Regulatory heterogeneity of neuroendocrine tumors

Ester Davis (1), Israel Antman (1), Shahd Abu Kamel (1), Merav Hecht (1), Shani Avniel-Polak (2), Simona Grozinsky-Glasberg (2), Yotam Drier (1)

(1) The Lautenberg Center for Immunology and Cancer Research, IMRIC, Faculty of Medicine, Hebrew University of Jerusalem, Israel.

(2) Neuroendocrine Tumor Unit, ENETS Center of Excellence, Department of Endocrinology and Metabolism, Hadassah-Hebrew University Medical Center, Jerusalem, Israel.

Treating patients with lung neuroendocrine tumors is challenging. The two main challenges are the lack of effective drug treatments and the lack of reliable biomarkers to guide management, since patients with the same tumor grade/stage often have different clinical courses. The genetic, epigenetic and developmental programs that drive lung neuroendocrine tumors remain obscure, limiting our abilities to suggest new biomarkers and drug targets. We are working to systematically characterize the regulatory programs of lung neuroendocrine tumors to address this gap. We will share our current progress on the project and plans for the future.

We are characterizing putative enhancers (by H3K27ac ChIP-seq) and transcriptomes (by RNA-seq) of neuroendocrine tumors, to identify the regulatory networks the underlies these tumors. Using these methods, we have recently successfully identified regulatory and developmental subtypes of pancreatic neuroendocrine tumors [1], and matching biomarkers that demonstrated clear clinical prognostic value. We are currently characterizing and analyzing lung neuroendocrine tumors, and comparing them amongst themselves as well as to existing profiles of other neuroendocrine tumors.

We have characterized putative enhancers and transcriptomes of 12 primary typical lung neuroendocrine tumors. Clustering of the enhancer profiles revealed two or three novel subtypes. Analysis of the regulatory and transcriptional programs of the different subtypes revealed they rely on different signaling pathway, suggesting they may respond to different tyrosine kinase inhibitors.

We are now in the process of validating the newly identified subtypes across larger cohorts, and setting up experimental models to test the response to different tyrosine kinase inhibitors. In addition, we are aiming to characterize atypical lung neuroendocrine tumors, and compare them to the typical tumors we profiled.

References

[1] Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. Cejas P*, Drier Y*#, Dreijerink KMA, Brosens LAA, Deshpande V, Epstein CB, Conemans EB, Morsink FHM, Graham MK, Valk GD, Vriens MR, Castillo CF, Ferrone CR, Adar T, Bowden M, Whitton HJ, Da Silva A, Font-Tello A, Long HW, Gaskell E, Shoresh N, Heaphy CM, Sicinska E, Kulke MH, Chung DC, Bernstein BE#, Shivdasani RA#. Nature Medicine, 25, 1260–1265 (2019).

LETTERS
https://doi.org/10.1038/s41591-019-0493-4

nature medicine

Corrected: Publisher Correction

Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors

Paloma Cejas[1,2,3,16], Yotam Drier[4,5,13,16*], Koen M. A. Dreijerink[6,7], Lodewijk A. A. Brosens[8], Vikram Deshpande[5], Charles B. Epstein[4], Elfi B. Conemans[6,7], Folkert H. M. Morsink[8], Mindy K. Graham[9], Gerlof D. Valk[6], Menno R. Vriens[10], Carlos Fernandez-del Castillo[11], Cristina R. Ferrone[11], Tomer Adar[12], Michaela Bowden[1], Holly J. Whitton[4], Annacarolina Da Silva[14], Alba Font-Tello[2], Henry W. Long[2], Elizabeth Gaskell[4], Noam Shoresh[4], Christopher M. Heaphy[9], Ewa Sicinska[14], Matthew H. Kulke[1,15], Daniel C. Chung[12], Bradley E. Bernstein[4,5*] and Ramesh A. Shivdasani[1,2,15*]

Most pancreatic neuroendocrine tumors (PNETs) do not produce excess hormones and are therefore considered 'non-functional'[1–3]. As clinical behaviors vary widely and distant metastases are eventually lethal[2,4], biological classifications might guide treatment. Using enhancer maps to infer gene regulatory programs, we find that non-functional PNETs fall into two major subtypes, with epigenomes and transcriptomes that partially resemble islet α- and β-cells. Transcription factors ARX and PDX1 specify these normal cells, respectively[5,6], and 84% of 142 non-functional PNETs expressed one or the other factor, occasionally both. Among 103 cases, distant relapses occurred almost exclusively in patients with ARX^{+}PDX1^{-} tumors and, within this subtype, in cases with alternative lengthening of telomeres. These markedly different outcomes belied similar clinical presentations and histology and, in one cohort, occurred irrespective of *MEN1* mutation. This robust molecular stratification provides insight into cell lineage correlates of non-functional PNETs, accurately predicts disease course and can inform postoperative clinical decisions.

Surgery is recommended for solitary PNETs larger than 2 cm and the World Health Organization (WHO) grade is the best current tool to predict metastasis[7,8]. Insulinomas resemble normal pancreatic β-cells and carry a good prognosis[1,2], but the lineage of most non-functional PNETs is obscure and about half the cases progress to lethal metastasis months to years after surgery[1,2,9]. Although 70% of PNETs carry *MEN1*, *ATRX* or *DAXX* gene mutations, and 15% activate mammalian target of rapamycin signaling[10,11], no mutation or biological feature is sufficiently correlated with clinical outcomes to guide prognosis or therapy[3,8]. Alternative lengthening of telomeres (ALT) is associated with ATRX or DAXX loss and elevated risk of recurrence[12–14], but is not a routine clinical test.

Cellular identities are encoded in chromatin states defined by the complement of active *cis*-elements, where nucleosomes bear H3K4me1/2, H3K27ac and other covalent marks[15,16]. 'Super' or 'stretch' enhancers, which control lineage-specifying genes[17,18], especially delineate cell-specific chromatin signatures[19] and help nominate tumor cell origins[20]. Chromatin immunoprecipitation sequencing (ChIP-seq)-derived profiles of H3K27ac-marked candidate enhancers in 8 frozen, primary non-functional, pancreatic and 22 intestinal neuroendocrine (NE; carcinoid) tumors (see Supplementary Table 1) differed substantially from those of gastrointestinal carcinomas and less so from each other (Fig. 1a). Loci expressed in the two NE tumor types, such as *SYP*, were similarly marked in both, whereas organ-restricted loci, such as *CDX2*, were marked selectively (Fig. 1b). Super-enhancer profiles, distinct from those of normal islets (Fig. 1c), revealed three PNET subtypes. Considering A and B types as separate groups, 288 enhancers showed more than twofold higher H3K27ac in type A (false discovery rate (FDR) < 0.05—see Supplementary Table 2), with sites in α-cell-specific loci *ARX* and *IRX2* showing especially large differences (Fig. 1d). Conversely, 104 regions showed notably more H3K27ac in B-type tumors, including enhancers over genes such as *PDX1* and *SLC17A6*, which are not expressed in α-cells[21,22].

Across the eight PNETs, A- and B-type tumors gave strong H3K27ac signals at *ARX* and *IRX2* or at *PDX1*, respectively, whereas

[1]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA. [2]Center for Functional Cancer Epigenetics, Dana-Farber Cancer Institute, Boston, MA, USA. [3]Translational Oncology Laboratory, Hospital La Paz Institute for Health Research, Madrid, Spain. [4]Broad Institute of Harvard and MIT, Cambridge, MA, USA. [5]Department of Pathology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [6]Department of Endocrine Oncology, UMC Utrecht Cancer Center, Utrecht, the Netherlands. [7]Department of Internal Medicine, Amsterdam UMC, Amsterdam, the Netherlands. [8]Department of Pathology, UMC Utrecht Cancer Center, Utrecht, the Netherlands. [9]Department of Pathology, Johns Hopkins University School of Medicine, Baltimore, MD, USA. [10]Department of Surgical Oncology, UMC Utrecht Cancer Center, Utrecht, the Netherlands. [11]Department of Surgery, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [12]Department of Gastroenterology Division, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [13]Present address: Lautenberg Center for Immunology and Cancer Research, Hebrew University, Faculty of Medicine, Jerusalem, Israel. [14]Present address: Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA. [15]Present address: Departments of Medicine, Brigham & Women's Hospital and Harvard Medical School, Boston, MA, USA. [16]These authors contributed equally: Paloma Cejas, Yotam Drier. *e-mail: yotam.drier@mail.huji.ac.il; bernstein.bradley@mgh.harvard.edu; ramesh_shivdasani@dfci.harvard.edu

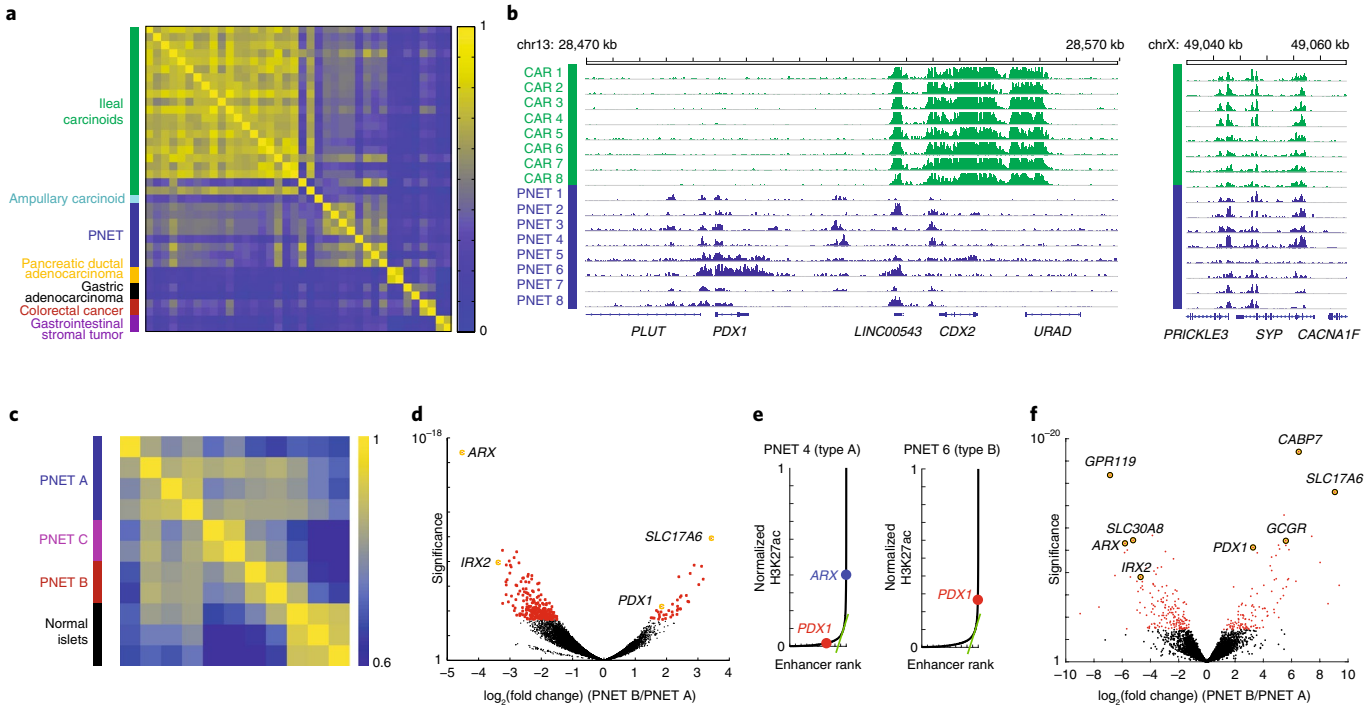


Fig. 1 | Distinctive PNET enhancer profiles. a, Pairwise Pearson's correlations of H3K27ac signals at distal regulatory elements in diverse gastrointestinal cancers. PNETs and carcinoid (small intestinal NE) tumors differ from others, reflecting distinct cell origins. Whereas intestinal NE enhancer landscapes are highly concordant, PNETs appear heterogeneous. **b**, Representative H3K27ac ChIP-seq tracks (16 of 30 samples), showing differential acetylation at the *CDX2* and *PDX1* loci; intronic *SYP* enhancers are marked in both intestinal (green) and pancreatic (blue) NETs. ChIP-seq read counts were scaled by DESeq2 normalization based on promoter signals (see Methods). **c**, Pearson's correlations of super-enhancer H3K27ac signals in PNETs and normal pancreatic islets suggest three possible disease subtypes. **d**, Significance ($\log_{10}$ scale, Wald's test) and $\log_2$ fold differences in H3K27ac ChIP-seq signals in A- and B-type PNETs, calculated by DESeq2 comparison of six biologically independent tumors. Each dot represents an individual site (red: FDR < 0.05). Selected H3K27ac sites are colored yellow and annotated by the nearest gene. **e**, Normalized H3K27ac signals across all sites (super-enhancers lie to the right of the inflection) in single representative type A and type B tumors from six samples. **f**, Significance ($\log_{10}$ scale, Wald's test) and $\log_2$ fold differences of normalized mRNA read counts in A- and B-type PNETs ($n = 8$ biologically independent tumors) considered as two groups. Each dot represents an individual gene (red: FDR < 0.05). Left and right: transcripts enriched in types A and B, respectively. Transcripts associated with selected super-enhancers, including those highlighted in **d**, are indicated (yellow).

