXX/ATRX or
177 the epigenetic status as discriminant (Fig. 3B and C). Epigenetic profiles could predict more accurately
178 the risk of relapse as shown in figures 3B and C.

179 Further, to identify a tumour dependent signature, suitable for stratification in clinical practice, we
180 have compared the three cell-type specific clusters, α-like, β-like and intermediate. The comparison
181 revealed 6364 unique differentially methylated sites (tables S4-S6; adj. p-value<0.001 and |Δβ|>0.2).
182 Consensus clustering of the 6364 CpG sites showed the maximal stability for *k*=4 (Fig. S6A). The
183 existence of α-like and β-like clusters was confirmed (Fig. 3D and tableS1). The intermediate subgroup
184 was divided into 2 further groups, one enriched for *DAXX/ATRX/MEN1* mutations, intermediate-ADM,
185 (61/76 had mutations in at least 1 of the 3 genes; Fig. 3D and tableS1) and the other enriched for
186 *DAXX/ATRX/MEN1* wild type samples, intermediate-WT (15/16, only 1 *DAXX/ATRX* mutated;, Fig.
187 3D and tableS1). PanNETs included in the intermediate-ADM and intermediate-WT showed
188 comparable risk to relapse (Fig. S6B). To prove the stability of the signature, we repeated the analysis
189 including 32 new G1/G2 PanNETs, for which DNA methylation data were public available[7] (table S7).
190 After filtering and normalization processes, 6359/6364 of the previously identified differentially
191 methylated sites were used in the consensus clustering algorithm. The results confirmed the 4 PanNET
192 subgroups with risks to relapse similar to previous analysis (Fig. S6 C-E).

193 **Discussion**

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

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

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

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

228 (table S1). Additionally, intermediate-ADM PanNETs were on average more similar to α- (53±17%)
229 than β-cells (12±13%). Ninety-two percent of the intermediate-ADM tumours were positive for ARX
230 (23/25) and none for PDX1. This might indicate that intermediate-ADM PanNETs are more related to
231 α-cells rather than β-cells. The PanNET methylation sub-groups that we have identified reflect the
232 groups A (tumours that most likely originate form α-cells), B (tumours that most likely originate form
233 β-cells) and C (tumours with intermediate phenotype) recently described by Cejas *et al.*, via the analysis
234 of super-enhancer signatures[8]. Comparable to our study, the analysis of TF expression via IHC
235 performed by Cejas *et al.* revealed mutually exclusive expression of either ARX or PDX1 in the majority
236 of the cases. Additionally, while PDX1 expression appears to be specific for benign and low stage
237 tumours, ARX expression is retained at early and at advanced stages and only DAXX/ATRX status
238 provide more information about stage and risk of disease progression[8].

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

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

256 While the vast majority of β-like PanNETs expressed PDX1 and insulin, the two malignant (N1
257 and/or M1) insulinomas of our cohort, showed one intermediate-ADM (*DAXX/ATRX* mutated) and one
258 α-like (*DAXX/ATRX* wild type) methylation signatures (table S1). Only for the intermediate-ADM
259 insulinoma, data on PDX1 and ARX were available; it resulted positive for ARX, and negative for PDX1
260 expression. Accordingly, the only malignant insulinoma of the Chan *et al.* cohort was mutated for *ATRX*
261 and clustered with the intermediate-ADM tumours. In line with this observation, a recent study including
262 37 sporadic insulinomas (35 primary and 2 liver metastases) showed that all the five insulinomas which

263 metastasized were ARX positive and 4/5 had ALT activation (3 primary and 2 liver metastases)[33]. These
264 data suggest that malignant insulinomas may arise from α-cells or stem-cells rather than β-cells. Under
265 certain conditions and stimuli α-cells are able to trans-differentiate into β-cells[35–39]. Additionally, α-cell
266 specific *Men1* knockout in mice leads to the development of glucagonomas which evolve into mixed
267 glucagonoma/insulinoma to ultimately become insulinomas, possibly via trans-differentiation of the
268 *Men1*-deleted α-cells[40,41]. Beta-like and/or PDX1 positive tumours are strongly enriched for benign
269 insulinomas in both the first and second cohort. They generally show no mutations in any of the most
270 commonly mutated genes for PanNETs (*DAXX*, *ATRX* and *MEN1*). These data confirmed the genetic
271 difference between non-functioning tumours and insulinomas[5,42–44]. Of the thirteen wild type PanNETs
272 obtained from Chan *et al.*[7], five were included in the β-like cluster (2 insulinomas and 3 NF-PanNETs),
273 6 in the intermediate-WT cluster and only 2 in the intermediate-ADM cluster. All the ADM-mutant
274 PanNETs were included in the intermediate-ADM cluster (table S7). These data and the few β-like NF-
275 PanNETs included in the cohort 1, demonstrate that occasionally also NF-PanNETs might originate
276 from β-cells.

