Title

An Integrative Genetic, Epigenetic and Proteomic Characterization of Pancreatic Neuroendocrine Neoplasms (PanNENs) defines Distinct Molecular Features of α- and β-cell like Subgroups

Authors

Tincy Simon[1,2], Soulafa Mamlouk[1,3], Pamela Riemer [1], Felix Bormann[4], Bertram Klinger[5], Andrea Menne[1,3], Daniel Teichmann[6], Kerstin Wanke-Möhr[1,3], Manuela Pacyna-Gengelbach[1], Slim Khouja[1], Dido Lenze[1], David Capper[3,6], Felix Ruoff[7], Markus. F. Templin[7], Ulf Leser[8], Katharina Detjen[9], Florian Rossner[1], Johannes Haybaeck[10,11], Markus Morkel[1], Nils Bluethgen[1,2,5], David Horst[1], Marianne Pavel[9], Christine Sers[1, 2,3,§]

Affiliations

1. Charité Universitätsmedizin Berlin, Institute of Pathology, Berlin, Germany.
2. CompCancer Research Training Group
3. German Cancer Consortium (DKTK); Partner Site Berlin, German Cancer Research Center (DKFZ), Heidelberg, Germany
4. Bioinformatics.Expert UG, Berlin, Germany
5. Integrative Research Institute (IRI) Life Sciences, Humboldt University Berlin, Germany
6. Charité Universitätsmedizin Berlin, Institute of Neuropathology, Berlin, Germany.
7. NMI Natural and Medical Sciences Institute, the University of Tübingen, Reutlingen, Germany
8. Humboldt-Universität zu Berlin, Knowledge Management in Bioinformatics
9. Charité Universitätsmedizin Berlin, Hepatology and Gastroenterology, Berlin, Germany
10. Department of Pathology, Neuropathology, and Molecular Pathology, Medical University of Innsbruck, Austria
11. Diagnostic & Research Center for Molecular BioMedicine, Institute of Pathology, Medical University of Graz, Austria

§ Corresponding author: Christine Sers: christine.sers@charite.de

Keywords

PanNEN, Methylation-profile, cell-of-origin

Abstract

Pancreatic Neuroendocrine Neoplasms (PanNENs) comprise a rare and heterogeneous group of tumors derived from neuroendocrine cells of the pancreas. Despite recent genetic and epigenetic characterization, biomarkers for improved patient stratification and personalized therapy are sparse and targeted therapies, including the mTOR inhibitor Everolimus, have shown limited success. To better define PanNENs tumors we performed multi-omic analyses on 59 tumors with varying grades (NET G1, NET G2, NET G3 and NEC), combining mutational profiling with epigenetic analysis and targeted proteomics. An unsupervised approach using DNA methylation profiles uncovered two major subgroups in PanNENs resembling α-like and β-like cells. DNA copy number analysis further divided the α-like subgroup into two distinct groups, indicating differential mechanisms of tumorigenesis from α-cells. β-like tumors however showed two distinct groups at the epigenetic level and suggest NET G3/NEC samples are of β-cell origin. DNA mutation profiling clearly separated α and β-cell derived PanNENs, whereby only α-like tumors had mutations within *MEN1/DAXX/ATRX* tumor suppressor genes. Targeted proteomic analysis further indicated overexpression of distinct components of the mTOR pathway. Our data provide further insights into the potential mechanisms behind PanNEN heterogeneity and form a basis for future diagnostic and therapy relevant patient stratification.

Introduction

Pancreatic Neuroendocrine Neoplasms (PanNENs) are rare tumors, with an incidence of approximately 0.5 in 100,000 cases they represent less than 3% of all pancreatic neoplasms [1]. PanNENs originate from the islets of Langerhans, which are composed of five hormone-producing cells; alpha (α), beta (β), delta (δ), epsilon(ε) and pancreas polypeptide (PP) cells. Functional PanNENs produce hormones which trigger medical conditions leading to the initial tumor diagnosis. However, the majority of PanNENs are non-functional; these tumors often remain undetected until their growth causes other symptoms[2]. PanNENs are classified as either well-differentiated Neuroendocrine Tumors (PanNETs) or poorly-differentiated Neuroendocrine Carcinomas (PanNECs PanNENs)[3]. PanNETs are graded by their proliferative index into G1, G2 or G3 tumors with an increasingly malignant nature. PanNECs are by default G3, since they have a high proliferation rate combined with de-differentiation of the cells, resulting in an aggressive phenotype and poor prognosis[3].

The mainstay for PanNEC systemic treatment is platinum-based chemotherapy[4,5], while additional targeted therapies have been approved for the well-differentiated PanNETs. These comprise Sunitinib, a multi-tyrosine kinase receptor inhibitor [6–8] and Everolimus, an inhibitor of the mammalian/mechanistic Target Of Rapamycin Complex 1 (mTORC1) [9,10]. Upregulation of the mTOR pathway has also been identified at the genetic and expression level in PanNENs, yet no biomarker for patient stratification has been implemented until now [11,12]. Clinical trials employing either Sunitinib or Everolimus often showed disease stabilization resulting in increased progression-free survival [13–15]. However, both therapy regimens do not result in complete remission, and secondary resistance and relapse inevitably arise in patients.

In general, PanNENs are characterized by few recurring driver mutations. The most frequent alterations are missense variants and large base-pair deletions detected within *MEN1,* encoding the Menin protein (37%) [12,16,17]. *MEN1* is a tumor suppressor involved in chromatin remodeling, but also regulating mTOR signaling, and DNA damage repair [18–20]. Additionally, PanNEN mutations are detected in *DAXX* and *ATRX* (25% and 18%,

respectively) [21,22], encoding a multifunctional protein and an ATPase/helicase, respectively. Both proteins maintain genomic stability by repressing telomeric lengthening. Furthermore, genes involved in mTOR/PI3K/AKT signaling axes such as *PTEN* and *TSC2* have been implicated in PanNET tumorigenesis[12,16]. In contrast, PanNECs are characterized by mutations in *KRAS, TP53*, and *RB* [4,23].

In addition to driver mutations affecting individual genes, copy number aberrations (CNA) have been used to characterize these tumors[16,24–26]. Recent work has revealed four distinct subtypes based on chromosome arm length CNA patterns: i) recurring copy number losses of specific chromosomes, ii) limited copy number events with mostly loss of chromosome 11, iii) polyploidy tumors and iv) aneuploidy tumors [16]. In addition, the first two groups harbor *MEN1* mutations in combination with loss of chromosome 11, resulting in a bi-allelic inactivation of this tumor suppressor gene[24]. The aneuploid tumors also show frequent *DAXX/ATRX* loss of expression along with an alternative telomerase lengthening phenotype [25].

Comparison of PanNEN *MEN1/DAXX/ATRX*-mutated and -wildtype subgroups by whole transcriptome profiling revealed a gene expression signature similar to α-cells of the Islets of Langerhans only in the mutant group[27]. The specific signature showed a strong expression of the well-established α-cell-specific transcription factor *ARX*. Classification of non-functional PanNENs into two major subgroups resembling islet α- and β-cells was further confirmed by histone modification-based epigenetic stratification and gene expression[28]. Active enhancer marks in combination with expression of the *ARX* gene specifically characterized α-cell-like samples, while *PDX1* exhibited similar epigenetic and expression profile in a subgroup characterizing β-cell-like samples. In addition to this subgrouping resembling distinct cell types of the Islets of Langerhans, it also became clear that α-like samples have a higher relapse rate and a worse prognosis[28].

Despite the growing body of information about PanNENs on multi-genomic (mutations and CNA) and epigenetic levels, these layers have not yet been considered in an integrated manner. Such an approach is challenging, because PanNENs represent a heterogeneous group in terms of prognosis and response to specific therapies. Here, we conducted a multi-genomic approach to provide a comprehensive characterization of

PanNENs. Additionally, and for the first time in the context of PanNENs, we also conducted a multi-protein study using patient samples. We identified α-like and β-like subgroups that strongly stratify patients at the multi-omics levels. Alpha-like tumors were *PDX1* hypermethylated and *IRX2* hypomethylated, harbored *MEN1/ATRX/DAXX* mutations and often exhibited an amplification-rich CNA signature. Beta-like tumors appeared more heterogeneous, displaying a *PDX1* hypomethylated and *IRX2* hypermethylated profile. Furthermore, NEC and NET G3 samples clustered within the β-like group and carried additional hypermethylation of various transcription factors involved in differentiation and maintenance of islet cells.