C-type tumors were variably marked at these loci and expressed low levels of *ARX* and *PDX1* messenger RNAs (mRNAs) (see Extended Data Fig. 1a). In individual A- and B-type tumors, *ARX* and *PDX1* met objective super-enhancer criteria[19], respectively (Fig. 1e). At these and other loci, ChIP-seq for H3K4me2 (ref. [23]) on the original eight and four additional PNETs (see Supplementary Table 1) revealed similar patterns to H3K27ac (see Extended Data Fig. 1a—H3K4me2 marks the *PDX1* promoter in all PNETs, but only B-type tumors carry locus-wide H3K4me2). Aggregate (Fig. 1f) and individual (see Extended Data Fig. 1b) comparisons of A- and B-type PNET transcriptomes (see Supplementary Table 3) showed differential *ARX* and *PDX1* expression.

Knowledge of pancreatic endocrine ontogeny derives largely from studies of mouse development[24]. NEUROG3 initiates the endocrine lineage and ARX specifies α-cell fate[5], whereas PAX4 drives β-cell differentiation[25]; absence of both factors favors the δ lineage[26]. ARX is necessary for α-cell differentiation[27], but other transcription factors (TFs, for example NKX2.2, PAX6, PDX1) sustain β-cells[6,28–31]. In embryos, early α-cell specification leads to *Insulin* and *Glucagon* co-expressing bipotential precursors, followed by β-cell differentiation[32]; forced transdifferentiation between α- and β-cells also occurs through intermediate *Insulin*$^+$ *Glucagon*$^+$ cells[29]. In this light, we compared A- and B-type PNET-restricted enhancers (Fig. 1d) with areas of chromatin selectively open in normal α- or β-cells[33]. Type A regions were highly enriched for α-cell-specific sites (62 shared, $P < 1.1 \times 10^{-18}$), whereas type B-specific enhancers were enriched for β-cell-restricted sites (15 shared, $P < 1.7 \times 10^{-5}$; see Supplementary Table 4). Comparison with RNA profiles of fractionated normal human islets[34] also revealed enrichment of α- and β-cell-specific transcripts in A- ($P < 1.6 \times 10^{-9}$) and B-type ($P < 6.2 \times 10^{-16}$) PNETs, respectively. Differential TF expression distinguished PNETs better than *Glucagon* and *Insulin* mRNA levels, which were generally low and similar in the two subtypes (see Extended Data Fig. 1b).

From H3K27ac ChIP-seq on 13 additional frozen PNETs (validation cohort), tumors were designated as: type A if signals were high (>500 reads per kb per million sequence tags (RPKM)) at *ARX* and *IRX2* but absent or low (<250 RPKM) at *PDX1*; type B if the *ARX* locus was largely unmarked (<250 RPKM); and type C if read counts at *ARX* and *PDX1* were comparable. Although A or C typing was ambiguous in two PNETs, the B type was easily identified by lack of H3K27ac at *ARX*. Similar to the discovery set, the validation cohort readily distinguished type B from other PNETs by H3K27ac signals at α-cell/type A- and β-cell/type B-specific enhancers (Fig. 2 and see Extended Data Fig. 1c). Moreover, ARX and PDX1 immunostains gave the expected signals in normal human islets (see Extended Data Fig. 2a,b) and, importantly, 15 independent, additional, non-functional PNETs showed mutually exclusive, nucleus-dominant, ARX or PDX1 expression in 10 tumors (67%); 5 cases lacked or co-expressed both TFs (Fig. 3a).

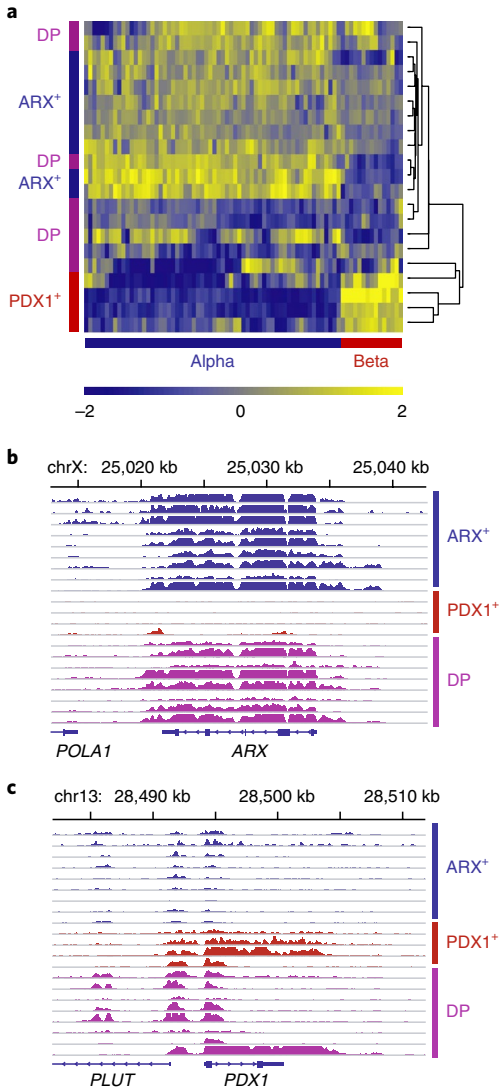


Fig. 2 | PNET subtypes represent distinct endocrine lineages. a, Relative H3K27ac ChIP-seq signals in all 21 PNETs (8 from the discovery set + 13 additional) of types A (ARX+), B (PDX1+) and C (DP) at enhancers that are specific to A versus B PNETs (Fig. 1d) and α- versus β-cells[33]. **b**,**c**, Genomic views of the *ARX* (**b**) and *PDX1* (**c**) loci, showing H3K27ac signals for all 21 PNETs. chr, chromosome. ChIP-seq signals are scaled by promoter-based DESeq2 normalization (see Methods).

Next tissue microarrays were examined that represented 77 Dutch PNETs[35], 61 of which (79%) had germline *MEN1* mutations, including 13 insulinomas associated with hyperinsulinemia (see Supplementary Table 5). Overall, 34 of the 77 cases showed nuclear expression of ARX only (31% type A), 31 expressed only PDX1 (37% type B) and 12 cases were either double-positive (DP) or double-negative (DN—8% each; Fig. 3b and see Extended Data Fig. 2b). When present, the immunohistochemistry (IHC) signals for PDX1 were uniformly strong and every insulinoma, scored blindly, expressed only PDX1. Although equally unambiguous, ARX signals varied in strength across type A tumors (see Extended Data Fig. 2c), so some DN tumors may represent technical failure of IHC, whereas others may represent type A PNETs with low ARX expression. DP tumor cells largely co-express ARX and PDX1 (Fig. 3c and see Extended Data Fig. 3a), possibly representing a dual lineage. *MEN1* mutations did not skew the proportions of A and B tumors (Fig. 3b).

H3K4me2 and mRNAs were profiled in selected tumors from the Dutch cohort using fixed-tissue chromatin immunoprecipitation sequencing (FiT-seq)[36] and RNA sequencing (RNA-seq). PNETs, identified by IHC as ARX+ or PDX1+, expressed exclusively those mRNAs and showed abundant H3K4me2 only at the corresponding super-enhancers; DP tumors (type C) expressed both *ARX* and *PDX1* mRNAs, with less pronounced enhancer marks at both loci (Fig. 3d). Of note, no PNETs in this group expressed somatostatin (SST) (see Extended Data Fig. 3b). FiT-seq for H3K27ac and immunostaining on representative formalin-fixed, paraffin-embedded (FFPE) samples from the discovery cohort also revealed concordance (see Extended Data Fig. 3c). Thus, IHC classifies PNETs robustly, reflecting lineage-specific gene and enhancer signatures that are henceforth called ARX+, PDX1+ and DP.

PNETs uniformly lacked H3K27ac and mRNA expression at early acting *NEUROG3* and *PAX4* loci (see Extended Data Fig. 4a), similar to isolated human islets, and IHC failed to detect NEUROG3 in any tumor (see Extended Data Fig. 4b). In contrast, enhancer marking and mRNA levels were high at other canonical endocrine loci (*ISL1* and *NEUROD1*—data not shown), but not at terminal differentiation genes (for example, *MAFA* and *FFAR1*; see Extended Data Fig. 4c). H3K4me2, but not H3K27ac, appeared at many such loci and, because PNETs may represent arrested differentiation, we compared tumor mRNA profiles with those of normal mature and progenitor islet cells[37] (see Supplementary Table 6). In all PNETs with mRNA data, the mature cell signature was stronger than that of progenitors (see Extended Data Fig. 5), suggesting that ARX+ and PDX1+ non-functional PNETs partially resemble mature α- and beta β-cells, respectively.

Detailed clinical follow-up was documented on the 61 Dutch *MEN1*-mutant cases[38] (see Supplementary Table 5). ARX+ and PDX1+ non-functional PNETs ($n = 47$) did not differ by tumor size or grade (Fig. 4a), and the range of sizes was similar to that of insulinomas (see Extended Data Fig. 6a). With 24 months of median follow-up (longest 8 years), all relapses occurred in the liver and only in ARX+ or DN cases; no PDX1+ or DP PNETs recurred (Fig. 4b and see Extended Data Fig. 6b). A study was then made of 67 unselected cases from Massachusetts (see Supplementary Table 5), where IHC revealed roughly equal fractions of ARX+ and PDX1+ PNETs, but a larger proportion of DN tumors than the previous cohorts (Fig. 4c). Tumor size was known for 61 cases and clinical outcomes for 56 of the 60 non-functional PNETs. ARX+ and PDX1+ tumors were similar in size (see Extended Data Fig. 6c). All relapses over 66 months of median follow-up (longest > 15 years) occurred at distant sites, mostly in patients with ARX+ or DN tumors (Fig. 4d and see Extended Data Fig. 6d; $P = 0.02$); only one PDX1+ and two DP tumors relapsed. Thus, PDX1 expression correlated with favorable prognosis in two distinct PNET cohorts.

Telomere-specific FISH was used to determine ALT status, which is associated with *ATRX* and *DAXX* mutations[12–14], in 50 Dutch *MEN1*-related and 62 American sporadic PNETs (see Extended Data Fig. 6e). Of 27 sporadic ARX+ and DN tumors, 13 tumors (48.1%) showed ALT compared with 14.3% of 35 PDX1+ and DP tumors ($P < 0.005$, Fisher's exact test); similar relationships appeared in *MEN1*-mutant cases (see Extended Data Fig. 6f). ALT was associated with disease relapse, as expected[12–14] (see Extended Data Fig. 6g), but was more informative when combined with PNET subtype: relapses occurred in every ARX+ALT+ tumor, only 9% of ARX+ALT− cases and just one PDX1+ALT− case (Fig. 4e).