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

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

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

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

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.

304 **Methods**

305 **Patient cohorts**

306 A cohort of primary PanNETs was assembled from two international centres; 19 samples previously
307 analyzed[51] from UCL Cancer Institute (London, UK) and 26 from Institute of Pathology (Bern,
308 Switzerland). All cases were classified according to WHO 2017 criteria[52]. TNM staging is based on the
309 8th edition UICC/AJCC. Inclusion criteria were histopathologic diagnosis of well-differentiated G1/G2
310 PanNETs, availability of tissue material and sufficient tumour purity (>70%). Seventeen of 45 tumour
311 samples (38%) were classified as G1, 28/45 (62%) as G2 (table 1). Two samples derived from MEN1-
312 patients. Median follow-up time is 51.5 months. All analyses were performed on Formalin-Fixed
313 Paraffin-Embedded (FFPE) tissue specimens obtained from routine pathological work-up. Additional
314 clinico-pathological characteristics are reported in table S1. The study on this cohort was approved by
315 the local ethics committees (Bern: number 105/2015; London: number 09/H0722/27).

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

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

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

326

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

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

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

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

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

345

346 **DNA methylation analysis**

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

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

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

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

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).

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

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*

432 **Table 3: Validation cohort**

	Patient cohort 433
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

434
435

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).

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.

479 **References**

480 1. Boninsegna, L. *et al.* Malignant pancreatic neuroendocrine tumour: Lymph node ratio and Ki67
481 are predictors of recurrence after curative resections. *European Journal of Cancer* **48**, 1608–1615
482 (2012).
483 2. Falconi, M. *et al.* ENETS Consensus Guidelines Update for the Management of Patients with
484 Functional Pancreatic Neuroendocrine Tumors and Non-Functional Pancreatic Neuroendocrine
485 Tumors. *Neuroendocrinology* **103**, 153–171 (2016).
486 3. Di Domenico, A., Wiedmer, T., Marinoni, I. & Perren, A. Genetic and epigenetic drivers of
487 neuroendocrine tumours (NET). *Endocr. Relat. Cancer* **24**, R315–R334 (2017).
488 4. Marinoni, I. *et al.* Loss of DAXX and ATRX are associated with chromosome instability and
489 reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology* **146**, 453-
490 460.e5 (2014).
491 5. Sadanandam, A. *et al.* A Cross-Species Analysis in Pancreatic Neuroendocrine Tumors Reveals
492 Molecular Subtypes with Distinctive Clinical, Metastatic, Developmental, and Metabolic
493 Characteristics. *Cancer Discov* **5**, 1296–1313 (2015).
494 6. Alvarez, M. J. *et al.* A precision oncology approach to the pharmacological targeting of
495 mechanistic dependencies in neuroendocrine tumors. *Nat Genet* **50**, 979–989 (2018).
496 7. Chan, C. S. *et al.* ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct
497 alpha-cell signature subgroup. *Nat Commun* **9**, 4158 (2018).
498 8. Cejas, P. *et al.* Enhancer signatures stratify and predict outcomes of non-functional pancreatic
499 neuroendocrine tumors. *Nat. Med.* **25**, 1260–1265 (2019).
500 9. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947
501 (2013).
502 10. Jennings, R. E. *et al.* Development of the human pancreas from foregut to endocrine
503 commitment. *Diabetes* **62**, 3514–3522 (2013).
504 11. Napolitano, T. *et al.* Pax4 acts as a key player in pancreas development and plasticity. *Seminars in*
505 *Cell & Developmental Biology* **44**, 107–114 (2015).