Results

Sample characteristics

To characterize and classify PanNEN tumors we performed methylation analysis, DNA sequencing and protein level analysis from 59 tumors in total. Samples used for each assay and their characteristics are shown in Figure 1a, further details are depicted in Supplementary Table 1. The cohort includes tumors of grades NET G1 (27) and NET G2 (25) NET G3 (4), and NEC (3), 45 samples were primary PanNENs and 14 were metastases (8 liver, 6 lymph nodes). As healthy control we used matched adjacent healthy (24) or distant healthy tissue (27) or blood (4) from the same patient.

DNA methylation classification identifies two PanNEN groups

To determine PanNEN subtypes we analyzed the methylome of 36 tumors from our cohort using the EPIC 850K bead chip array. We utilized a previously published method to first determine groups within the cohort by clustering probes found in close proximity together[29] (Supplementary Figure 1a). We used the 8000 most variable probe clusters to perform unsupervised class discovery (Supplementary Figure 1b and methods for details). The 8000 most variable probe clusters defined two main subgroups at the methylation level; group A and group B (Figure 1b, Supplementary Figure 1c). Methylation beta values from the 8000 probe clusters revealed notable distinction between group A and group B (Figure 1c). Group A had both hypermethylated and hypomethylated patterns, while group B carried hypermethylated features across all variable probe clusters. The single sample that formed group C showed overall hypomethylation across the probe clusters; this sample was considered an outlier and removed from further analysis. A principal component analysis (PCA) of the 8000 probe cluster methylation β values further highlights the distinction between group A and group B (Figure 1c). Next, differentially methylated probe (DMP) analysis of the two groups led to discovery of more than 50,000 probes distinguishing the groups (Supplementary Table 2). Hierarchical clustering of samples using DMPs reconstructed the groups defined by the probe clusters

(Supplementary Figure 1d). PCA of DMP confirmed two discrete clusters based upon the identified groups, with 40.71% variation described by PC1 (Supplementary Figure 1e). The majority of the DMPs were associated with gene body and intergenic gene regional probes located in open sea CpG positions of the genome (Supplementary Figure 1f). Pearson correlation analysis revealed strong correlation between samples within group A and, albeit to a lesser extent, samples within group B (Supplementary Figure 1g).

We next examined pathways of genes associated with DMPs. Reactome pathway analysis identified 22 significantly enriched pathways (padj < 0.05) (Figure 1d), the most significant was transcription factor AP-2 alpha (TFAP2) regulating receptors and ligands. A closer look at the genes belonging to this pathway identified several genes strongly different between group A and group B (mean abs(Δβ) ≥ 0.2) (Figure 1e, Supplementary Figure 1h). These are epidermal growth factor family receptors and ligands (EGFR, ERBB2, TGFA) and Glycoprotein Hormones, Alpha Polypeptide (CGA) as well as tyrosine-protein kinase KIT and vascular endothelial growth factor A (VEGFA). Probes within the identified genes were found predominantly hypomethylated in group A and strongly hypermethylated in group B. With the exception of *ERBB2* and *CGA*, the majority of probes that showed differential methylation were located in gene bodies (Figure 1e).

In summary, using differential methylation analysis of PanNENs we identified two clear subgroups of tumors. The tumor samples in group A and group B differed strongly from each other in methylation patterns, particularly by key regulators in growth factors involved in tumor growth and angiogenesis.

PanNEN groups show distinct cell-of-origin methylation features belonging to α and β pancreatic islet cells

We next sought to identify and characterize PanNENs belonging to the identified subgroups using the imprinted cell-of-origin pattern persistent in the tumor entity. We first explored recently published single-cell sequencing data and Gene Ontology database to identify islet cell associated marker genes[30–32] (marker list in Supplementary Table 3). We investigated markers which define islet cell differentiation and maintenance at the expression level and which are both early and late islet cell differentiation markers. In

total, we identified 63 markers from which we detected 44 markers associated with the DMPs from our analysis (Figure 2a) (Supplementary Table 4). The DMPs associated with the cell differentiation markers showed enrichment in gene bodies (Supplementary Figure 2a). Genes with differentially methylated regions (mean abs(Δβ) > 0.15) (Supplementary Figure 2b) and those represented in more than one publication/database (in total 15 genes) were further evaluated in our cohort (Figure 2b). The resulting distribution of probe beta values between group A and group B within these 15 genes showed hypermethylation in group B in the majority of genes except *PDX1*, *ETV1* and *BMP5*, which were hypermethylated in group A.

Methylated gene probes in *PDX1*, encoding a β-cell maintenance marker, showed the largest difference between group A and group B tumors (mean abs(Δβ) = 0.34) (Figure 2b). *PDX1* is found hypermethylated within group A tumors, with the exception of 3 cases, PNET52, PNET25 and PNET91, which are the only known insulinomas of the cohort (Figure 2c). These samples showed gene body hypermethylation of *HHEX*, a δ-cell maintenance factor. *HHEX* maintained hypomethylated state in the remaining group A sample set. The known β-cell marker *MEG3* was found with patterns of hypomethylation in a subset of group A (Figure 2c). In contrast, *IRX2*, an α-specific transcription factor, is largely hypermethylated in tumors belonging to group B (mean abs(Δβ) = 0.26) (Figure 2b). Based on this pattern of differentially methylated lineage-specific genes in group A and B, and the exclusive expression pattern of the aforementioned genes by α- and β-cells of the pancreas, we conclude that group A and group B profiles resemble α- and β-cells, respectively.

Eight additional genes harboring DMPs included further islet cell-type specific markers (also representing δ-cells, and PP-cells) as well as pan-endocrine markers of various developmental stages of islet formation (*NKX6-1, SIX3, MEIS1, MEIS2, ID4, NKX2-2, MAFB, PAX6 and NEUROD1*). Interestingly, these markers further subdivided group B into hypo- and hypermethylated subclasses. In general, NET G2, NET G3 and NEC samples clustered within the hypermethylation subclass of group B, while G1 grade samples were hypomethylated for these markers.

In conclusion, we have identified epigenetic patterns related to α-cell differentiation in tumors of group A and patterns related to β-cell differentiation in tumors of group B. We find high grade PanNET and PanNEC tumors, in addition to maintaining a β-cell signature, to be further distinguished by global hypermethylation of islet cell lineage markers.

MEN1/DAXX/ATRX mutations were an exclusive feature of group A subgroup

To link differential methylation profiles of PanNENs with gene mutations we designed a custom panel targeting 47 PanNEN-specific genes (Supplementary Table 5; details of panel design in material and methods). Additionally, a sample harboring no mutation in PanNEN panel genes was sequenced using the Comprehensive Cancer Panel (CCP). In total, 35 samples harbored alterations among the targeted genes (Supplementary Figure 3a). Among the samples which were analyzed at both the DNA and methylation level, we find 62% (n=22) of samples with DNA mutations belonged to either group A or B (Figure 3). The most frequently altered gene in our cohort was *MEN1* (34%) followed by *DAXX* (14%) and *ATRX* (7%) (Supplementary Figure 3a shows 59 samples investigated by DNA sequencing). Alteration in these tumor suppressor genes were overrepresented in group A while the majority of group B samples did not show known driver mutations associated with PanNENs (Figure 3). *TSC2* alterations were also present in three samples, of which two belonged to group A. The third case was a NEC, which explains its clustering to group B. In total, 24 samples contained no known driver mutations (Supplementary Figure 3a) (note all mutations discussed are non-synonymous, deletions or indels, for intronic and synonymous mutations; see original data). With the exception of five variants, all identified driver alterations had an allele frequency above 10% (Supplementary Figure 3b; list of all mutations and allele frequencies in Supplementary Table 6).