Among the 103 total cases with clinical follow-up, 83 had data on subtype, size, ALT and WHO grade. The odds ratio (OR) for relapse of ARX+ or DN cases was higher (14.45, 95% confidence interval (CI) 1.79–116.61) than ORs for 2-cm size (1.14, CI 0.12–11.07) and even 3 cm (8.47, CI 2.11–34.02). Even excluding DN and DP cases (remaining $n = 64$), the OR for relapse of ARX+ PNETs (10.31, CI 1.25–84.73), was higher than for tumors > 2 cm (8.13, CI 0.99–66.97)

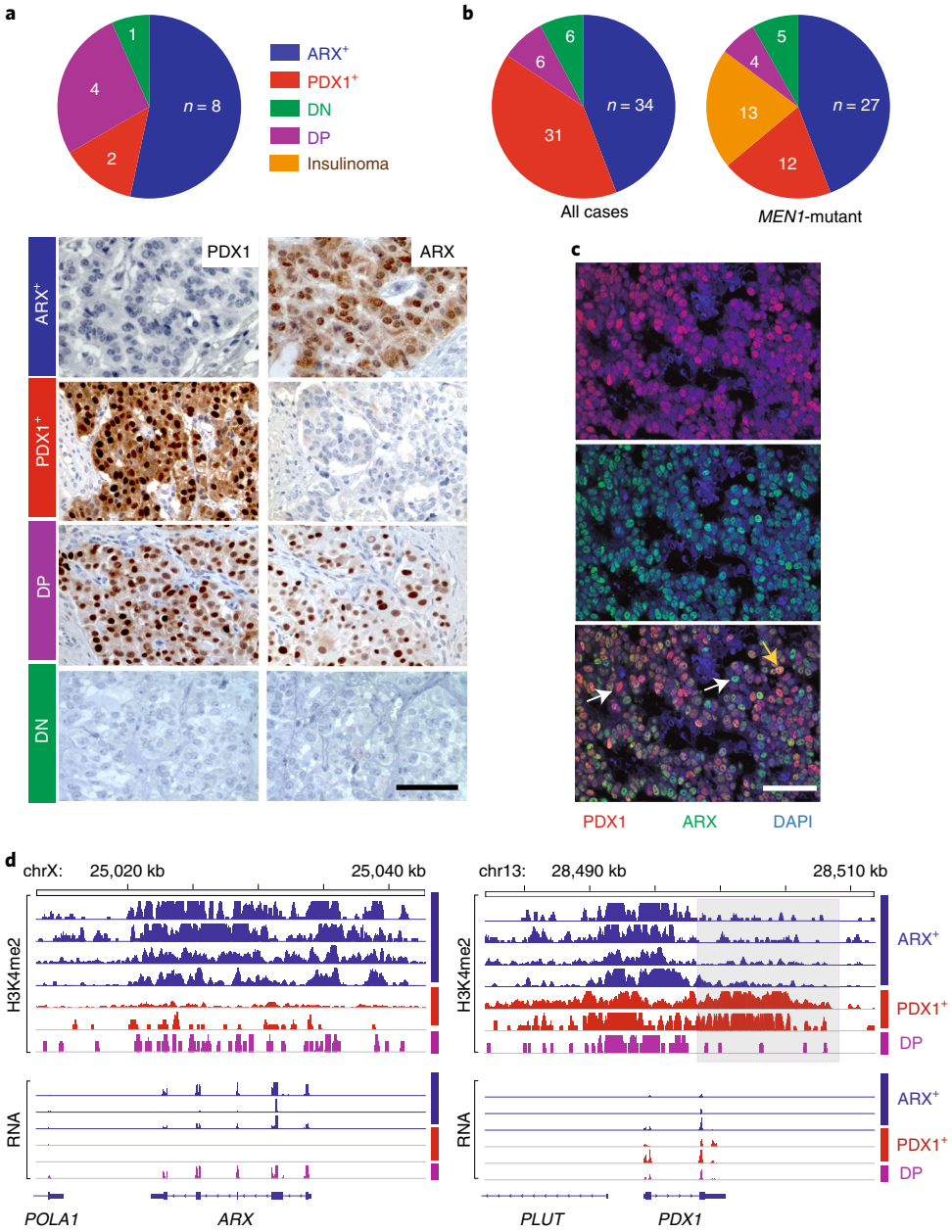
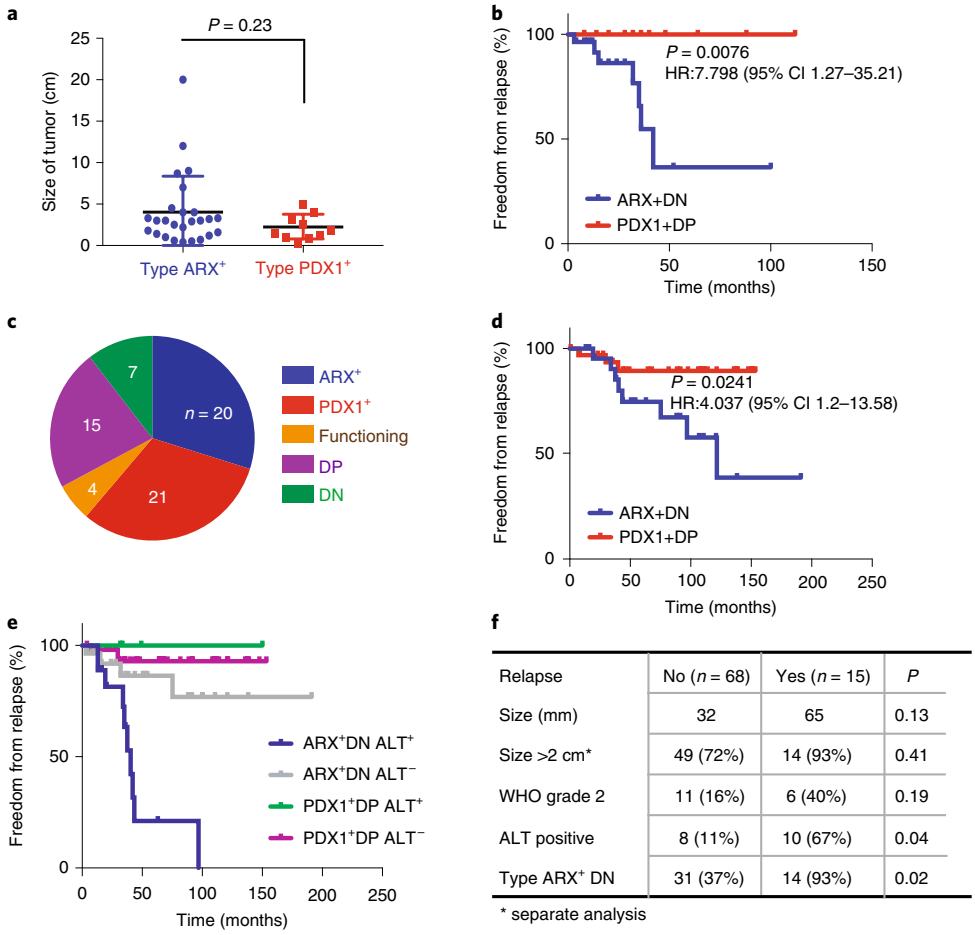


Fig. 3 | ARX and PDX1 immunostains distinguish PNET subtypes. a, Assignment of PNETs in a small independent cohort to subtypes based on immunostaining for ARX (type A), PDX1 (type B), both TFs (DP) or neither (DN). Numbers in each slice indicate independent biological samples, and representative data from each subtype illustrate specificity, which allowed unambiguous assignment. **b**, PNET subtype distribution in a large ($n=77$) cohort, dominated by 61 *MEN1*-mutant cases. Numbers in each slice indicate independent biological samples. **c**, Double immunofluorescence for PDX1 (red) and ARX (green) in a DP tumor, showing TF co-expression in most cells (yellow arrow), with solitary ARX or PDX1 staining in the minority. Scale 50 µm. Similar results were evident in all four examples tested. **d**, H3K4me2 and mRNA levels in PNETs sub-typed by immunostaining, showing that differential histone modifications and transcript levels match the assignments. Every evaluated case is shown. ChIP-seq signals are scaled by promoter-based DESeq2 normalization. RNA-seq signal is scaled to 0–2 RPKM. chr, chromosome.

or >3 cm (9.6, CI 1.94–47.44). Relapses occurred in patients with WHO grade 1 PNETs, and considering all variables by multiple logistic regression, tumor grade was a poor independent risk factor (Fig. 4f). Only ALT and especially absence of PDX1 correlated independently with relapse (Fig. 4f); PDX1$^+$ PNETs rarely relapsed.

In summary, enhancer profiles in overtly similar non-functional PNETs revealed superficial similarities with islet α- or β-cells, reflected in IHC delineation of ARX$^+$, PDX1$^+$ and fewer ARX$^+$PDX$^+$ (DP) tumors. Among 103 cases followed at length after surgery, distant relapses predominated in patients with ARX$^+$ tumors, occurring in only three PDX1$^+$ or DP cases. This favorable association with PDX1 expression can be applied rapidly in the clinic. Specifically, patients with small PDX1$^+$ tumors may be reassured and followed conservatively, whereas vigilant monitoring in patients with PDX1$^-$ tumors may detect early metastases amenable to surgical or medical treatment. Consideration of ALT$^+$ status, which correlates with ARX expression, adds prognostic information, but is less practical than IHC in clinical laboratories. The superior prognosis of PDX1$^+$ over ARX$^+$ PNETs matches the indolent and aggressive disease courses, respectively, of



f

Relapse	No ($n=68$)	Yes ($n=15$)	P
Size (mm)	32	65	0.13
Size >2 cm*	49 (72%)	14 (93%)	0.41
WHO grade 2	11 (16%)	6 (40%)	0.19
ALT positive	8 (11%)	10 (67%)	0.04
Type ARX+ DN	31 (37%)	14 (93%)	0.02

* separate analysis

Fig. 4 | Different prognosis of PNET subtypes. a, Size of all 36 primary tumors with data available for size and PNET type in the Dutch cohort ($P=0.23$, two-tailed Mann–Whitney U-test). Bars represent mean ± s.d. **b**, Kaplan–Meier analysis of disease-free survival in all 47 Dutch cases with available data, showing significant recurrence of ARX+ and DN tumors considered together ($P=0.0076$, two-sided log-rank test; hazard ratio (HR) = 7.8, Mantel–Haenszel test; CI is indicated). No PDX1+ tumor recurred, and outcomes of ARX+ tumors (ungrouped from DN) are shown in Extended Data Fig. 6b. **c**, Distribution of PNET subtypes in the MGH cohort ($n=67$; numbers in each slice indicate independent biological samples). **d**, Kaplan–Meier analysis of disease-free survival in all 55 MGH cases with available data, confirming the worse prognosis associated with ARX+ and DN tumors considered together ($P=0.024$, two-sided log-rank test; HR = 4.04, CI 1.2–13.58). Data for ARX+ and PDX1+ tumors, separated from DP and DN, are in Extended Data Fig. 6d. **e**, Kaplan–Meier analysis of disease-free survival in both cohorts, with cases sub-typed for both TF expression and ALT. All ARX+ and DN tumors with ALT recurred. Although PDX1+ tumors gave excellent outcomes irrespective of ALT status (the only recurrence lacked ALT), ARX+ and DN tumors without ALT had an intermediate prognosis. **f**, Multiple logistic regression analysis of the MGH and Dutch cohorts combined ($n=83$—all cases with relevant data, including 15 relapses). ARX+ and DN status are strong prognostic indicators.

metastatic insulinomas and glucagonomas[1,3]; thus, PNETs that resemble β-cells have better clinical outcomes irrespective of hormonal activity and disease extent. RNA analyses in mouse and human PNETs previously found a group with high *Insulin* expression[39], but did not identify tumor subtypes or different clinical outcomes. Conversely, Chan et al. found enriched α-cell RNA signatures and worse prognosis in *ATRX*-, *DAXX*- or *MEN1*-mutant PNETs[40]. The new, clinically actionable differences reported among PNET types illustrate a general strategy to stratify cancers by epigenetic landscapes and cell lineage, with prognostic implications.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and associated accession codes are available at https://doi.org/10.1038/s41591-019-0493-4.

