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

Most pancreatic neuroendocrine tumors (PNETs) do not produce excess hormones and are therefore considered ‘non-functional’. As clinical behaviors vary widely and distant metastases are eventually lethal, 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, 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- or PDX1- tumors and, within this subtype, in cases with alternative lengthening of telomeres. These markedly different outcomes bely similar clinical presentations and histology and, in one cohort, occurred irrespective of MEN1 mutation. This robust molecular stratification provides insight into cell lineage correlations 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. Insulinomas resemble normal pancreatic β-cells and carry a good prognosis, but the lineage of most non-functional PNETs is obscure and about half the cases progress to lethal metastasis months to years after surgery. Although 70% of PNETs carry MEN1, ATRX or DAXX gene mutations, and 15% activate mammalian target of rapamycin signaling, no mutation or biological feature is sufficiently correlated with clinical outcomes to guide prognosis or therapy. Alternative lengthening of telomeres (ALT) is associated with ATRX or DAXX loss and elevated risk of recurrence, 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. Super- or ‘stretch’ enhancers, which control lineage-specifying genes, especially delineate cell-specific chromatin signatures and help nominate tumor cell origins. 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 SYIP, 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.

Across the eight PNETs, A- and B-type tumors gave strong H3K27ac signals at ARX and IRX2 or at PDX1, respectively, whereas
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, respectively (Fig. 1c). At these and other loci, ChIP-seq for H3K4me2 (ref. 17) 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 development14. NEUROG3 initiates the endocrine lineage and ARX specifies α-cell fate, whereas PAX4 drives β-cell differentiation15; absence of both factors favors the δ lineage16. ARX is necessary for α-cell differentiation17, but other transcription factors (TFs, for example NKX2.2, PAX6, PDX1) sustain β-cells18–21. In embryos, early α-cell specification leads to Insulin and Glucagon co-expressing bipotential precursors, followed by β-cell differentiation22; forced transdifferentiation between α- and β-cells also occurs through intermediate Insulin+ Glucagon+ cells23. 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-cells24. Type A regions were highly enriched for α-cell–specific sites (62 shared, P < 1.1 × 10⁻¹⁰), whereas type B-specific enhancers were enriched for β-cell–restricted sites (15 shared, P < 1.7 × 10⁻⁴; see Supplementary Table 4). Comparison with RNA profiles of fractionated normal human islets25 also revealed enrichment of α- and β-cell-specific transcripts in A- (P < 1.6 × 10⁻⁴) and B-type (P < 6.2 × 10⁻⁶) 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 ChIP-seq signals at the ARX locus (Fig. 1e). 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).
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 (ISLI and NEUROD1—data not shown), but not at terminal differentiation genes (for example, MAF 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 \(\alpha\) and \(\beta\) 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). Within 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\(^{15-19}\), 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\(^{15-19}\) (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. 4c).

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 for 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. 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 \(\alpha\)- versus \(\beta\)-cells\(^{30,31}\). 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).
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. 76x135). Only ALT and especially absence of PDX1 correlated independently with relapse (Fig. 164x124); 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
metastatic insulinomas and glucagonomas1',2; 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 expression39, 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 PNETs40. 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


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 immunoperoxidase staining of FFPE samples from BWH, tissue microarrays (TMA)s representing 77 cases (61 with clinical information) from a Dutch national registry and 67 PNETs (54 with clinical information) from MGH. The TMA’s included replicate sections and surrounding normal pancreas. All patients gave informed consent for the use of tumor tissues shown, clinical data, and 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 rinsed (3 × 10 ml 0.1 M sodium decyl sulfate (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 1 mm thick. Sections were washed three times with xylene to remove paraffin, rehydrated in an ethanol/water series and prepared as described, with modified buffer and sonication conditions (Covaris E220 instrument, 20 min, 5% duty cycle, 105 peak incident power, 200 cycles per burst, 1 μl Fiber Filament). Soluble chromatin was immunoprecipitated with 10 μg H3K27ac (Active Motif 39131) antibody for ChiP-seq and H3K4me2 (Millipore 07-0370 or H3K27ac (Diagenode C15410196) antibody for FIT-seq. ChiP-seq libraries were constructed by end repair (EpirecQ 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 TruSeq FX-CD 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. M2212L) on dry ice, followed by implementation of Agencourt RNAAdvantage (Beckman Coulter, catalog no. A32645) on a Biomek FXP automatic workstation (Beckman Coulter) to isolate RNA. We used 150 ng RNA to prepare libraries with TruSeq Stranded Total RNA kits (Illumina, catalog no. RS-1222301). 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 Agilent RNA 6000 Nano Kit. Ribosomal (rRNA) and mitochondrial RNA 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 xylene and hydrated through an ethanol and water series. Antigens were in 10 mM sodium citrate (pH 6) and 0.65% Tween-20 in a pressure cooker for 3 min. After cooling and rinsing with PBS, slides were treated with 3% H2O2 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 1:30), PDX1 (Abcam ab151155, 1:100), NEUROG3 (Fishier Scientific, AB5684, 1:50) or SST (Agilent A056601-2, 1:10,000) antibody diluted in blocking solution for 1 h at ambient temperature. 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 cross-reactive rabbit 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% BSA 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-rabbit serum antibody (Life Technologies A10016, 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.

ChIP-seq libraries were enriched using the T-prep method (Covaris, catalog no. 520097) on dry ice, followed by Sonication (Millipore 07-030) or H3K27ac (Diagenode C15410196) antibody for FiT-seq. ChIP-seq libraries were constructed by end repair (Epicentre ER0720) of Agilent RNA 6000 Nano Kit, followed by annealing of 15 bp single-stranded DNA using Agilent Alu6000 Nanotools (Qiagen). Ribosomal (rRNA) and mitochondrial RNA 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).

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


Nature medicine

LETTERS

NATURE


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⁺/−cells scattered across islets and fewer ARX⁺/−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. The results represent hundreds of normal islets from multiple individuals, which revealed no ARX⁺ PDX1⁺ DP cells. Thus, although described in rodent embryos, 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 (β-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\textsuperscript{5.} 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.
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.
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
  - 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

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).
### Life sciences study design

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

<table>
<thead>
<tr>
<th>Sample size</th>
<th>We did not perform sample size calculation. Rather, we used the largest combined tumor sample and clinical datasets that were available to us.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data exclusions</td>
<td>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.</td>
</tr>
<tr>
<td>Replication</td>
<td>All attempts of replication are described in the manuscript.</td>
</tr>
<tr>
<td>Randomization</td>
<td>The study described does not include any intervention. Therefore randomization of patients does not apply to this study.</td>
</tr>
<tr>
<td>Blinding</td>
<td>Immunohistochemical scoring was done blinded to the clinical information.</td>
</tr>
</tbody>
</table>

### 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

#### Antibodies

- 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

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

- tumor grade/miotic 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

| 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_20_h3k27ac_8688.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 |
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 enhancers 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