506 12. Guz, Y. *et al.* Expression of murine STF-1, a putative insulin gene transcription factor, in beta cells
507 of pancreas, duodenal epithelium and pancreatic exocrine and endocrine progenitors during
508 ontogeny. *Development* **121**, 11–18 (1995).
509 13. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330
510 (2015).
511 14. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468**,
512 1095–1099 (2010).
513 15. Northcott, P. A. *et al.* The whole-genome landscape of medulloblastoma subtypes. *Nature* **547**,
514 311–317 (2017).
515 16. Azzarelli, R., Simons, B. D. & Philpott, A. The developmental origin of brain tumours: a cellular
516 and molecular framework. *Development* **145**, (2018).
517 17. Visvader, J. E. Cells of origin in cancer. *Nature* **469**, 314–322 (2011).
518 18. Bormann, F. *et al.* Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal
519 Carcinogenesis. *Cell Reports* **23**, 3407–3418 (2018).
520 19. Gaiti, F. *et al.* Epigenetic evolution and lineage histories of chronic lymphocytic leukaemia.
521 *Nature* **569**, 576–580 (2019).
522 20. Jäkel, C. *et al.* Genome-wide genetic and epigenetic analyses of pancreatic acinar cell carcinomas
523 reveal aberrations in genome stability. *Nat Commun* **8**, 1–10 (2017).
524 21. Xiao, Q. *et al.* Cancer-Associated Fibroblasts in Pancreatic Cancer Are Reprogrammed by Tumor-
525 Induced Alterations in Genomic DNA Methylation. *Cancer Res.* **76**, 5395–5404 (2016).
526 22. Reinius, L. E. *et al.* Differential DNA Methylation in Purified Human Blood Cells: Implications for
527 Cell Lineage and Studies on Disease Susceptibility. *PLOS ONE* **7**, e41361 (2012).
528 23. Muraro, M. J. *et al.* A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Systems* **3**,
529 385-394.e3 (2016).
530 24. Segerstolpe, Å. *et al.* Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and
531 Type 2 Diabetes. *Cell Metabolism* **24**, 593–607 (2016).

532 25. Ramond, C. *et al.* Reconstructing human pancreatic differentiation by mapping specific cell
533 populations during development. *eLife* **6**, e27564 (2017).
534 26. Thurner, M. *et al.* Integration of human pancreatic islet genomic data refines regulatory
535 mechanisms at Type 2 Diabetes susceptibility loci. *eLife* **7**, e31977 (2018).
536 27. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA
537 methylation sequencing. *Nature* **510**, 537–541 (2014).
538 28. Capper, D. *et al.* DNA methylation-based classification of central nervous system tumours.
539 *Nature* **555**, 469–474 (2018).
540 29. Houseman, E. A. & Ince, T. A. Normal Cell-Type Epigenetics and Breast Cancer Classification: A
541 Case Study of Cell Mixture–Adjusted Analysis of DNA Methylation Data from Tumors. *Cancer*
542 *Inform* **13**, 53–64 (2014).
543 30. Queirós, A. C. *et al.* Decoding the DNA Methylome of Mantle Cell Lymphoma in the Light of the
544 Entire B Cell Lineage. *Cancer Cell* **30**, 806–821 (2016).
545 31. Fernandez, A. F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**,
546 407–419 (2012).
547 32. Perren, A. *et al.* Multiple Endocrine Neoplasia Type 1 (MEN1): Loss of One MEN1 Allele in Tumors
548 and Monohormonal Endocrine Cell Clusters But Not in Islet Hyperplasia of the Pancreas. *J Clin*
549 *Endocrinol Metab* **92**, 1118–1128 (2007).
550 33. Hackeng, W. M. *et al.* Aberrant Menin expression is an early event in pancreatic neuroendocrine
551 tumorigenesis. *Hum. Pathol.* **56**, 93–100 (2016).
552 34. Karnik, S. K. *et al.* Menin regulates pancreatic islet growth by promoting histone methylation and
553 expression of genes encoding p27Kip1 and p18INK4c. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 14659–
554 14664 (2005).
555 35. Thorel, F. *et al.* Conversion of adult pancreatic alpha-cells to beta-cells after extreme beta-cell
556 loss. *Nature* **464**, 1149–1154 (2010).
557 36. Bramswig, N. C. *et al.* Epigenomic plasticity enables human pancreatic α to β cell reprogramming.
558 *J Clin Invest* **123**, 1275–1284 (2013).