Received: 28 June 2018; Accepted: 21 May 2019;
Published online: 1 July 2019

References

1. Metz, D. C. & Jensen, R. T. Gastrointestinal neuroendocrine tumors: pancreatic endocrine tumors. *Gastroenterology* **135**, 1469–1492 (2008).
2. Yao, J. C. et al. One hundred years after 'carcinoid': epidemiology of and prognostic factors for neuroendocrine tumors in 35,825 cases in the United States. *J. Clin. Oncol.* **26**, 3063–3072 (2008).
3. Kulke, M. H. et al. NANETS treatment guidelines: well-differentiated neuroendocrine tumors of the stomach and pancreas. *Pancreas* **39**, 735–752 (2010).
4. Lawrence, B. et al. The epidemiology of gastroenteropancreatic neuroendocrine tumors. *Endocrinol. Metab. Clin. North Am.* **40**, 1–18 (2011). vii.
5. Collombat, P. et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev.* **17**, 2591–2603 (2003).
6. Gannon, M. et al. Pdx-1 function is specifically required in embryonic beta cells to generate appropriate numbers of endocrine cell types and maintain glucose homeostasis. *Dev. Biol.* **314**, 406–417 (2008).
7. Rindi, G. et al. Nomenclature and classification of neuroendocrine neoplasms of the digestive system. In *WHO Classification of Tumours of the Digestive System* 4th edn (eds Bosman, F. T. et al.) (International Agency for Research on Cancer, 2017).
8. 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).

9. Kouvaraki, M. A. et al. Surgical treatment of non-functioning pancreatic islet cell tumors. *J. Surg. Oncol.* **89**, 170–185 (2005).
10. Jiao, Y. et al. DAXX/ATRX, MEN1, and mTOR pathway genes are frequently altered in pancreatic neuroendocrine tumors. *Science* **331**, 1199–1203 (2011).
11. Scarpa, A. et al. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature* **543**, 65–71 (2017).
12. Heaphy, C. M. et al. Altered telomeres in tumors with ATRX and DAXX mutations. *Science* **333**, 425 (2011).
13. Marinoni, I. et al. Loss of DAXX and ATRX are associated with chromosome instability and reduced survival of patients with pancreatic neuroendocrine tumors. *Gastroenterology* **146**, 453–460 (2014).
14. Singhi, A. D. et al. Alternative lengthening of telomeres and loss of DAXX/ATRX expression predicts metastatic disease and poor survival in patients with pancreatic neuroendocrine tumors. *Clin. Cancer Res.* **23**, 600–609 (2017).
15. Zhou, V. W., Goren, A. & Bernstein, B. E. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* **12**, 7–18 (2011).
16. Rivera, C. M. & Ren, B. Mapping human epigenomes. *Cell* **155**, 39–55 (2013).
17. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
18. Parker, S. C. et al. Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl Acad. Sci. USA* **110**, 17921–17926 (2013).
19. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
20. Wang, X. et al. SMARCB1-mediated SWI/SNF complex function is essential for enhancer regulation. *Nat. Genet.* **49**, 289–295 (2017).
21. Ackermann, A. M., Wang, Z., Schug, J., Naji, A. & Kaestner, K. H. Integration of ATAC-seq and RNA-seq identifies human alpha cell and beta cell signature genes. *Mol. Metab.* **5**, 233–244 (2016).
22. Wang, H. et al. Insights into beta cell regeneration for diabetes via integration of molecular landscapes in human insulinomas. *Nat. Commun.* **8**, 767 (2017).
23. Barski, A. et al. High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837 (2007).
24. Larsen, H. L. & Grapin-Botton, A. The molecular and morphogenetic basis of pancreas organogenesis. *Semin. Cell Dev. Biol.* **66**, 51–68 (2017).
25. Sosa-Pineda, B., Chowdhury, K., Torres, M., Oliver, G. & Gruss, P. The Pax4 gene is essential for differentiation of insulin-producing beta cells in the mammalian pancreas. *Nature* **386**, 399–402 (1997).
26. Collombat, P. et al. The simultaneous loss of Arx and Pax4 genes promotes a somatostatin-producing cell fate specification at the expense of the alpha- and beta-cell lineages in the mouse endocrine pancreas. *Development* **132**, 2969–2980 (2005).
27. Collombat, P. et al. Embryonic endocrine pancreas and mature beta cells acquire alpha and PP cell phenotypes upon Arx misexpression. *J. Clin. Invest.* **117**, 961–970 (2007).
28. Sussel, L. et al. Mice lacking the homeodomain transcription factor Nkx2.2 have diabetes due to arrested differentiation of pancreatic beta cells. *Development* **125**, 2213–2221 (1998).
29. Yang, Y. P., Thorel, F., Boyer, D. F., Herrera, P. L. & Wright, C. V. Context-specific alpha-to-beta-cell reprogramming by forced Pdx1 expression. *Genes Dev.* **25**, 1680–1685 (2011).
30. Gutierrez, G. D. et al. Pancreatic beta cell identity requires continual repression of non-beta cell programs. *J. Clin. Invest.* **127**, 244–259 (2017).
31. Swisa, A. et al. PAX6 maintains beta cell identity by repressing genes of alternative islet cell types. *J. Clin. Invest.* **127**, 230–243 (2017).
32. Johansson, K. A. et al. Temporal control of neurogenin3 activity in pancreas progenitors reveals competence windows for the generation of different endocrine cell types. *Dev. Cell* **12**, 457–465 (2007).
33. Arda, H. E. et al. A chromatin basis for cell lineage and disease risk in the human pancreas. *Cell Syst.* **7**, 310–322 e314 (2018).
34. Wang, Y. J. et al. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes* **65**, 3028–3038 (2016).
35. Conemans, E. B. et al. Expression of p27(Kip1) and p18(Ink4c) in human multiple endocrine neoplasia type 1-related pancreatic neuroendocrine tumors. *J. Endocrinol. Invest.* **41**, 655–661 (2018).
36. Cejas, P. et al. Chromatin immunoprecipitation from fixed clinical tissues reveals tumor-specific enhancer profiles. *Nat. Med.* **22**, 685–691 (2016).
37. Ramond, C. et al. Understanding human fetal pancreas development using subpopulation sorting, RNA sequencing and single-cell profiling. *Development* **145**, 1–15 (2018).
38. de Laat, J. M. et al. Long-term natural course of pituitary tumors in patients with MEN1: results from the DutchMEN1 study group (DMSG). *J. Clin. Endocrinol. Metab.* **100**, 3288–3296 (2015).
39. 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).
40. 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).

Acknowledgements

The present study has been supported by the Neuroendocrine Tumor Research Foundation (R.A.S., B.E.B., M.H.K. and D.C.C.), the SPORE program in gastrointestinal cancers (P50CA127003—National Cancer Institute, R.A.S.), the North American Neuroendocrine Tumor Society (C.M.H.) and a grant (no. PI18-01604 to P.C.) from Instituto de Salud Carlos III of the Spanish Economy and Competitiveness Ministry.

C.R.C. Pieterman, B. Havekes, A.R. Hermus, O.M. Dekkers, W.W. de Herder, M.L. Drent, A.N.A. van der Horst-Schrivers and P.H. Bisschop contributed to the Dutch MEN1 Study Group database and tissue repository. We thank J. Chan for critical reading of the manuscript.

Author contributions

P.C., Y.D., C.B.E., E.S., D.C.C., B.E.B. and R.A.S. designed the study. P.C. performed the experiments. Y.D. performed the computational analyses. P.C., L.A.A.B. and V.D. analyzed immunohistochemistry data. C.B.E., M.B., E.G., H.J.W., N.S., A.F.-T. and H.W.L. coordinated ChIP- and RNA-seq efforts. K.M.A.D., E.B.C., L.A.A.B., F.H.M.M., G.D.V., M.R.V., C.F.-d.C., C.F., T.A., A.D.S., E.S., M.H.K. and D.C.C. obtained and curated tissue collections and clinical data. P.C. and K.M.A.D. analyzed clinical data. M.K.G. and C.M.H. performed and scored telomere-specific FISH for ALT. B.E.B. and R.A.S. supervised the study. Y.D., P.C. and R.A.S. wrote the first manuscript draft. K.M.A.D., V.D., M.H.K., D.C.C., B.E.B. and R.A.S. revised the paper.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s41591-019-0493-4.

Supplementary information is available for this paper at https://doi.org/10.1038/s41591-019-0493-4.

Reprints and permissions information is available at www.nature.com/reprints.

Correspondence and requests for materials should be addressed to Y.D., B.E.B. or R.A.S.

Peer review information: J. Carmona was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Methods

Clinical materials. Fresh-frozen pancreatic and intestinal NE tumor specimens were obtained from tissue banks at Massachusetts General Hospital (MGH) and Brigham and Women's Hospital (BWH). FiT-seq and RNA-seq occurred on seven and eight additional FFPE PNET specimens, respectively, in the Dutch cohort, and four FFPE specimens from the discovery set. The epigenome classification was validated by ChIP-seq on 13 additional fresh frozen PNETs and by immunophenotyping 15 FFPE samples from BWH, tissue microarrays (TMAs) representing 77 cases (61 with clinical information) from a Dutch national registry and 67 PNETs (54 with clinical information) from MGH. The TMAs included replicate sections and surrounding normal pancreas. All patients gave informed consent for use of primary tumor tissues and, where shown, clinical outcomes. Institutional Review Boards at Dana-Farber/Harvard Cancer Center and University Medical Center Utrecht approved the studies. Pathologists reviewed H&E-stained tissue sections to confirm diagnoses. See Supplementary Tables 1 and 5.

ChIP-seq and FiT-seq. Sections (30 µm) from frozen samples with >90% tumor enrichment were collected in microtubes, washed in phosphate-buffered saline (PBS), cross-linked with 1% formaldehyde for 5 min, and quenched with 0.125 M glycine for 5 min at room temperature. Cross-linked material was resuspended in 0.1% sodium dodecylsulfate (50 mM Tris-HCl pH 8, 10 mM ethylenediaminetetraacetic acid) and sonicated for 40 min in a Covaris E210 instrument (duty cycle 20%; intensity 8; cycles per burst 200). For FiT-seq, FFPE sections were macrodissected to obtain >80% tumor cell enrichment and 10 sections were used, each 10-µm thick. Sections were washed three times with xylenes to remove paraffin, rehydrated in an ethanol/water series and prepared as described[36], with modified buffer and sonication conditions (Covaris E220 instrument, 20 min, 5% duty cycle, 105 peak incident power, 200 cycles per burst, 1 ml AFA Fiber milliTUBEs). Soluble chromatin (5 µg) was immunoprecipitated with 10 µg H3K27ac (Active Motif 39133) antibody for ChIP-seq and H3K4me2 (Millipore 07-030) or H3K27ac (Diagenode C15410196) antibody for FiT-seq. ChIP-seq libraries were constructed by end repair (Epicentre ER0720) of precipitated DNA, base extension with Klenow fragment (New England Biolabs M0212L) and ligation (New England Biolabs M220S) of Y-form-sequencing adapters (Illumina), followed by 14 cycles of PCR. Between enzymatic steps, fragments of the desired size were enriched using AMPure XP beads (Beckman Coulter). FiT-seq libraries were prepared using ThruPLEX-FD kits (Rubicon Genomics) following the manufacturer's protocols. Then, 36-base-pair (bp) paired-end reads were sequenced on a HiSeq 2500 instrument (Illumina).