DNA copy number analysis identified three signatures in PanNENs

To explore copy number alterations (CNAs) in PanNENs we first inferred signal intensities from methylation log2 ratios. Next, in order to identify CNA signatures, we generated a

cumulative chromosomal aberration score (C-CAS), which is the sum of log2 ratios of the chromosomal intensities per patient (see method for details). Based on the C-CAS values, the samples in our cohort were divided into amplification-rich CNA ($C-CAS > 1$), deletion-rich CNA ($C-CAS < 0$) and low-CNA ($0 \leq C-CAS < 1$) groups (Figure 4a). Euclidean hierarchical clustering of the data with no prior knowledge of C-CAS values generated clustering of samples similar to C-CAS directed clustering (Supplementary Figure 4b). Using CNA profiles, we found NET G1 samples were enriched in the low-CNA signature group (Figure 4a). The amplification-rich CNA signature predominantly harbored recurring copy number gain in multiple chromosomes, where gain in chromosomes 5, 7 and 14 was found in 90% of samples, and gain in chromosomes 18, 4 and 9 was found in 72- 63% of the samples. The tumors within the amplification-rich CNA signature group predominantly belonged to group A. The deletion-rich signature group harbored losses in multiple chromosomes, with recurring loss seen mainly in chromosome 11, 21 and 16. Loss of chromosome 11 was present in 69% of the deletion-rich CNA group, followed by loss of chromosome 21 and 16, each occurring in 46% of the deletion-rich CNA samples. Tumor suppressor gene mutations combined with CNA loss in the corresponding chromosome was a predominant feature of the deletion-rich signature. This was evident in *MEN1, DAXX, TSC2, VHL* and *TP53* together with loss in chromosome 11, 6, 16, 3, and 17 respectively (Figure 4a, mutation status panel). Tumors in low-CNA and deletion-rich signatures harbored samples from both groups A and group B.

We performed fluorescent in situ hybridization (FISH) analysis on 18 samples to validate CNA results. We studied CNA gains using FISH probes for *RICTOR* and *TGFBR1* for chromosome 5 and 9, respectively. CNA loss in chromosome 11 was investigated with a *MEN1* FISH probe for chromosome 11 (Figure 4b). Signal counts per cell from FISH and log2 ratio from CNA analysis were correlated and show a regression coefficient of R^2 0.6531 and p=6.243 x 10^{7} (Figure 4c). Representative images are displayed in Figure 4d. Taken together, we have identified three distinct CNA signatures based on overall copy number changes. Amplification-rich CNA samples, with predominant gains were mainly in group A. Deletion-rich CNA samples with predominant chromosome losses and a tumor

suppressor inactivation phenotype were seen both in group A and B. Low-CNA samples with few copy number aberrations were found specifically in low grade samples belonging to both group A and B.

Proteomic analysis revealed differences in mTOR pathway component levels in the PanNEN groups.

To further characterize the PanNEN tumor groups identified by methylome analysis, we investigated cellular signaling in 14 tumor samples with matched normal tissues (nine tumors from group A and five from group B). We utilized DigiWest, a high-throughput western blot approach, to quantify proteins and to test differential abundance between α-like (group A) and β-like (group B) PanNEN tumors.

Initially, 86 analytes were tested with our cohort using the DigiWest analysis tool (Supplementary Table 7)[33]. We extracted 49 analytes which displayed less than 10% background (Supplementary Table 8 and see methods for details). Hierarchical clustering of fold change of these 49 analytes clustered group A and group B (Figure 5a). Differential expression analysis of group A and group B tumors identified 8 candidate proteins (padj < 0.05) (Figure 5b). Tumors in group A showed a significantly increased level of GSK3β and eIF4E, both members of the mTOR pathway, as compared to group B (>2.5 fold increase in group A). In addition, ILK1, an integrin-linked kinase upstream of GSK3ß and AKT, showed significant abundance in group A compared to group B. Total proteins levels of the Hippo pathway component RASSF2 as well as the apoptosis regulator MCL-1 were also increased in group A compared to group B. Thus, our proteomic analysis identified differential protein abundance between α-like (group A) and β-like (group B) PanNEN tumors.

Discussion

Our study presents novel subgroups within PanNENs and highlights that inter-patient heterogeneity is evident at multiple biological levels. Methylation-based identification of cell-of-origin has proven to be the method of choice for cancer subtyping in the past years[34,35]. Using genome-wide methylation profiling we identified two main cell-of-origin based tumor subtypes: α-like tumors, with a low inter-patient heterogeneity and beta (β)-like tumors displaying a stronger heterogeneity.

To gain insight into the islet cell type underlying PanNEN samples, information from single cell sequencing approaches was employed to identify highly expressed marker genes for individual islet cells[30–32,36]. Cross referencing the list with differentially methylated probes indicated gene body and promoter methylation of these marker genes (Figure 2). Single cell analysis of Islets of Langerhans has shown that expression of *PDX1* and *ARX/IRX2* is exclusive to β-cells and α-cells, respectively. *PDX1* drives differentiation of the early pancreatic epithelium and maintains the hormone-secreting phenotype of β-cells[37,38]. Our data showed hypomethylation within promoter and gene-body of *PDX1* in the β-like samples, while maintaining a hypermethylated state in α-like samples. In contrast, *IRX2*, a target gene of the α-cell maintenance transcription factor *ARX*[39], reveals hypomethylation within the α-like subgroup, but hypermethylation within the β-like samples. The differential methylation within the *IRX2* gene between the groups was predominant in gene-body as compared to the promoter probes of *IRX2*. Interestingly, differential gene body methylation was previously characterized as a mechanism of lineage-specific gene regulation for insulin, glucagon and a large number of additional genes in α- and β-cell of the pancreas[40]. It is currently unclear why DNA methylation within gene body enhancers is decisive for islet cell-specific gene expression. This effect was suggested to contribute to high plasticity in cellular reprogramming within pancreatic islets and potentially expands to lineage specific markers within the remaining islet cell types (δ- and PP-cells)[40]. Nevertheless, the methylation pattern of these master transcriptional regulators previously identified by single cell sequencing and genetic analysis defined islet cell type-specific subgroups of PanNENs within our cohort. Similar

subgrouping of PanNEN has been established at the histone modification level, wherein, one subgroup of samples showed enrichment of the "super" enhancer mark H3K27ac at the *ARX* and *IRX2* locus, while the second subgroup showed enrichment in the *PDX1* locus[28].

DNA methylation patterns characterizing α- or β-cell lineage of our PanNEN samples were rather consistent for *PDX1* and *IRX2*. Genes encoding other cell type markers, such as the homeobox factors *MEIS1* and *NKX2-2*, as well as the paired box transcription factor *PAX6*, displayed a differential methylation profile within the β-like samples. All three of them exhibited hypomethylation in β-like NET G1, in contrast to the high-grade tumors in this group, which showed a hypermethylated feature. The distinct methylation pattern within the aforementioned genes indicates an epigenetic branching of low and high-grade PanNENs with a β-cell history. Apart from this, other cell markers critical for islet cell maintenance were also seen hypermethylated in NET G3 and NEC and further expanded to the β-like NET G2 and NET G1 at varying rates. This could suggest that NET G1/G2 samples, albeit currently of a low-grade phenotype, are amassing features of aggressive tumors at the methylation level.

Beyond the lineage-specific gene methylation patterns, Reactome pathway analysis uncovered differential methylation profiles within a small group of growth factor receptor and ligand genes between these groups. Among those, *VEGFA* and *cKIT* genes exhibited strong gene-body hypermethylation in β-like tumors, contrasted by hypomethylation in the α-like group. Regulation of the *VEGFA* gene via DNA methylation in the promoter region has been described and was correlated with invasive properties of transition cell carcinomas[41]. To what extent the differential gene body methylation detected in our samples correlates with differential expression of VEGFA and other genes needs to be determined.

In addition to the two main islet-cell specific epigenetic profiles of PanNEN, these groups also differ with respect to tumor suppressor gene mutations, copy number alterations (CNV) and oncogenic pathway activation. Mutational signatures detected by targeted sequencing were similar to previously reported alterations in PanNENs[16,42,43], and identified two subgroups within α-like samples: one group is characterized by *MEN1,* and

DAXX or *ATRX* mutations, the second group is devoid of any PanNEN-specific mutations detected by this approach. In contrast to these two α-like mutation signatures, β-like samples had few and variable mutations and *MEN1/DAXX/ATRX* alterations were completely absent (Figure 3).