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.

585 48. Sturm, D. *et al.* Hotspot Mutations in H3F3A and IDH1 Define Distinct Epigenetic and Biological
586 Subgroups of Glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
587 49. Ceccarelli, M. *et al.* Molecular Profiling Reveals Biologically Discrete Subsets and Pathways of
588 Progression in Diffuse Glioma. *Cell* **164**, 550–563 (2016).
589 50. Pixberg, C. F., Schulz, W. A., Stoecklein, N. H. & Neves, R. P. L. Characterization of DNA
590 Methylation in Circulating Tumor Cells. *Genes (Basel)* **6**, 1053–1075 (2015).
591 51. Pipinikas, C. P. *et al.* Epigenetic dysregulation and poorer prognosis in DAXX-deficient pancreatic
592 neuroendocrine tumours. *Endocr. Relat. Cancer* **22**, L13-18 (2015).
593 52. RV, L., RY, O., G, K. & J, R. *WHO Classification of Tumours of Endocrine Organs*.
594 53. *AJCC Cancer Staging Manual*. (Springer International Publishing, 2017).
595 54. Thirlwell, C. *et al.* Genome-wide DNA methylation analysis of archival formalin-fixed paraffin-
596 embedded tissue using the Illumina Infinium HumanMethylation27 BeadChip. *Methods* **52**, 248–
597 254 (2010).
598 55. Morris, T. J. *et al.* ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428–430
599 (2014).
600 56. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of
601 Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
602 57. Maksimovic, J., Gordon, L. & Oshlack, A. SWAN: Subset-quantile Within Array Normalization for
603 Illumina Infinium HumanMethylation450 BeadChips. *Genome Biology* **13**, R44 (2012).
604 58. Zhou, W., Laird, P. W. & Shen, H. Comprehensive characterization, annotation and innovative use
605 of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* **45**, e22 (2017).
606 59. Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using
607 empirical Bayes methods. *Biostatistics* **8**, 118–127 (2007).
608 60. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing
609 batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**,
610 882–883 (2012).