RNA-seq. A small piece of each snap-frozen primary tumor was pulverized using the T-prep method (Covaris, catalog no. 520097) on dry ice, followed by implementation of Agencourt RNAadvance (Beckman Coulter, catalog no. A32645) on a Biomek FXP automation workstation (Beckman Coulter) to isolate RNA. We used 150 ng RNA to prepare libraries with TruSeq Stranded Total RNA kits (Illumina, catalog no. RS-122-2301). From FFPE tissues, four 10-µm sections were deparaffinized, rehydrated in an ethanol/water series, and RNA extracted using AllPrep DNA/RNA FFPE kits (Qiagen). Concentrations were measured using the Quant-iT RiboGreen RNA assay (Thermo Fisher) and the quality assessed on an Agilent 2100 Bioanalyzer using the Agilent RNA 6000 Nano kit. Ribosomal (rRNA) and mitochondrial RNAs were removed using biotinylated, target-specific oligonucleotides combined with Ribo-Zero rRNA removal beads and the TruSeq Stranded Total RNA kit (Illumina). The 75-bp single-end reads were sequenced on a NextSeq 500 Instrument (Illumina).

Immunostaining. Tissue sections were deparaffinized in xylenes and hydrated through an ethanol and water series. Antigens were retrieved in 10 mM sodium citrate (pH 6) and 0.05% Tween-20 in a pressure cooker for 3 min. After cooling and rinsing with PBS, slides were treated with 3% H_2O_2 in PBS for 10 min to quench endogenous peroxidases, washed and incubated in blocking solution (PBS containing 1% BSA and 1% Tween-20) for 1 h at ambient temperature. Slides were incubated with ARX (Thermo Fisher AF7068SP, dilution 1:300), PDX1 (Abcam ab134150, 1:300), NEUROG3 (Fisher Scientific AB5684, 1:50) or SST (Agilent A056601-2, 1:10,000) antibody diluted in blocking solution for 1 h. Slides stained for PDX1, SST or NEUROG3 were washed in PBS and incubated with the peroxidase-based EnVision Kit (Dako). Slides stained for ARX were incubated with horse radish peroxidase-conjugated anti-sheep antibody (Santa Cruz sc-2473, 1:500 in blocking solution) for 30 min, followed by VectaStain Elite ABC kit (Vector) for 30 min, washed, developed using 3,3′-diaminobenzidine (Dako) and counterstained with Mayer's hematoxylin. For double immunostaining, slides were blocked with 5% FBS in PBS for 1 h at room temperature and incubated overnight at 4 °C with ARX (1:100 in PBS, 1% BSA, 0.3% Triton X-100) and PDX1 (1:500) antibody. Slides were then incubated with a biotinylated anti-sheep antibody (Life Technologies A16045, 1:250), followed by a mix of Alexa Fluor 488-conjugated streptavidin (Life Technologies S11223) and Alexa Fluor 546-conjugated anti-rabbit serum antibody (Life Technologies A11035, 1:250) for 1 h, washed in PBS and mounted in DAPI$^+$ Vectashield medium (Vector). Associations of TF expression with clinical variables were calculated using the statistical package SPSS v.19.0.0.

Telomere-specific FISH and assessment of ALT. Deparaffinized TMA slides were hydrated, steamed for 25 min in citrate buffer (Vector Laboratories), dehydrated, and hybridized with a Cy3-labeled peptide nucleic acid probe complementary to the mammalian telomere repeat sequence. An Alexa Fluor 488-labeled peptide nucleic acid probe specific to human centromeric DNA was included as a positive control. After post-hybridization washes, slides were counterstained with DAPI. ALT$^+$ tumors were identified by telomere length heterogeneity and the presence of large, ultra-bright foci of nuclear FISH signals in ≥1% of tumor cells[12,14]. Necrotic areas were excluded from consideration and two individuals blinded to the PNET subtype scored the slides, with 100% concordance.

Computational and statistical analyses. ChIP-seq reads were aligned to the reference genome (hg19) using BWA 0.7.10 (ref. [41]). Reads with mapping scores <10 were discarded, and those aligned to the same position and strand were counted only once. Density signals were calculated using igvtools and visualized on the Integrated Genome Viewer[42]. H3K27ac and H3K4me2 peaks (style 'histone' for variable length, local filtering disabled) were called using HOMER[43]. A union set of H3K27ac$^+$ peaks was generated by taking the top 50,000 from each sample and merging those peaks using bedtools merge[44]. Reads were acetylation signals at enhancers, estimated from ChIP-seq reads that fell within this union set and counted using featureCounts[45].

H3K27ac profiles were compared using DESeq2 (ref. [46]). Promoter signals (<2 kb downstream and <2.5 kb upstream of transcription start sites) vary less than those at enhancers; therefore, only promoter signals were used to estimate normalization (size) factors for each library by DESeq2. When comparing libraries normalized by these size factors, only enhancers with average normalized read counts ≥100 were considered. *P* values were calculated using Wald's test and FDRs using the Benjamini–Hochberg method[47]. Super-enhancers were called by ROSE[19] from H3K27ac peaks using default parameters. Public H3K27ac ChIP-seq data, obtained from the following sources, were re-analyzed using the above methods:

Pancreatic ductal adenocarcinoma[48]: Gene Expression Omnibus (GEO) accessions GSM2131266 and GSM2131280

Gastric adenocarcinoma[49]: GEO accessions GSM1969645 and GSM1969657

Colorectal cancer[50]: GEO accessions GSM2058055 and GSM2058056

Normal islets[51]: ArrayExpress accession E-MTAB-1919

α-Cell- and β-cell-specific enhancers were obtained from Supplementary Table 4 in Arda et al.[33]. Enhancers shared with PNET types A (ARX$^+$) and B (PDX1$^+$) were determined by bedtools intersect[44], and enrichment *P* values were calculated using Fisher's exact test. To cluster PNETs by enhancer signatures (see Fig. 2a and Extended Data Fig. 1c), hierarchical clustering of Pearson's correlations of H3K27ac profiles (quantified by $\log_2$ read counts) was performed, using the shortest distance linkage.

RNA-seq reads were mapped to the hg19 reference genome using STAR 2.5.2 (ref. [52]) and GENCODE v.19 annotations; reads with mapping scores <10 were discarded. To compare RNA profiles, reads were counted with featureCounts[45] and compared using DESeq2. *P* values were calculated using Wald's test and FDRs using the Benjamini–Hochberg method, considering only genes with an average normalized read count ≥100 and using a Cook's cutoff of 20. RNA-seq data on normal pancreatic islets[53] were obtained from GEO accessions GSM1303932 and GSM1303934. α-Cell- and β-cell-specific genes were obtained from Supplementary Table 2 in Wang et al.[34]. Enrichment *P* values of α- and β-beta cell-specific genes were calculated by Fisher's exact test (two-sided). Pancreatic endocrine cell mRNA signatures were obtained from Supplementary Table 1 in Ramond et al.[37], considering populations A and B as mature, and populations C and D as progenitors (see Extended Data Fig. 5). Correlations of tumors to these Mature and Progenitor signatures were based on Spearman's correlation of $\log_2$(transcript count per million (TPM) + 1) values of the tumor and the average $\log_2$(TPM + 1) values of Mature and Progenitor islet cell populations.

ALT status and PNET immunohistochemistry were compared using two-sided Fisher's exact test. Tumor sizes between groups were compared using the two-tailed Mann–Whitney *U*-test. Disease-free survival was analyzed using the log-rank test and Kaplan–Meier plots were generated using GraphPad Prism v.7. To assess the prognostic value of different variables, we applied SPSS Statistics for multiple logistic regression analysis.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

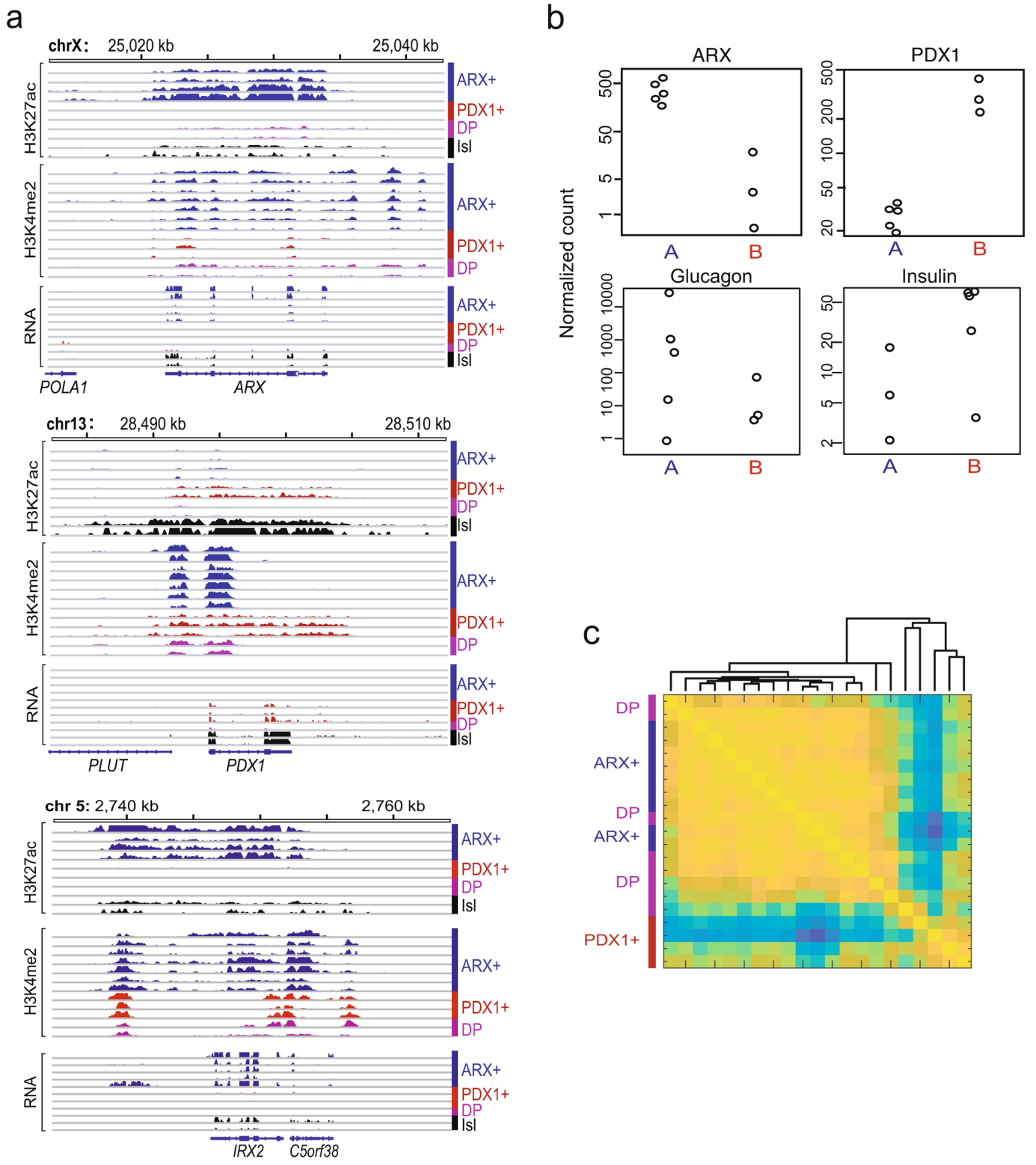
Data availability

All relevant data are included in the manuscript and/or in its supplementary information files. ChIP-seq and RNA-seq data have been deposited in the National Center Biotechnology Information's GEO under GSE116356. Other original data that support the findings of this study have been uploaded as Source Data.