A strong indicator of different tumorigenic mechanisms both within α-like PanNEN subgroups and between α- and β-like PanNENs became apparent within the CNV signatures. An amplification-rich signature targeting several chromosomes together with *MEN1/DAXX/ATRX* mutations defines one α-like subgroup. The signature is characterized by co-occurring gains in chromosomes 5, 7, 14, 18 and 19 identified as synchronous events with a single peak in amplification[44]. This signature is identified as a late event in PanNENs following mutations in the core tumor suppressor genes associated with PanNEN[44]. Thus, it is likely that the mutations in *MEN1/DAXX/ATRX* preceded whole chromosome amplifications in the amplification-rich α-like subgroup. Within the deletion-rich CNV group of PanNENs, we found both α- and β-like samples. These tumors are characterized by a deletion-rich signature showing loss of chromosomes 6, 11, 16 and 22 combined with mutations in *MEN1, DAXX, VHL* and *TSC2* tumor suppressor genes. The key features of deletion-rich CNA signature are early events during PanNEN evolution[44]. Consequently, the deletion-rich α-and β-like PanNENs are likely to separate early. Chromosomal losses harboring tumor suppressor genes including *MEN1* (Chr11), *TSC2* (Chr 16), *VHL* (Chr 3) and *DAXX* (Chr 6) may therefore act as early driver events in these tumors. While deletion-rich and amplification-rich subgroups harbor the majority of NET G2 samples, 8 out of the 14 NET G1 samples are characterized by the absence of large CNV alterations. Considering the frequent benign nature of NET G1, its association with small tumor size and very few alterations, a low-CNA signature is plausible. The novel distribution of these genetic alterations within the specific α- and β-like PanNEN subgroups illustrate different mechanism of tumorigenesis likely acting within α- and β-like PanNENs.

To take molecular characterization of the PanNEN subgroups one step further, we looked at the protein levels of major players in key biological processes known to play a role in tumor development. This analysis revealed increased levels of mTOR signaling

components including GSK3β, eIF4E, in α-like PanNENs. Upon clustering the samples, additional components of the mTOR pathway including AKT, p70S6K and NPM1, involved in ribosome biogenesis, appeared higher expressed in α-like samples. However, due the heterogeneous expression and the low samples size they did not appear significant between the two groups. This observation, nevertheless, indicates a distinct mTOR phenotype in α-like PanNENs, which is largely independent of mutations. MTOR pathway activation can occur in PanNENs via deletions and mutations in PTEN and TSC2 genes, respectively, yet these are rather low frequency events[12,16,42,43]. Based on these observations, the mTOR inhibitor Everolimus has been approved for well-differentiated PanNENs and has shown efficacy in several clinical trials[45]. Despite positive responses in terms of mainly stable disease, the fraction of patients with a long-term response is low. Increased activation of the mTOR pathway has been suggested by several groups as a valuable biomarker for Everolimus response in pancreatic and other gastroenteropatic NETs, and was suggested as an indicator of shorter disease-free survival[11,46,47]. At the same time, patients with increased levels of phospho-mTOR/S6K are likely to show a better response towards Everolimus compared to patients without mTOR pathway activity. Our data indicate that stratification of patient groups in terms of α- and β-like PanNEN subgroups is likely to provide results that are more concise in regard to pathway activation and treatment response.

Taken together our multi-level approach uncovered at least two main subgroups within PanNENs (Figure 6). Distinct tumor progression trajectories within α-like subgroups were apparent at the genetic level, where samples carrying low-CNA or deletion-rich CNA phenotype are within the early stages of tumor development, while a subset of α-like tumors harboring an amplification-rich CNA phenotype are in a much later stage. The development trajectories of β-like tumors were evident at the epigenetic level. The aggressive G3 tumors are characterized by global hypermethylation of genes playing a fundamental role in islet cell differentiation. Less aggressive tumors of the β-like subgroup, although likely harboring the same cell-of-origin, maintain a hypomethylated state within these differentiation-associated genes. Based on our protein analysis, we have gained first indications that a differential mTOR signature, which we could not

deduce from epigenomic and genomic analyses, characterizes α-like PanNENs. This clearly needs to be confirmed by more detailed investigations. In the future, this could be a starting point for an improved stratification of PanNEN patients prior to pathway-directed therapy.

Materials and Methods

Patient Cohort and experimental representation

PanNEN samples were collected from 58 patients. For one patient, both primary and metastasis was obtained. The sample cohort was provided by the Institute of Pathology at Charité Universitätsmedizin, Berlin and the Medical University of Graz. Four lymph node metastases of PanNEN origin used for this project have been provided by Biobank Graz, Austria. The cohort carried well-differentiated tumors of grades G1, G2 and G3 grades as well as poorly differentiated G3 tumors. Both genders were equally represented within the cohort (table 1). Primary tumors consisted of 76% of the cohort, followed by liver metastasis and lymph node metastasis, at 14% and 10%, respectively. The primary tumors were resected 36% from the head, 31% from the tail and 7% from the body of the pancreas. Two samples had lesions in multiple sections of the pancreas. Clinical report on the tumor was collected and is presented in Supplementary table 1. All samples were collected as formaldehyde fixed paraffin embedded (FFPE) blocks, and normal control for the respective patients were collected for all patients with the exception of 5 cases (Figure 1a). Normal tissue sections were obtained as either tissue adjacent to the tumor (as per the pathologist's examination) (normal adjacent n= 24), or as a completely separate block harboring only normal tissue (normal distant n= 27). Normal blood samples were available in 4 cases.

All patients provided a signed consent as part of the clinical documentation protocol of the Charité Universitätsmedizin, Berlin and the Medical University of Graz.

Table 1. Cohort characteristics

feature	variables	n (%)
Gender	female	27 (46)
	male	27 (46)
	unknown	5 (8)
Tumor type	liver metastasis	8 (14)
	lymph node	6 (10)
	primary	45 (76)
Grade	NET G1	27 (46)
	NET G2	24(41)
	NET G3	4(7)
	NEC	3(5)
	unknown	1(2)
location	Head	21 (36)
	Body	4 (7)
	tail	18 (31)
	Multiple pancreatic sites	2 (3)
	liver	7 (12)
	lymph node	6 (10)
	unknown	1 (2)

PanNEN panel design

We designed a custom panel targeting PanNEN relevant genes, covering all mutations associated with PanNENs. To make the panel, we first performed a text mining approach to extract high quality information regarding genes associated with PanNENs from GeneView[48], the Catalogue of Somatic Mutations in Cancer (COSMIC)[49] and mutations collected from PanNEN publications[12,50–52]. From this list, 47 PanNEN likely driver genes were then extracted with further focus on MAPK and mTOR pathways. Amplicon design was then performed via Ion AmpliSeq Designer tool (Life Technologies) under the criteria "DNA Gene design (multi-pool)" with GRCh37 genome. The panel was designed to generate primers targeting 125-bp stretches of exon regions of the selected genes. The complete panel included 1175 amplicons, divided into four pools.

DNA isolation

DNA from all samples was isolated from tissue that had been formalin-fixed and embedded in paraffin (FFPE). Tissue samples were sectioned and stained with Hematoxylin and Eosin (H&E). Pathologists demarcated tumor and healthy tissue areas in the H&E slides and depending on the size of the marked area, 12 sections of 5μm each were used for DNA isolation from tumor samples and 6 section of 5μm for control normal tissue. The tissue was macro-dissected from the slides and DNA was prepared using the GeneRead DNA FFPE kit (Qiagen, Netherlands). Quality and quantity of DNA was determined by RNAse P quantification (Thermo Fisher Scientific, USA). Microsatellite instability was identified using a mononucleotide marker panel (MSI Analysis System, Promega, Germany).