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

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

Figure 1

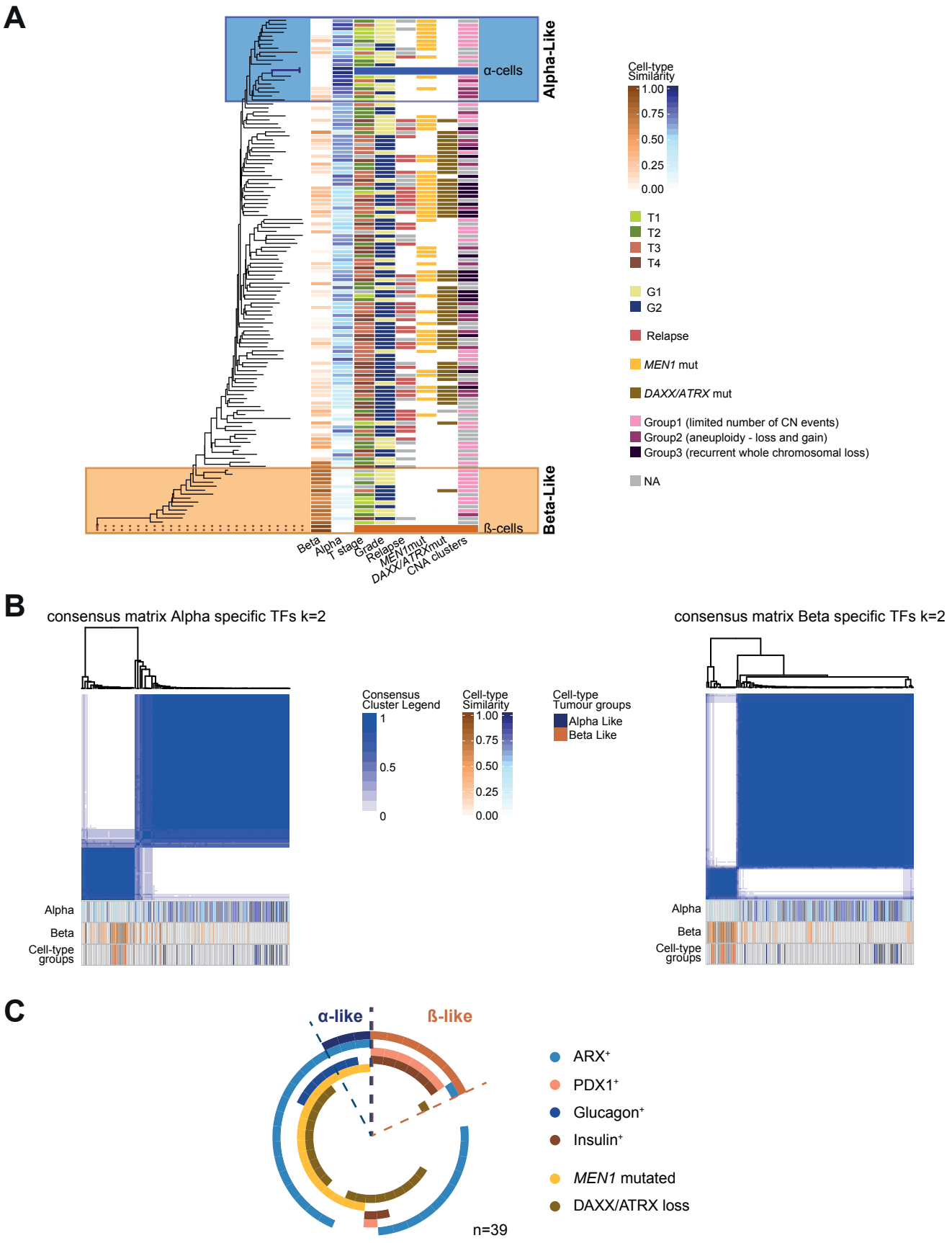

A
α-cells
Alpha-Like
ß-cells
Beta-Like
Beta
Alpha
T stage
Grade
Relapse
MEN1mut
DAXX/ATRXmut
CNA clusters
Cell-type Similarity
1.00
0.75
0.50
0.25
0.00
T1
T2
T3
T4
G1
G2
Relapse
MEN1 mut
DAXX/ATRX mut
Group1 (limited number of CN events)
Group2 (aneuploidy - loss and gain)
Group3 (recurrent whole chromosomal loss)
NA
B
consensus matrix Alpha specific TFs k=2
consensus matrix Beta specific TFs k=2
Consensus Cluster Legend
1
0.5
0
Cell-type Similarity
1.00
0.75
0.50
0.25
0.00
Cell-type Tumour groups
Alpha Like
Beta Like
Alpha
Beta
Cell-type groups
C
α-like
ß-like
ARX+
PDX1+
Glucagon+
Insulin+
MEN1 mutated
DAXX/ATRX loss
n=39