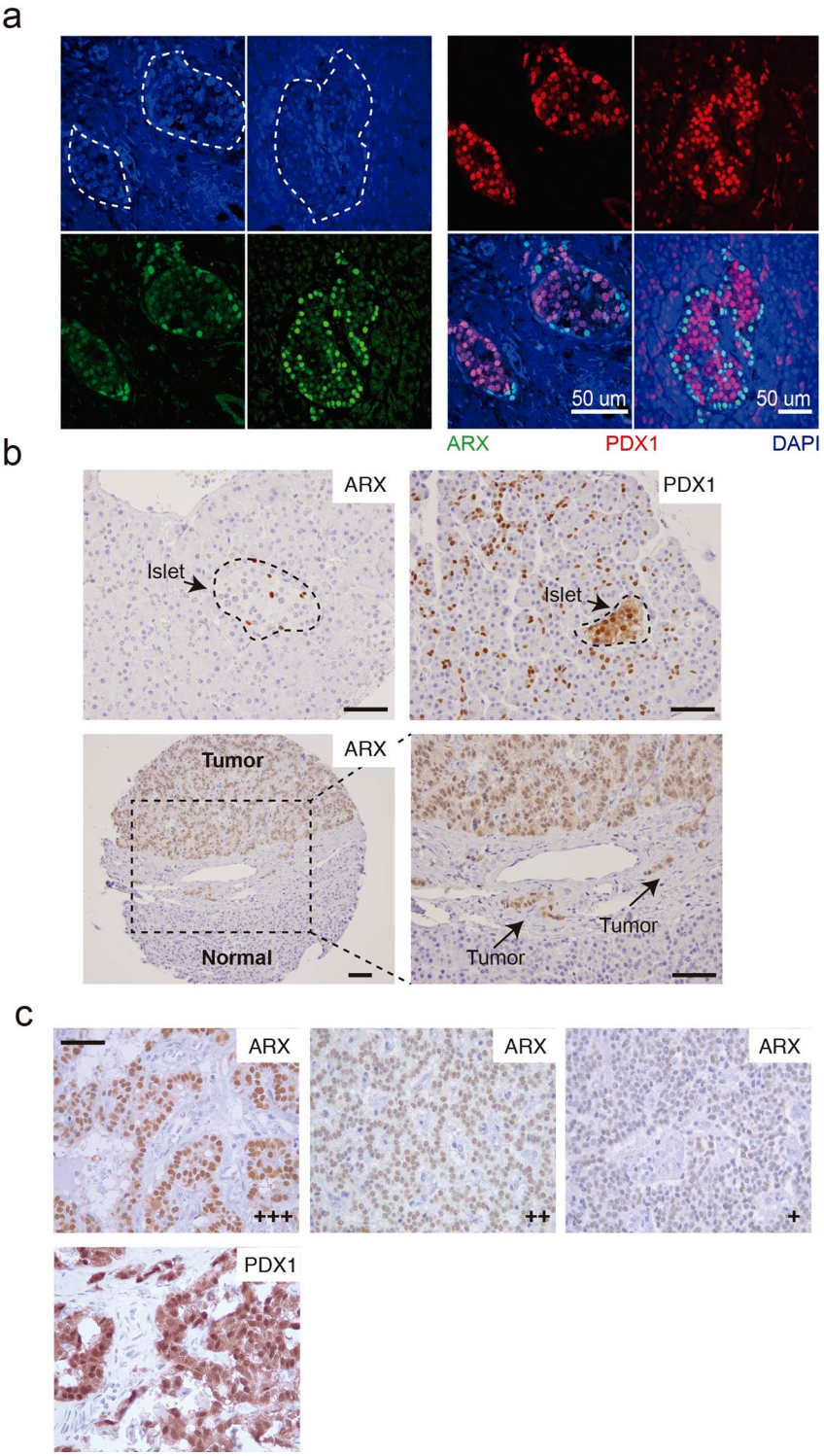
References

41. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

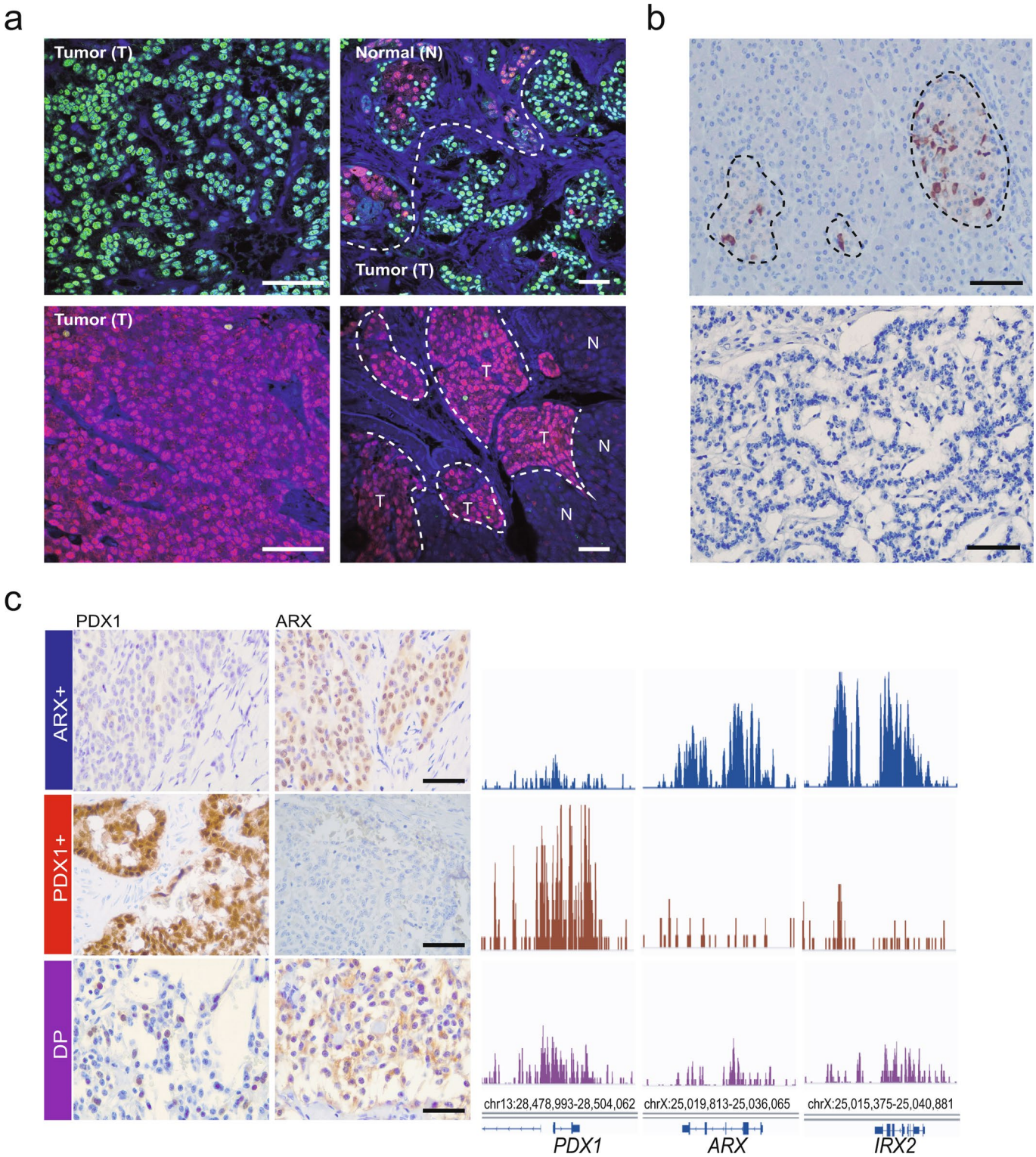
42. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).
43. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
44. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
45. Liao, Y., Smyth, G. K. & Shi, W. Feature counts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
46. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
47. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* **57**, 289–300 (1995).
48. McDonald, O. G. et al. Epigenomic reprogramming during pancreatic cancer progression links anabolic glucose metabolism to distant metastasis. *Nat. Genet.* **49**, 367–376 (2017).
49. Ooi, W. F. et al. Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. *Nat. Commun.* **7**, 12983 (2016).
50. Cohen, A. J. et al. Hotspots of aberrant enhancer activity punctuate the colorectal cancer epigenome. *Nat. Commun.* **8**, 14400 (2017).
51. Pasquali, L. et al. Pancreatic islet enhancer clusters enriched in type 2 diabetes risk-associated variants. *Nat. Genet.* **46**, 136–143 (2014).
52. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
53. Cnop, M. et al. RNA sequencing identifies dysregulation of the human pancreatic islet transcriptome by the saturated fatty acid palmitate. *Diabetes* **63**, 1978–1993 (2014).



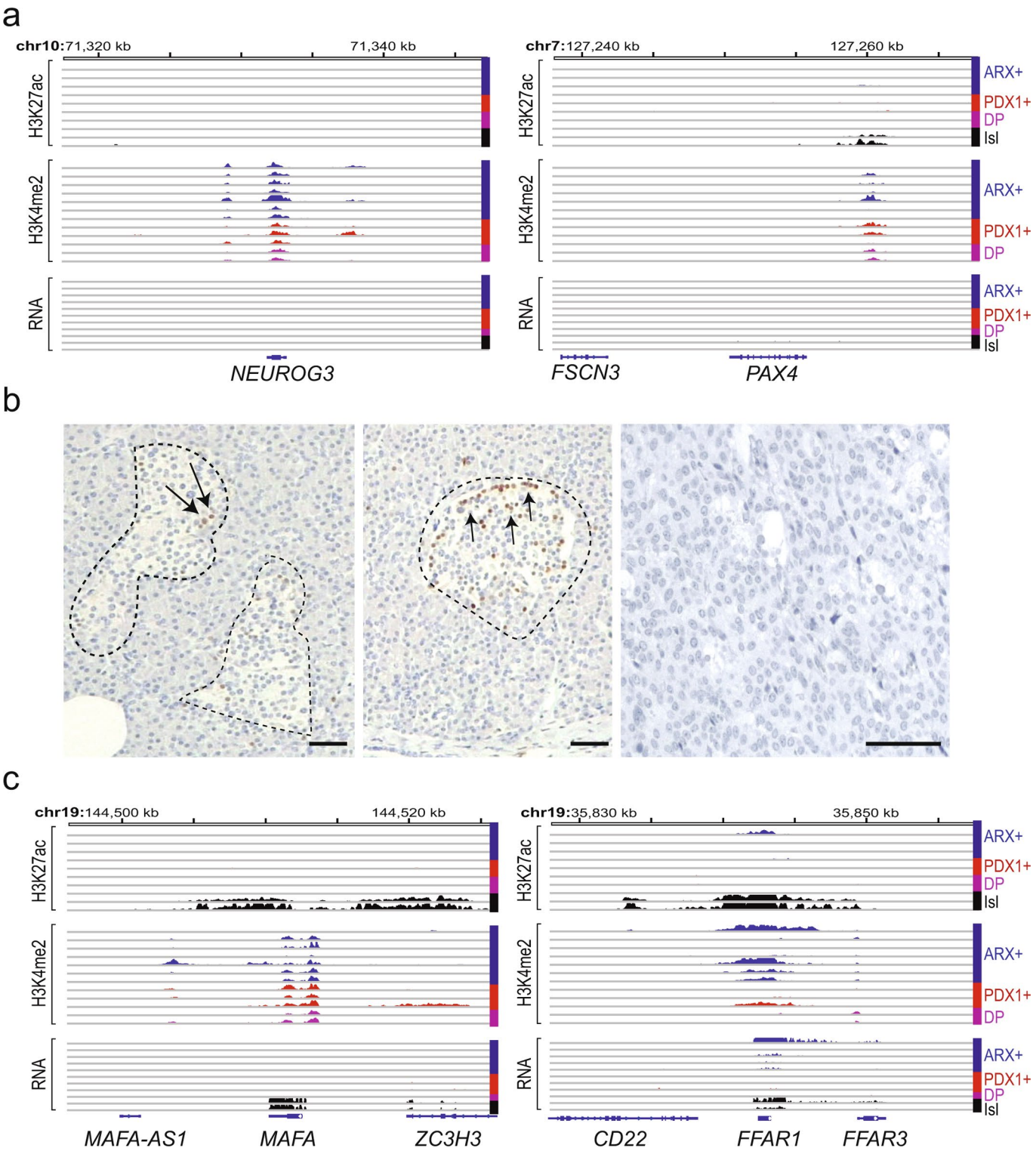
Extended Data Fig. 1 | PNET subtypes are associated with distinct enhancers of lineage-restricted TFs. a, H3K27ac, H3K4me2 and mRNA data tracks at *ARX* and *PDX1* in all eight PNETs from the discovery set and from two samples of normal islets of Langerhans (Isl). ChIP-seq signals are scaled by promoter-based DESeq2 normalization (see Methods) and mRNA read counts are normalized by total read numbers (*y* axis represents 0–2 fragments per million reads). **b**, Distributions of *ARX* and *PDX1* mRNA levels in A- and B-type PNETs. **c**, Pearson's correlations of H3K27ac signals at PNET type A/α-cell and type B/β-cell enhancers in all 21 tumors from the discovery and validation cohorts ($n=8$ and $n=13$ biologically independent samples, respectively).