DNA Sequencing

We used 20ng of DNA for library preparation using Ion Ampliseq Library kit (Thermo Fisher Scientific). Regions were targeted by primers distributed into 2 amplicon pools per DNA sample. Upon ligation to Ion Xpress Barcode Adapters (Thermo Fisher Scientific) and purification using Agencourt AMPure beads (Beckman Coulter), two samples were mixed at equal ratio on 318v2 sequencing chip. Using the Ion Torrent PGM (Thermo

Fisher Scientific), the samples were sequenced at an average read depth of 1158 reads in order to generate the raw intensity data.

Sanger Sequencing

To validate results from targeted massive parallel sequencing we performed Sanger sequencing on specific mutations which we found required validation, for example, sub-optimal amplicon performance or location within long nucleotide repeat area. (Supplementary Table 6).

Fluorescence in-situ hybridization (FISH)

Fluorescence in situ hybridization (FISH) was performed on 3µm tumor sections from 23 samples. We used commercially available, standardized probes for detecting chromosome 5 with a *RICTOR* gene probe, chromosome 9 with *TGFBR1* gene, and chromosome 11 with *MEN1* (Empire Genomics, USA). Hybridization was performed according to manufacturer's instructions. Where possible, we scored 40 cells per sample using an Olympus microscope. Analysis was conducted using 'BioView solo' (Abbott Molecular).

DNA methylation

DNA methylation profiling of all samples was performed with 200–500 ng of DNA using the Infinium MethylationEPIC BeadChip array (850 k; Illumina, Inc., San Diego, CA, USA) according to the protocols provided by the manufacturer.

High throughput proteomics

The DigiWest technology is a bead-based microarray platform that combines gel-based protein separation followed by immobilization on microspheres in order to semi-quantitatively analyze via Luminex technology hundreds of specific proteins and protein modifications in a given sample. Tumor sections of selected patients were marked by a pathologist for tumor and normal tissue and 6x of 5 μm section were macro dissected for

tumor and normal tissue of each patient. To extract proteins, the Qproteome FFPE tissue kit (Qiagen, Venlo, NLD) was used. Deparaffinization was performed using Heptane and incubation took place for 1h at 23°C. Protein extraction was performed according to manufacturer's protocol. Protein precipitation was performed using the 2D Clean-Up kit (GE Healthcare, Chicago, IL, USA) and proteins were resuspended in 20 µl of loading buffer (Life Technologies, Carlsbad, USA). DigiWest was performed as described previously[33]. Briefly, the NuPAGE system (Life Technologies, Carlsbad, USA) with a 4-12% Bis-Tris gel was used for gel electrophoresis and Western blotting onto PVDF membranes. After washing with PBST, proteins were biotinylated by adding 50 µM NHS-PEG12-Biotin in PBST for 1 h to the Membrane. After washing in PBST Membranes were dried overnight. Each Western-Blot lane was cut into 96 stripes of 0.5 mm each. Strips of one Western blot lane were sorted into a 96-Well plate (Greiner Bio-One, Frickenhausen, GER) according to their molecular weight. Protein elution was performed using 10 µl of elution buffer (8 M urea, 1% Triton-X100 in 100 mM Tris-HCl pH 9.5). Neutravidin coated MagPlex beads (Luminex, Austin, TX, USA) of a distinct color ID were added to the proteins of a distinct molecular weight fraction and coupling was performed overnight. Leftover binding sites were blocked by adding 500 µM deactivated NHS-PEG12-Biotin for 1 h. To reconstruct the original Western blot lane, the beads were pooled, at which the color IDs represent the molecular weight fraction of the proteins.

For antibody incubation 5 µl of the DigiWest Bead mixes were added to 50 µl assay buffer (Blocking Reagent for ELISA (Roche, Rotkreuz, Switzerland) supplemented with 0.2% milk powder, 0.05% Tween-20 and 0.02% sodium azide) in a 96 Well plate. Assay buffer was discarded and 30 µl of primary antibody diluted in assay buffer was added per Well. Primary antibodies were incubated overnight at 15 °C on a shaker. Subsequently, primary antibodies were discarded, and beads were washed twice with PBST. After washing, 30 µl of species specific-secondary antibody diluted in assay buffer labeled with Phycoerithrin was added and incubation took place for 1 h at 23°C. Before the readout on a Luminex FlexMAP 3D beads were washed twice with PBST.

Data processing and analysis

Mutational analysis

The raw reads were first aligned to GRCh37 sequence using the Torrent Mapping and Alignment Program (TMAP; Life Technologies), placing a cut off of > 50 nucleotides in aligned reads and a mapping quality of > 4 using an in-house python script. The processed bam files were then utilized for variant calling using TS Variant Caller plugin under the "strict" setting per Ion Suite (Ion Torrent platform) parameter profiles. The generated variant call format (VCF) files of the tumor-normal pair per patient was merged and reference and alteration read count for all variants within the two VCF files were extracted to determine the representation of the variants in both cases. The merged VCF file was then annotated using the annotation framework SoFIA[53]. Downstream filtering included removing variants that were positive for the following set of parameters: intronic and synonymous variants, 1000Genome variants with frequency greater than 1% in the population, variants within homopolymer regions > 4 nucleotides and finally, variants that were represented at comparable allelic frequencies in the matched normal tissue.

DNA methylation preprocessing

Raw idat files were preprocessed using subset-quantile within array normalization (SWAN) provided through the minif package[54 55]. Probes performing poorly in the analysis were further filtered out. The failed probes were identified if their detection p value was > 0.01 in at least one sample. We also filtered out probes showing cross reactivity to multiple sites in the genome[56]. We removed sex chromosome probes and probes containing SNPs with allele frequency ≥ 0.01. Both features were identified via minfi package.

Unsupervised class discovery

Probe clusters were determined using the boundedClusterMaker function of the R package “minfi”. A maximum cluster width of 1500bp and a maximum gap of 500bp was applied. Using the 8000 most variable probe-clusters identified by determining row (probe) variance (σ^2), DNA methylation-based classes of PanNEN were identified with ConsensusClusterPlus package under the following parameters: maxK = 12, reps=1000, pItem=0.8 and pFeature=1. The function performed agglomerative hierarchical clustering after performing 1-Pearson correlation distance and consensus matrix carrying pairwise consensus values is generated for 12 clusters. Most stable number of clusters was determined based on the cumulative distribution score curve (CDF) that reached approximate maximum (k=3). The third cluster identified was formed from one sample and was removed from further analysis. Hierarchical clustering of 8000 variable probe clusters was done by first obtaining dissimilarity matrix using Euclidean algorithm and then performing the clustering using complete linkage. PCA for the data was also performed. Both analyses were done using R package “Stats”.

Differentially methylated probes (DMP) and associated analysis

Upon extracting and assigning the samples to the identified stable clusters group A and group B, differentially methylated probes (DMP) were identified out of all the CpG sites (upon the aforementioned preprocessing) using CHAMP package function champ.DMP under the following parameters: adjPVal = 0.05, and adjust.method = "BH". Principal component analysis (PCA) was performed to determine the variance between group A and group B samples. Hierarchical clustering using the beta values of DMP was done by first obtaining dissimilarity matrix using Euclidean algorithm and then performing the clustering using complete linkage. Both analyses were done using R package “Stats”. Pearson correlation of the DMP beta values was done using Hmisc package.

Genes associated with DMPs were evaluated for Reactome pathway[57] term enrichment using ReactomePA, clusterprofiler R packages. The analysis was done under the following parameters: pAdjustMethod = “BH” (Benjamini and Hochberg), pvalueCutoff =

0.05, qvalueCutoff = 0.2, minGSSize = 10, and maxGGSize = 500. All genes represented in the Illumina EPIC array was used as background. Final set of terms was curated by filtering only those that showed an adjusted p-value less than 0.05 and fold enrichment of greater than 1.5. Probes within genes associated with top Reactome terms were further evaluated. Highly differentially methylated gene set in the top enriched Reactome pathway term was determined accordingly. First, the absolute value differences between group A and group B per probe was determined. Mean of absolute difference in beta value across all probe per gene was calculated. Highly differentially methylated genes were then defined by the mean | delta beta value | greater than 0.20.