Figure 2

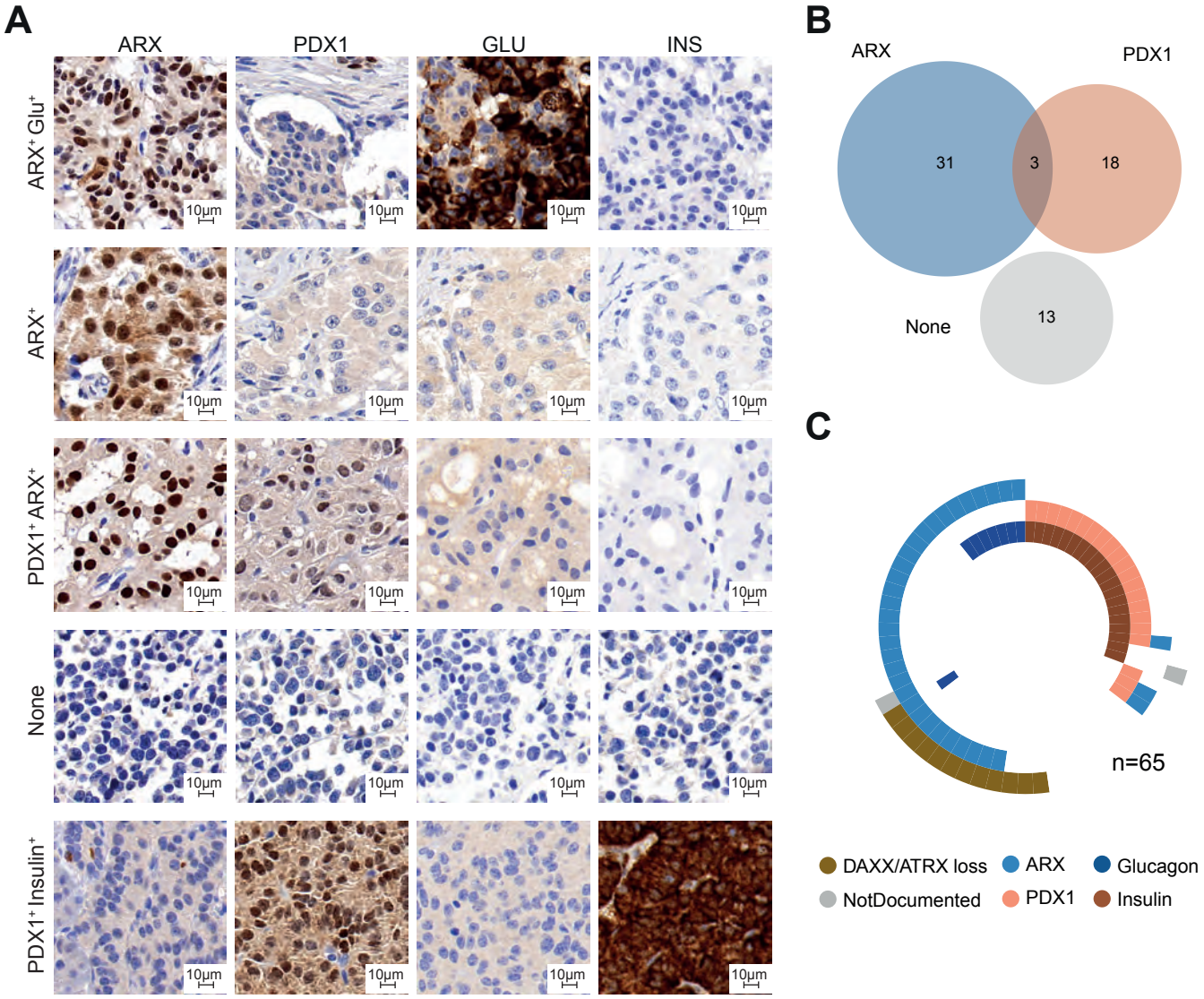
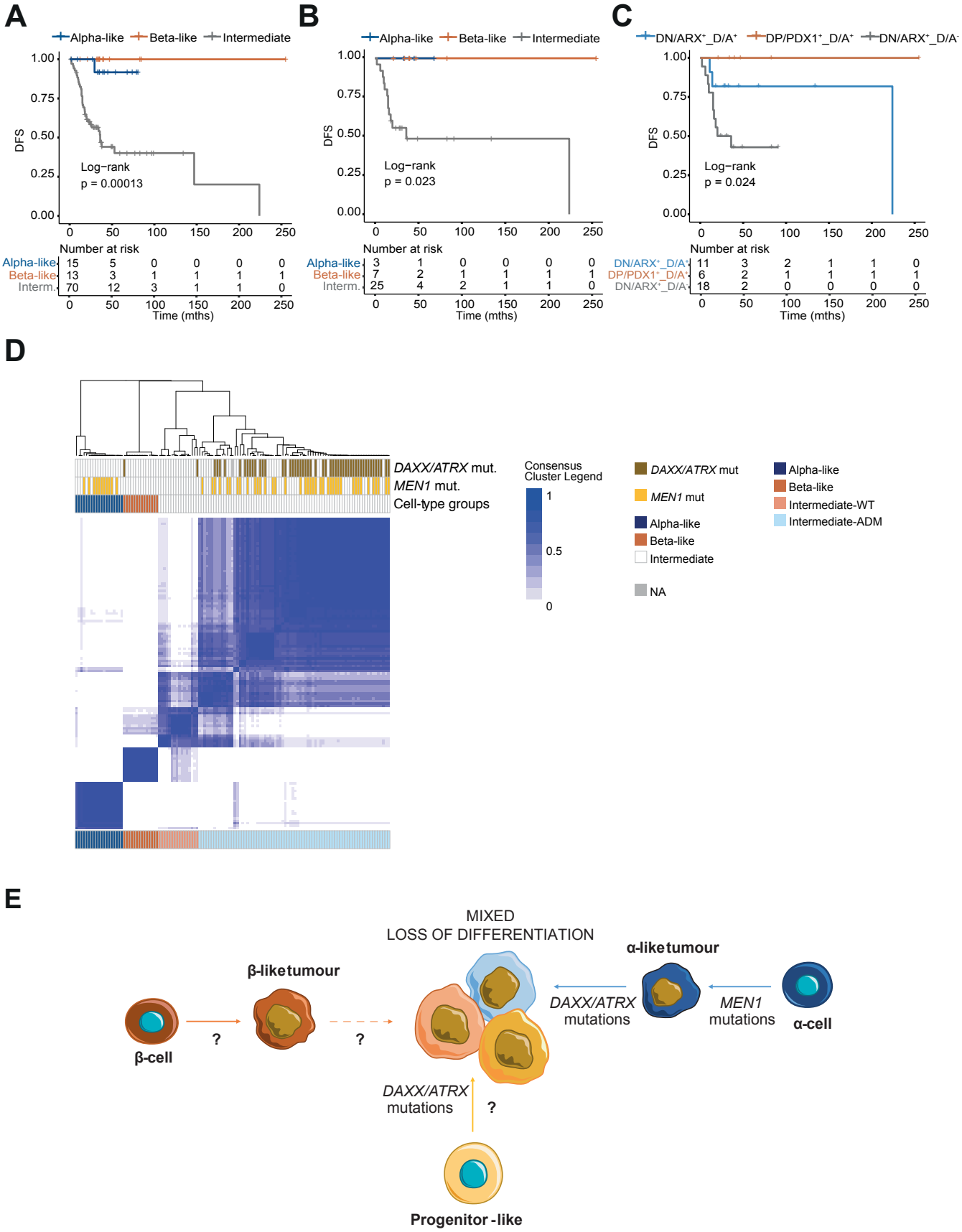