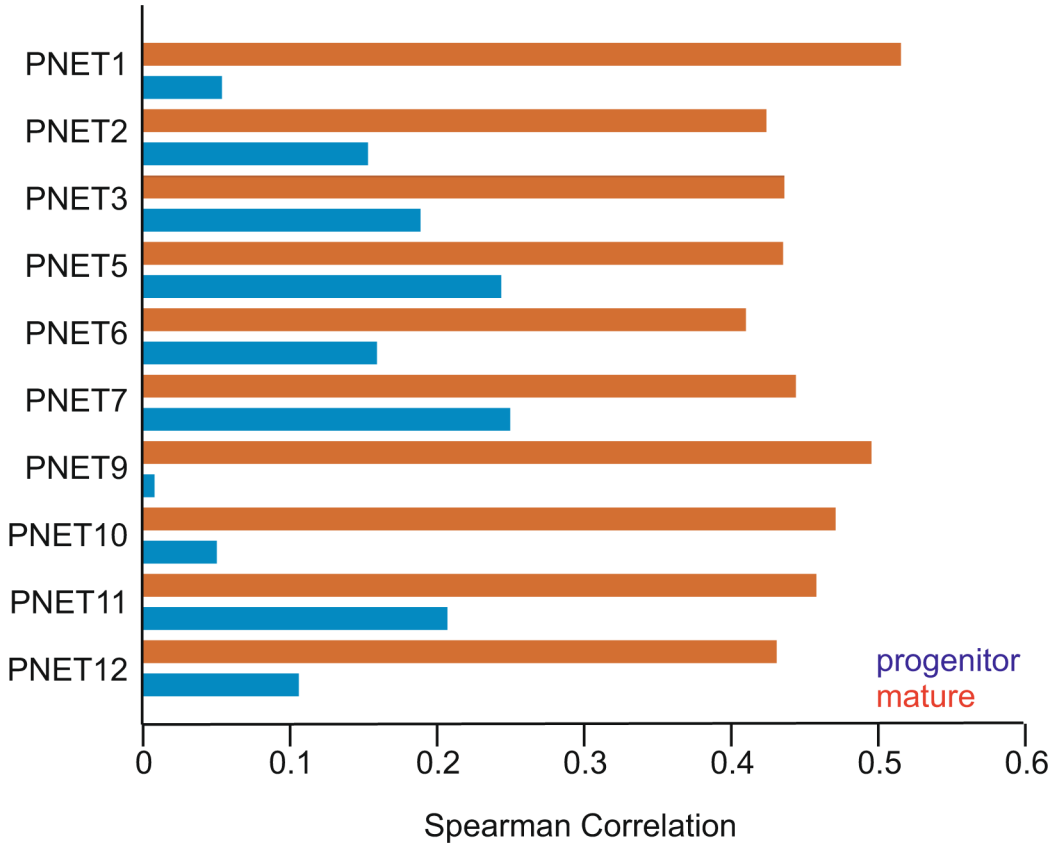
Extended Data Fig. 2 | ARX and PDX1 immunostain in human normal islets and PNETs. a, Double immunofluorescence for PDX1 (red) and ARX (green) in normal islets (marked by dashed white outlines). Scale bar, 50 μm. The results, representing hundreds of islets, verify antibody specificity, lineage-restricted expression and cell distributions: abundant PDX1$^+$ β-cells scattered across islets and fewer ARX$^+$ α-cells enriched in the islet periphery. **b**, Top: ARX and PDX1 IHC selectively mark endocrine α- and β-cells, respectively, in normal human islets. Many exocrine and ductal cells also express PDX1, as is well known[24]. The results represent hundreds of normal islets from multiple individuals, which revealed no ARX$^+$ PDX1$^+$ DP cells. Thus, although described in rodent embryos[24], such cells are absent or extremely rare in the adult human pancreas. Bottom: IHC for ARX in a representative PNET and surrounding normal cells on TMAs from the Dutch cohort. The area boxed in the left image is magnified on the right. ARX$^+$ cells dominate in the tumor and mark invasive foci (arrows). **c**, Range of IHC signal strength in ARX$^+$ PNETs (+weak, ++ moderate, +++ strong), contrasted with uniformly robust PDX1 staining. Images are examples selected from 34 ARX$^+$ and 31 PDX$^+$ cases (Fig. 3b). Scale bars, 50 μm.



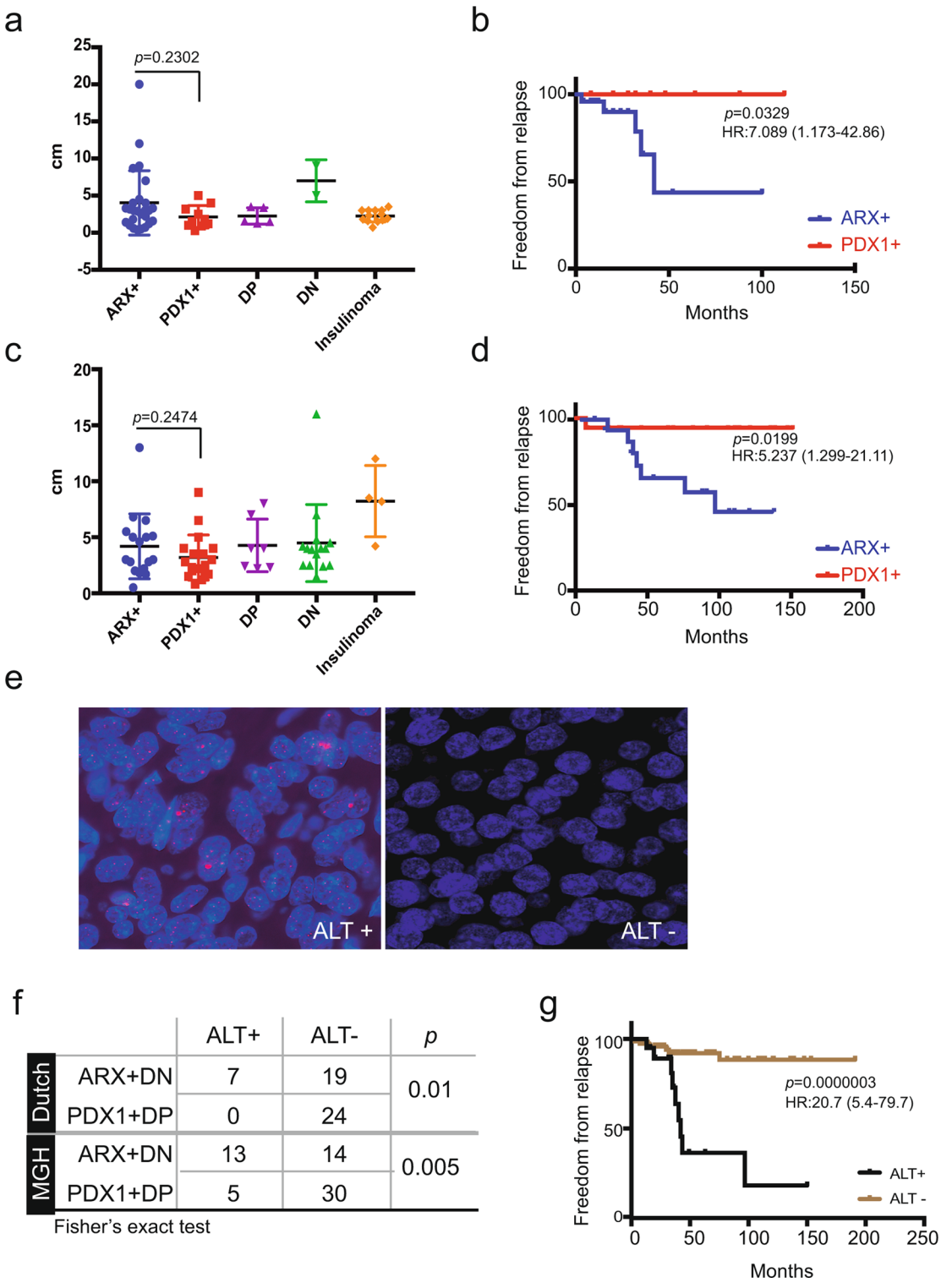
Extended Data Fig. 3 | Additional IHC and enhancer characterization of PNETs. a, Double immunofluorescence of representative ARX$^+$ (type A, $n=34$ biologically independent samples) and PDX1$^+$ (type B, $n=31$ biologically independent samples) tumors (T) adjacent to normal islets (N), showing selective detection of ARX (green) and PDX1 (red), respectively. Lack of antibody cross-reactivity controls for ARX and PDX1 co-staining (Fig. 3c) in DP tumors. **b**, SST expression in normal islets (δ-cells) and absence in all 77 Dutch PNETs, including the representative DN tumor ($n=6$ biologically independent samples) shown here. **c**, IHC results for ARX and PDX1 shown alongside H3K27ac FiT-seq data from the same samples in three of the four cases (one of each subtype) from the discovery cohort where both FFPE and frozen samples were available.



Extended Data Fig. 4 | Other endocrine-specific loci in PNETs. a, H3K27ac, H3K4me2 and mRNA data tracks from all eight PNETs in the discovery set and from two normal islet samples at loci that control early pancreas ontogeny: *NEUROG3* and *PAX4*. Histone marks and RNA-seq data are scaled as in Extended Data Fig. 1a. **b**, IHC for NEUROG3 in rare normal islets (dashed outlines), showing scarce NEUROG3^{+} endocrine cells (arrows). Hundreds of normal islets and all 19 biologically independent PNETs represented on one TMA (one example is shown) lacked expression. **c**, H3K27ac, H3K4me2 and mRNA data tracks from all eight PNETs in the discovery set and from two normal islet samples at loci that control terminal endocrine cell maturation, *MAFA* and *FFAR1*. Histone marks and RNA-seq data are scaled as in Extended Data Fig. 1a. A single outlier showed strong H3K27ac and mRNA at *FFAR1*.



Extended Data Fig 5 | Differentiation status of PNETs. a, Correlations of mRNA profiles in individual PNETs with those of pancreatic endocrine progenitor and mature cells[37]. *x* axis: Spearman's correlations between $\log_2$(TPM + 1) values of each tumor and the average $\log_2$(TPM + 1) values of mature and progenitor populations.



		ALT+	ALT-	p
Dutch	ARX+DN	7	19	0.01
	PDX1+DP	0	24	
MGH	ARX+DN	13	14	0.005
	PDX1+DP	5	30	

Fisher's exact test

Extended Data Fig 6 | Association of PNET subtypes with ALT status. a,**c**, Tumor size in all PNET subtypes in the Dutch (**a**) ($n=56$ independent tumors) and the MGH (**c**) ($n=61$ independent tumors) cohorts. Bars represent mean ± s.d. *P* values for differences in size of primary ARX$^+$ and PDX1$^+$ tumors determined by the two-sided Mann–Whitney *U*-test. **b**,**d**, Analyses of recurrence-free survival in the Dutch (**b**) ($n=30$ cases) and MGH (**d**) ($n=35$ cases) cohorts when ARX$^+$ and PDX1$^+$ tumors were considered separately, ungrouped from DP and DN tumors. *P* values and HRs were determined using two-sided log-rank and Mantel–Haenszel tests, respectively. **e**,**f**, Representative (**e**) (1 example each from 25 independent ALT$^+$ and 87 independent ALT$^-$ cases) and aggregate (**f**) ($n=112$ biologically independent cases) results of telomere-specific FISH in cases classified as positive or negative for ALT. The statistical test was two-sided. **g**, Kaplan–Meier analysis of disease-free survival in all 112 cases with ALT data from both cohorts, without consideration of PNET subtype.

nature research

Corresponding author(s): Ramesh A. Shivdasani

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Standard Illumina NextSeq 500 and HiSeq 2500 processing.
Data analysis	BWA 0.7.10, igvtools 2.3, featureCounts 1.6.2, bedtools 2.26, DESeq 2, HOMER 4.6, ROSE

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

ChIP-seq and RNA-seq data have been deposited in in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE116356 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE116356).