Islet cell marker were curated from recently published four single cell sequencing publications of islet cell[30–32,36] as well as from Gene Ontology database (http://geneontology.org). Gene Ontology terms used included “pancreatic A cell differentiation”,” “type B cell differentiation”, “pancreatic D cell differentiation”, and “pancreatic PP cell differentiation”. DMP associated genes that overlapped with curated Islet cell markers were filtered. Final lists of DMP associated Islet cell markers were identified if they met the follow criteria: marker mentioned more than once within the five references for a particular cell/cells and the mean abs(Δβ) the gene is greater than 0.15. To determine closely related samples within each group, hierarchical clustering with complete linkage was performed using beta values of the identified Islet cell markers associated with DMPs before visualization.

Copy Number Aberrations (CNA)

CNA was identified from EPIC array data using conumee[58]. Upon raw preprocessing, mean of all segments of autosomal regions were obtained, and a mean value per autosome was calculated for each sample to determine the log2 ratio of intensities across the chromosome. A cut-off of $x > 0.15$ and $x < 0.15$ was placed to limit the number of false positive obtained upon comparing log2 ratio values to FISH count (Supplementary Figure 4a). To determine the subgroups within the cohort, we generated a metrics termed cumulative chromosomal aberration score, which calculated the sum of intensity across all autosome per sample. If a sample did not harbor any detectable aberration, a score of

0 was given for this patient. $x > 1$ represented amplification-rich CNA samples, $x < 0$ were deletion-rich CNA samples and $0 \leq x < 1$ represents low-CNA subgroup. Euclidean hierarchical clustering was also performed on the data.

Proteome data analysis

The DigiWest analysis tool described by Treindl *et. al.*[33], was used to identify antigen-specific peaks. Average fluorescence intensity (AFI) values were calculated by integration of peak areas. For comparative analysis, AFI values were normalized to the total protein per Western blot lane. In order to determine outlier samples, we evaluated the distribution of signal intensities across all samples. If the median of a sample was close to the population standard deviation, then the sample was eliminated. This resulted in elimination of one sample pair (PNET14) from further analysis In order to eliminate poorly performing analytes, we looked at the noise level of the analytes across the sample set. The value 33 was the threshold signal at which background noise was established. After identifying the percentage of samples carrying ≤ background noise value per analyte, a 10% cut off was placed to eliminate false positive analyte data. This resulted in 49 analytes with less than 10% or no background signal across all samples (both tumor and normal), which were utilized for downstream analysis. Tumor specific protein quantity was identified upon normalizing to the respective normal protein quantity per patient. Differentially expressed proteins between the groups were identified using student t-test and was corrected using Benjamini and Hochberg (BH) test in R. The identified candidates had an adjusted p-value < 0.05.

Data visualization and statistics

All data visualization was performed in R (version 3.6.3) using base R plotting function, ggplot2 package or ComplexHeatmap package. The appropriate statistics mentioned above were all performed using respective R packages or base R functions.

Figure legends

Fig. 1 PanNEN methylome profiles identify two distinct tumor subgroups. a. Summary of tumor characteristics and molecular data analyzed; DNA sequencing, Methylation, DNA copy number, and Proteomics. **b.** Consensus matrix using 8000 most variable probe clusters. Scale indicates the probability of cluster assignment from 0 (blue) to 1 (red), PanNEN group A (blue) and group B (red) and group C (black). **c.** Left: Methylation β value heatmap using 8000 most variable probe clusters. Right: Principal component analysis based on 8000 probe clusters separated group A and group B. **d.** Reactome Pathway analysis of DMP associated genes (fold enrichment >1.5, padj <0.05. P value corrected using Benjamini and Hochberg). **e.** Beta value heatmap of TFAP2 family regulated transcription of growth factors and their receptors (top enriched Reactome pathway) grouped by PanNEN groups A and B. Displayed genes have mean abs(Δβ) ≥ 0.2. Probe features of DMPs are shown on the right. Tumor grades are annotated in the lower panel.

Fig. 2 Differentially methylated islet cell markers distinguish group A and group B carrying α- and β-cell related features. a. Scheme of islet cell marker identification in the DMPs. Gene ontology and four publications were utilized to select 63 markers alpha (α), beta (β), delta (δ) and pancreas polypeptide cells (PP) endocrine cells. From these, 44 markers were found within the genes associated with the DMPs of our study. Markers found in more than one publication and with mean abs(Δβ) ≥ 0.15 were further assessed. **b.** Distribution of probe beta values of the 15 islet cell markers identified in (a) between group A and group B. **c.** Beta value heatmap of the 15 islet cell marker probes in samples of group A and B. Islet cell types associated with the cell markers are graphically represented by circles: green = alpha, orange = beta, blue = PP, brown = delta islet cells. Probe features of DMPs are shown on the right. Tumor characteristics are annotated in the lower panel.

Fig. 3 DNA mutations in *MEN1*, *ATRX* and *DAXX* PanNEN driver genes found in group A tumors. Mutations identified by panel sequencing within each individual sample from group A and B. Heatmap colors represent mutation type. Tumor characteristics and panel utilized for sequencing are annotated in the lower panel. Bar chart (top) displays mutation count per sample, (right) displays mutation count per gene. Only non-synonymous mutations are displayed.

Fig. 4 DNA Copy Number Alteration (CNA) profiles separate PanNEN samples into three subsets. a. Representation of copy number alterations from samples of group A

and group B. Top panel: dot plot of CCAS (cumulative chromosomal aberration score) values: the sum of log2 ratios intensity across all autosome per sample. Heatmap of mean inferred signal intensity log2 ratios (tumor/normal) per autosome across all samples from group A and group B. Copy number gain is displayed in red and copy number loss in blue. Bullets indicate mutations found in gene of the respective autosome. Tumor characteristics, driver mutation status and fluorescent *in situ* hybridization (FISH) validation status annotated in lower panels. **b.** Quantification of FISH for chromosomes 5, 9 and 11 using probes in *RICTOR*, *TGFBR*1 and *MEN1* respectively. Forty cells randomly selected were counted in each sample. The distribution of signals per sample is depicted. Mean and standard deviation represented by error bars. **c.** Linear regression analysis of EPIC log2 ratio and mean copy number from FISH analysis. Regression coefficient R^2 = 0.6531 and p = 6.243 $x10^{-07}$. d. Representative FISH for *RICTOR*, *TGFBR1* and *MEN1* displayed in orange and for centromeres of chromosomes 5, 9 and 11 in green. Scale bar depicts 10 mm.

Fig. 5 Protein level analysis indicates mTOR signaling in group A tumors. a. Heatmap representing log2 ratio of average fluorescent intensities (tumor / normal) of proteins from group A and group B tumors. Euclidean hierarchical clustering using complete linkage displayed. **b.** Eight proteins with significantly different levels of log2 fold change between group A and group B (from 12 samples) padj < 0.05. p value was BH corrected.

Fig. 6 Schematic representation of PanNEN tumor model and genetic / epigenetic characteristics. Group A tumors show methylation features of alpha-cell origin with distinct genetic and protein characteristics, while group B displays methylation features of beta-cell origin. High grade PanNET and PanNECs of beta-cell origin, displayed global hypermethylation features various islet cell markers.

Supplementary Fig. 1 a. Schematic defining probe clusters. Probes in close proximity are combined to generate mean Beta value per probe cluster. **b.** Outline of methodology followed for group discovery and differential methylation investigation. **c.** Cumulative distribution function (CDF) curve of the resulting cluster counts. Cluster count of 'three' resulted in the most stable number subgroups. **d.** Hierarchical clustering using differentially methylated probes (DMPs) clustered samples into two main clusters: group A (blue) and group B (red). **e.** Principal component analysis based on DMPs of samples **f.** Pie chart of localization of differentially methylated probes (DMPs) from Figure 1e. **g.** Heatmap of Pearson correlation coefficient of group A and group B samples. **h.** Distribution of mean | delta Beta valuel of DMP associated genes of TFAP2 family

regulated transcription of growth factors and their receptors. Dotted red line indicates cut-off used to select genes.