A
ARX
PDX1
GLU
INS
ARX+ Glu+
ARX+
PDX1+ ARX+
None
PDX1+ Insulin+
10µm
B
ARX
PDX1
31
3
18
None
13
C
n=65
DAXX/ATRX loss
ARX
Glucagon
NotDocumented
PDX1
Insulin

Figure 3


A
Alpha-like
Beta-like
Intermediate
DFS
Log−rank
p = 0.00013
Number at risk
Alpha-like 15 5 0 0 0 0
Beta-like 13 3 1 1 1 1
Interm. 70 12 3 1 1 0
Time (mths)
B
Alpha-like
Beta-like
Intermediate
DFS
Log−rank
p = 0.023
Number at risk
Alpha-like 3 1 0 0 0 0
Beta-like 7 2 1 1 1 1
Interm. 25 4 2 1 1 0
Time (mths)
C
DN/ARX+_D/A+
DP/PDX1+_D/A+
DN/ARX+_D/A-
DFS
Log−rank
p = 0.024
Number at risk
DN/ARX+_D/A+ 11 3 2 1 1 0
DP/PDX1+_D/A+ 6 2 1 1 1 1
DN/ARX+_D/A- 18 2 0 0 0 0
Time (mths)
D
DAXX/ATRX mut.
MEN1 mut.
Cell-type groups
Consensus Cluster Legend
1
0.5
0
DAXX/ATRX mut
MEN1 mut
Alpha-like
Beta-like
Intermediate
NA
Alpha-like
Beta-like
Intermediate-WT
Intermediate-ADM
E
MIXED
LOSS OF DIFFERENTIATION
β-liketumour
α-liketumour
β-cell
?
?
DAXX/ATRX
mutations
MEN1
mutations
α-cell
DAXX/ATRX
mutations
?
Progenitor-like