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We did not perform sample size calculation. Rather, we used the largest combined tumor sample and clinical datasets that were available to us.
Data exclusions	Failed chip-seq tracks were excluded (IP efficiency less than 5%, as estimated by ratio of reads within peaks) and ovarian metastatic samples were excluded.
Replication	All attempts of replication are described in the manuscript.
Randomization	The study described does not include any intervention. Therefore randomization of patients does not apply to this study.
Blinding	Immunohistochemical scoring was done blinded to the clinical information

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☐	☒ Unique biological materials
☐	☒ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

Methods

n/a	Involved in the study
☐	☒ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	n/a

Antibodies

Antibodies used	H3K27ac (Active Motif 39133; lot #01613007), H3K27ac (Diagenode C15410196; lot#A1723-0041D), H3K4me2 (Millipore 07-030; lot #2477948), ARX (R&D, AF7068; lot #CFOM0217021), PDX1 (Abcam ab134150; lot #GR97323-11), NEUROG3 (EMD Millipore AB5684; lot # 2519231), SST (Agilent A056601-2). All antibodies with human reactivity.
Validation	H3K27ac (Active Motif 39133), H3K27ac (Diagenode C15410196) and H3K4me2 (Millipore 07-030) are validated for ChIP-seq application as stated on Manufacturer's website. In addition, H3K4me2 (Millipore 07-030) is validated for FiT-seq (Nat. Med 2016, PMID:27111282). ARX (R&D, AF7068), PDX1 (Abcam ab134150), NEUROG3 (EMD Millipore AB5684), SST (Agilent A056601-2) are all validated for immunohistochemistry as stated on the corresponding Manufacturer's websites. We further validated the performance by analysis of the expression in normal pancreatic islets (described in the manuscript).

Human research participants

Policy information about studies involving human research participants

Population characteristics	The relevant characteristics of the participants include tumor site, MEN1 status, ALT status, size of the primary tumor at surgery,

Population characteristics	tumor grade/ mitotic index, time between surgery and relapse/ follow-up and whether the patient had a hormone producing tumor or not. We did also report age and sex of the patients, although these parameters were not assessed as variables in the study.
Recruitment	All patients that underwent surgery of a primary pNET in the hospitals involved in the study during the sample correction period and gave informed consent were included in the study.

ChIP-seq

Data deposition

☒ Confirm that both raw and final processed data have been deposited in a public database such as GEO.

☒ Confirm that you have deposited or provided access to graph files (e.g. BED files) for the called peaks.

Data access links *May remain private before publication.*	GEO accession GSE116356, token sdcjkgmenputxkp
Files in database submission	CAR_10_h3k27ac_13811.bw CAR_11_h3k27ac_8691.bw CAR_12_h3k27ac_8701.bw CAR_13_h3k27ac_13810.bw CAR_14_h3k27ac_13809.bw CAR_15_h3k27ac_13832.bw CAR_16_h3k27ac_13807.bw CAR_17_h3k27ac_13833.bw CAR_18_h3k27ac_7487.bw CAR_19_h3k27ac_7478.bw CAR_1_h3k27ac_8688.bw CAR_20_h3k27ac_13836.bw CAR_21_h3k27ac_8485.bw CAR_2_h3k27ac_13835.bw CAR_3_h3k27ac_13838.bw CAR_4_h3k27ac_13834.bw CAR_5_h3k27ac_8702.bw CAR_6_h3k27ac_7479.bw CAR_7_h3k27ac_8704.bw CAR_8_h3k27ac_8699.bw CAR_9_h3k27ac_13839.bw CAR_A_22_h3k27ac_8679.bw PNET_10_h3k4me2_16537.bw PNET_11_h3k4me2_16544.bw PNET_12_h3k4me2_16549.bw PNET_1_h3k27ac_14046.bw PNET_1_h3k4me2_16548.bw PNET_2_h3k27ac_8649.bw PNET_2_h3k4me2_16520.bw PNET_3_h3k27ac_8656.bw PNET_3_h3k4me2_16530.bw PNET_4_h3k27ac_8482.bw PNET_4_h3k4me2_16512.bw PNET_5_h3k27ac_14048.bw PNET_5_h3k4me2_16524.bw PNET_6_h3k27ac_8644.bw PNET_6_h3k4me2_16536.bw PNET_7_h3k27ac_8658.bw PNET_7_h3k4me2_16519.bw PNET_8_h3k27ac_8657.bw PNET_8_h3k4me2_16547.bw PNET_9_h3k4me2_16590.bw pnet_ffpe1_172_h3k4me2_16102.bw pnet_ffpe2_4532_h3k4me2_16103.bw pnet_ffpe3_14475_h3k4me2_16106.bw pnet_ffpe5_6978_h3k4me2_16104.bw pnet_ffpe6_3685_h3k4me2_16105.bw pnet_ffpe7_14645_h3k4me2_16108.bw pnet_ffpe8_8642_h3k4me2_16107.bw CAR_10_h3k27ac_13811.bed CAR_11_h3k27ac_8691.bed CAR_12_h3k27ac_8701.bed CAR_13_h3k27ac_13810.bed CAR_14_h3k27ac_13809.bed CAR_15_h3k27ac_13832.bed CAR_16_h3k27ac_13807.bed

CAR_17_h3k27ac_13833.bed
CAR_18_h3k27ac_7487.bed
CAR_19_h3k27ac_7478.bed
CAR_1_h3k27ac_8688.bed
CAR_20_h3k27ac_13836.bed
CAR_21_h3k27ac_8485.bed
CAR_2_h3k27ac_13835.bed
CAR_3_h3k27ac_13838.bed
CAR_4_h3k27ac_13834.bed
CAR_5_h3k27ac_8702.bed
CAR_6_h3k27ac_7479.bed
CAR_7_h3k27ac_8704.bed
CAR_8_h3k27ac_8699.bed
CAR_9_h3k27ac_13839.bed
CAR_A_22_h3k27ac_8679.bed
PNET_10_h3k4me2_16537.bed
PNET_11_h3k4me2_16544.bed
PNET_12_h3k4me2_16549.bed
PNET_1_h3k27ac_14046.bed
PNET_1_h3k4me2_16548.bed
PNET_2_h3k27ac_8649.bed
PNET_2_h3k4me2_16520.bed
PNET_3_h3k27ac_8656.bed
PNET_3_h3k4me2_16530.bed
PNET_4_h3k27ac_8482.bed
PNET_4_h3k4me2_16512.bed
PNET_5_h3k27ac_14048.bed
PNET_5_h3k4me2_16524.bed
PNET_6_h3k27ac_8644.bed
PNET_6_h3k4me2_16536.bed
PNET_7_h3k27ac_8658.bed
PNET_7_h3k4me2_16519.bed
PNET_8_h3k27ac_8657.bed
PNET_8_h3k4me2_16547.bed
PNET_9_h3k4me2_16590.bed
pnet_ffpe1_172_h3k4me2_16102.bed
pnet_ffpe2_4532_h3k4me2_16103.bed
pnet_ffpe3_14475_h3k4me2_16106.bed
pnet_ffpe5_6978_h3k4me2_16104.bed
pnet_ffpe6_3685_h3k4me2_16105.bed
pnet_ffpe7_14645_h3k4me2_16108.bed
pnet_ffpe8_8642_h3k4me2_16107.bed
PNET_10_RNAseq.counts.tsv
PNET_11_RNAseq.counts.tsv
PNET_12_RNAseq.counts.tsv
PNET_1_RNAseq.counts.tsv
PNET_3_RNAseq.counts.tsv
PNET_5_RNAseq.counts.tsv
PNET_6_RNAseq.counts.tsv
PNET_7_RNAseq.counts.tsv
PNET_9_RNAseq.counts.tsv
pnet_ffpe10_3536_KL4872.RNAseq.counts.tsv
pnet_ffpe1_172_KL4872.RNAseq.counts.tsv
pnet_ffpe2_4532_I_15_KL4872.RNAseq.counts.tsv
pnet_ffpe3_14475_I3_KL4872.RNAseq.counts.tsv
pnet_ffpe5_6978_II_1_KL4872.RNAseq.counts.tsv
pnet_ffpe7_14645_KL4872.RNAseq.counts.tsv
pnet_ffpe8_8642_I4_KL4872.RNAseq.counts.tsv
pnet_ffpe9_10674_KL4872.RNAseq.counts.tsv
pnet_131_h3k27ac_20226.bw
pnet_131_h3k27ac_20226_fp-stylehistone-L0.bed
pnet_136_h3k27ac_20233.bw
pnet_136_h3k27ac_20233_fp-stylehistone-L0.bed
pnet_14_h3k27ac_20234.bw
pnet_14_h3k27ac_20234_fp-stylehistone-L0.bed
pnet_31_h3k27ac_20232.bw
pnet_31_h3k27ac_20232_fp-stylehistone-L0.bed
pnet_42_h3k27ac_20228.bw
pnet_42_h3k27ac_20228_fp-stylehistone-L0.bed
pnet_46_h3k27ac_20224.bw
pnet_46_h3k27ac_20224_fp-stylehistone-L0.bed
pnet_56_h3k27ac_20230.bw
pnet_56_h3k27ac_20230_fp-stylehistone-L0.bed
pnet_61_h3k27ac_20220.bw
pnet_61_h3k27ac_20220_fp-stylehistone-L0.bed

pnet_70_h3k27ac_20229.bw
pnet_70_h3k27ac_20229_fp-stylehistone-L0.bed
pnet_75_h3k27ac_20221.bw
pnet_75_h3k27ac_20221_fp-stylehistone-L0.bed
pnet_77_h3k27ac_20225.bw
pnet_77_h3k27ac_20225_fp-stylehistone-L0.bed
pnet_80_h3k27ac_20227.bw
pnet_80_h3k27ac_20227_fp-stylehistone-L0.bed
pnet_81_h3k27ac_20231.bw
pnet_81_h3k27ac_20231_fp-stylehistone-L0.bed

Genome browser session
(e.g. UCSC)

Carcinoid data: https://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=yotamd&hgS_otherUserSessionName=carcinoid_h3k27ac
PNET data: https://genome-euro.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=yotamd&hgS_otherUserSessionName=pnet

Methodology

Replicates

N/A

Sequencing depth

All ChIP-seq data are paired-end. Library information is included in Suppl. Table 1.

Antibodies

For ChIP: H3K27ac (Active Motif 39133; lot #01613007), H3K27ac (Diagenode C15410196; lot#A1723-0041D), H3K4me2 (Millipore 07-030; lot #2477948). All antibodies have human reactivity and are validated for ChIP.
For immunohistochemistry and immunofluorescence: ARX R&D Systems AF7068), PDX1 (Abcam ab134150), NEUROG3 (Millipore AB5684), and Somatostatin (Agilent A056601-2) antibodies were used to stain tissue sections after appropriate blocking of specimens to reduce non-specific background signals.

Peak calling parameters

Peaks were called using homer 4.6 with the command findPeaks using parameters -style histone -L 0.
Super enhnacers were called using ROSE (Whyte et al, Cell 2013; 153:307-319)

Data quality

We required IP efficiency of at least 5%, with at least 20,000 peaks detected at FDR of 0.001

Software

BWA 0.7.10, igvtools 2.3, featureCounts 1.6.2, bedtools 2.26, DESeq 2, HOMER 4.6, ROSE