Supplementary Fig. 2 a. Pie chart of localization of differentially methylated probes (DMPs) within the identified 44 islet cell makers from Figure 2a. **b.** Distribution of mean Idelta Beta valuel of 44 islet cell markers from Figure 2a. Dotted red line indicates cut off used to select genes.

Supplementary Fig. 3 a. Mutations identified by PanNEN panel sequencing within each individual sample from the entire cohort (59 samples). Heatmap colors represent mutation type. Tumor characteristics are annotated in the lower panel. Bar chart (top) displays mutation count per sample, (right) displays mutation count per gene. b. Allele frequency heat map of all identified mutations from PanNEN panel sequencing (samples without mutations are not displayed).

Supplementary Fig. 4 a. Graphical representation of mean FISH count and log2 ratios from EPIC. Dotted red line indicates cut off placed to discard false positives. **b.** Euclidean hierarchical clustering using complete linkage of copy number alteration data. Copy number gain is displayed in red and loss in blue.

References

1. Yadav S, Sharma P, Zakalik D. Comparison of Demographics, Tumor Characteristics, and Survival between Pancreatic Adenocarcinomas and Pancreatic Neuroendocrine Tumors: A Population-based Study. *Am J Clin Oncol Cancer Clin Trials*. 2018;41(5):485-491. doi:10.1097/COC.0000000000000305
2. Halfdanarson TR, Rabe KG, Rubin J, Petersen GM. Pancreatic neuroendocrine tumors (PNETs): Incidence, prognosis and recent trend toward improved survival. *Ann Oncol*. 2008;19(10):1727-1733. doi:10.1093/annonc/mdn351
3. Inzani F, Petrone G, Rindi G. The New World Health Organization Classification for Pancreatic Neuroendocrine Neoplasia. *Endocrinol Metab Clin North Am*. 2018;47(3):463-470. doi:10.1016/j.ecl.2018.04.008
4. Hijioka S, Hosoda W, Matsuo K, et al. Rb loss and KRAS mutation are predictors of the response to platinum-based chemotherapy in pancreatic neuroendocrine neoplasm with grade 3: A Japanese multicenter pancreatic NEN-G3 study. *Clin Cancer Res*. 2017;23(16):4625-4632. doi:10.1158/1078-0432.CCR-16-3135
5. Okuyama H, Ikeda M, Okusaka T, et al. A phase II trial of everolimus in patients with advanced pancreatic neuroendocrine carcinoma refractory or intolerant to platinum-containing chemotherapy (NECTOR trial). *Neuroendocrinology*. January 2020. doi:10.1159/000505550
6. Fjällskog MLH, Lejonklou MH, Öberg KE, Eriksson BK, Janson ET. Expression of molecular targets for tyrosine kinase receptor antagonists in malignant endocrine pancreatic tumors. *Clin Cancer Res*. 2003;9(4):1469-1473.
7. Casanovas O, Hicklin DJ, Bergers G, Hanahan D. Drug resistance by evasion of antiangiogenic targeting of VEGF signaling in late-stage pancreatic islet tumors. *Cancer Cell*. 2005;8(4):299-309. doi:10.1016/j.ccr.2005.09.005
8. Hansel DE, Rahman A, Hermans J, et al. Liver metastases arising from well-differentiated pancreatic endocrine neoplasms demonstrate increased VEGF-C expression. *Mod Pathol*. 2003;16(7):652-659. doi:10.1097/01.MP.0000077416.68489.50
9. Metz DC, Jensen RT. Gastrointestinal Neuroendocrine Tumors: Pancreatic Endocrine Tumors. *Gastroenterology*. 2008;135(5):1469-1492. doi:10.1053/j.gastro.2008.05.047
10. Kulke MH, Bendell J, Kvols L, Picus J, Pommier R, Yao J. Evolving diagnostic and treatment strategies for pancreatic neuroendocrine tumors. *J Hematol Oncol*. 2011;4:2-9. doi:10.1186/1756-8722-4-29
11. Missiaglia E, Dalai I, Barbi S, et al. Pancreatic endocrine tumors: Expression profiling evidences a role for AKT-mTOR pathway. *J Clin Oncol*. 2010;28(2):245-255. doi:10.1200/JCO.2008.21.5988
12. Jiao Y, Shi C, Edil BH, et al. DAXX/ATRX, MEN1, and mTOR Pathway Genes Are Frequently Altered in Pancreatic Neuroendocrine Tumors. *Science (80-)*. 2011;331(March):1199-1204. https://doi.org/10.1126/science.1200609.
13. Yao JC, Phan AT, Chang DZ, et al. Efficacy of RAD001 (everolimus) and

octreotide LAR in advanced low- to intermediate-grade neuroendocrine tumors: Results of a phase II study. *J Clin Oncol*. 2008;26(26):4311-4318. doi:10.1200/JCO.2008.16.7858

14. Yao JC, Shah MH, Ito T, et al. Everolimus for advanced pancreatic neuroendocrine tumors. *N Engl J Med.* 2011;364(6):514-523. doi:10.1056/NEJMoa1009290
15. Yao JC, Lombard-Bohas C, Baudin E, et al. Daily oral everolimus activity in patients with metastatic pancreatic neuroendocrine tumors after failure of cytotoxic chemotherapy: A phase II trial. *J Clin Oncol*. 2010;28(1):69-76. doi:10.1200/JCO.2009.24.2669
16. Scarpa A, Chang DK, Nones K, et al. Whole-genome landscape of pancreatic neuroendocrine tumours. *Nature*. 2017;543(7643):65-71. doi:10.1038/nature21063
17. Corbo V, Dalai I, Scardoni M, et al. MEN1 in pancreatic endocrine tumors: Analysis of gene and protein status in 169 sporadic neoplasms reveals alterations in the vast majority of cases. *Endocr Relat Cancer*. 2010;17(3):771-783. doi:10.1677/ERC-10-0028
18. Karnik SK, Hughes CM, Gu X, et al. Menin regulates pancreatic islet growth by promoting histone methylation and expression of genes encoding p27Kip1 and p18INK4c. *Proc Natl Acad Sci U S A*. 2005;102(41):14659-14664. doi:10.1073/pnas.0503484102
19. Razmara M, Monazzam A, Skogseid B. Reduced menin expression impairs rapamycin effects as evidenced by an increase in mTORC2 signaling and cell migration 06 Biological Sciences 0601 Biochemistry and Cell Biology. *Cell Commun Signal*. 2018;16(1):1-12. doi:10.1186/s12964-018-0278-2
20. Jin S, Mao H, Schnepp RW, et al. Menin associates with FANCD2, a protein involved in repair of DNA damage. *Cancer Res*. 2003;63(14):4204-4210.
21. Goldberg AD, Banaszynski LA, Noh KM, et al. Distinct Factors Control Histone Variant H3.3 Localization at Specific Genomic Regions. *Cell*. 2010;140(5):678-691. doi:10.1016/j.cell.2010.01.003
22. Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. *Genes Dev*. 2010;24(12):1253-1265. doi:10.1101/gad.566910
23. Yachida S, Vakiani E, White CM, et al. Small cell and large cell neuroendocrine carcinomas of the pancreas are genetically similar and distinct from well-differentiated pancreatic neuroendocrine tumors. *Am J Surg Pathol*. 2012;36(2):173-184. doi:10.1097/PAS.0b013e3182417d36
24. Lawrence B, Blenkiron C, Parker K, et al. Recurrent loss of heterozygosity correlates with clinical outcome in pancreatic neuroendocrine cancer. *npj Genomic Med*. 2018;3(1). doi:10.1038/s41525-018-0058-3
25. Pea A, Yu J, Marchionni L, et al. Genetic Analysis of Small Well-differentiated Pancreatic Neuroendocrine Tumors Identifies Subgroups With Differing Risks of Liver Metastases. *Ann Surg*. 2018;XX(Xx):1. doi:10.1097/sla.0000000000003022

26. Hong X, Qiao S, Li F, et al. Whole-genome sequencing reveals distinct genetic bases for insulinomas and non-functional pancreatic neuroendocrine tumours: leading to a new classification system. *Gut*. 2019:gutjnl-2018-317233. doi:10.1136/gutjnl-2018-317233
27. Chan CS, Laddha S V., Lewis PW, et al. ATRX, DAXX or MEN1 mutant pancreatic neuroendocrine tumors are a distinct alpha-cell signature subgroup. *Nat Commun*. 2018;9(1):1-10. doi:10.1038/s41467-018-06498-2
28. Cejas P, Drier Y, Dreijerink KMA, et al. Enhancer signatures stratify and predict outcomes of non-functional pancreatic neuroendocrine tumors. *Nat Med*. 2019. doi:10.1038/s41591-019-0493-4
29. Bormann F, Rodríguez-Paredes M, Lasitschka F, et al. Cell-of-Origin DNA Methylation Signatures Are Maintained during Colorectal Carcinogenesis. *Cell Rep*. 2018;23(11):3407-3418. doi:10.1016/j.celrep.2018.05.045
30. Muraro MJ, Dharmadhikari G, Grün D, et al. A Single-Cell Transcriptome Atlas of the Human Pancreas. *Cell Syst*. 2016;3(4):385-394.e3. doi:10.1016/j.cels.2016.09.002
31. Segerstolpe Å, Palasantza A, Eliasson P, et al. Single-Cell Transcriptome Profiling of Human Pancreatic Islets in Health and Type 2 Diabetes. *Cell Metab*. 2016;24(4):593-607. doi:10.1016/j.cmet.2016.08.020
32. Wang YJ, Schug J, Won KJ, et al. Single-cell transcriptomics of the human endocrine pancreas. *Diabetes*. 2016;65(10):3028-3038. doi:10.2337/db16-0405
33. Treindl F, Ruprecht B, Beiter Y, et al. A bead-based western for high-throughput cellular signal transduction analyses. *Nat Commun*. 2016;7. doi:10.1038/ncomms12852
34. Jurmeister P, Schöler A, Arnold A, et al. DNA methylation profiling reliably distinguishes pulmonary enteric adenocarcinoma from metastatic colorectal cancer. *Mod Pathol*. 2019;32(6):855-865. doi:10.1038/s41379-019-0207-y
35. Capper D, Jones DTW, Sill M, et al. DNA methylation-based classification of central nervous system tumours. *Nature*. 2018;555(7697):469-474. doi:10.1038/nature26000
36. Li J, Klughammer J, Farlik M, et al. Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep*. 2016;17(2):178-187. doi:10.15252/embr.201540946
37. Holland AM, Hale MA, Kagami H, Hammer RE, MacDonald RJ. Experimental control of pancreatic development and maintenance. *Proc Natl Acad Sci U S A*. 2002;99(19):12236-12241. doi:10.1073/pnas.192255099
38. Zhu Y, Liu Q, Zhou Z, Ikeda Y. PDX1, Neurogenin-3, and MAFA: Critical transcription regulators for beta cell development and regeneration. *Stem Cell Res Ther*. 2017;8(1):1-7. doi:10.1186/s13287-017-0694-z
39. Collombat P, Mansouri A, Hecksher-Sørensen J, et al. Opposing actions of Arx and Pax4 in endocrine pancreas development. *Genes Dev*. 2003;17(20):2591-2603. doi:10.1101/gad.269003
40. Neiman D, Moss J, Hecht M, et al. Islet cells share promoter hypomethylation independently of expression, but exhibit cell-type-specific methylation in

enhancers. *Proc Natl Acad Sci U S A*. 2017;114(51):13525-13530. doi:10.1073/pnas.1713736114

41. Ping SY, Shen KH, Yu DS. Epigenetic regulation of vascular endothelial growth factor a dynamic expression in transitional cell carcinoma. *Mol Carcinog*. 2013;52(7):568-579. doi:10.1002/mc.21892
42. Di Domenico A, Wiedmer T, Marinoni I, Perren A. Genetic and epigenetic drivers of neuroendocrine tumours (NET). *Endocr Relat Cancer*. 2017;24(9):R315-R334. doi:10.1530/ERC-17-0012
43. Pavel ME, Sers C. Systemic therapies in neuroendocrine tumors and novel approaches toward personalized medicine. *Endocr Relat Cancer*. 2016;23(11):T135-T154. doi:10.1530/ERC-16-0370
44. Gerstung M, Jolly C, Leshchiner I, Dentro SC. The evolutionary history of 2 , 658 cancers. 2020;578(August 2017).
45. Yao JC, Pavel M, Lombard-Bohas C, et al. Everolimus for the treatment of advanced pancreatic neuroendocrine tumors: Overall survival and circulating biomarkers from the randomized, Phase III RADIANT-3 study. *J Clin Oncol*. 2016;34(32):3906-3913. doi:10.1200/JCO.2016.68.0702
46. Kasajima A, Pavel M, Darb-Esfahani S, et al. mTOR expression and activity patterns in gastroenteropancreatic neuroendocrine tumours. *Endocr Relat Cancer*. 2011;18(1):181-192. doi:10.1677/ERC-10-0126
47. Gelsomino F, Casadei-Gardini A, Caputo F, et al. Mtor pathway expression as potential predictive biomarker in patients with advanced neuroendocrine tumors treated with everolimus. *Cancers (Basel)*. 2020;12(5):1-12. doi:10.3390/cancers12051201
48. Thomas P, Starlinger J, Vowinkel A, Arzt S, Leser U. GeneView: A comprehensive semantic search engine for PubMed. *Nucleic Acids Res*. 2012;40(W1):585-591. doi:10.1093/nar/gks563
49. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: Exploring the world's knowledge of somatic mutations in human cancer. *Nucleic Acids Res*. 2015;43(D1):D805-D811. doi:10.1093/nar/gku1075
50. Banck MS, Ames MM, Andreas S, et al. The genomic landscape of small intestine neuroendocrine tumors Find the latest version : The genomic landscape of small intestine neuroendocrine tumors. 2013;123(6):2502-2508. doi:10.1172/JCI67963DS1
51. Cao Y, Gao Z, Li L, et al. Whole exome sequencing of insulinoma reveals recurrent T372R mutations in YY1. *Nat Commun*. 2013;4. doi:10.1038/ncomms3810
52. Yuan F, Shi M, Ji J, et al. KRAS and DAXX/ATRX gene mutations are correlated with the clinicopathological features, advanced diseases, And poor prognosis in Chinese patients with pancreatic neuroendocrine tumors. *Int J Biol Sci*. 2014;10(9):957-965. doi:10.7150/ijbs.9773
53. Childs LH, Mamlouk S, Brandt J, Sers C, Leser U. SoFIA: A data integration framework for annotating high-throughput datasets. *Bioinformatics*. 2016;32(17):2590-2597. doi:10.1093/bioinformatics/btw302

54. Maksimovic J, Gordon L, Oshlack A. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol*. 2012;13(6):1-12. doi:10.1186/gb-2012-13-6-r44
55. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al. Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics*. 2014;30(10):1363-1369. doi:10.1093/bioinformatics/btu049
56. Chen YA, Lemire M, Choufani S, et al. Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. *Epigenetics*. 2013;8(2):203-209. doi:10.4161/epi.23470
57. Croft D, O'Kelly G, Wu G, et al. Reactome: A database of reactions, pathways and biological processes. *Nucleic Acids Res*. 2011;39(SUPPL. 1):691-697. doi:10.1093/nar/gkq1018
58. Hovestadt V, Zapatka M. conumee: Enhanced copy-number variation analysis using Illumina DNA methylation arrays. R package version 1.9.0, http://bioconductor.org/packages/conumee/.

Acknowledgement

This work was supported by the BMBF e:Med initiative MAPTor-NET (031A426A, grant to CS, KD, MP), the Deutsche Krebshilfe (grant to CS), the CompCancer Graduate School (DFG, TS), the BSIO (DFG; TS). DigiWES proteomic data was generated by Natural and Medical Sciences Institute, at University of Tübingen. Methylation data generation was performed at Institute of Neuropathology, Charite University of Medicine, Berlin. We would like to acknowledge Torsten Gross, and Thomas Sell for their helpful tips in coding and data analysis.

Contributions

TS: study concept and design; interpretation of data; statistical analysis; technical, or material support; writing of the manuscript. SM: study concept and design, writing of the manuscript; study supervision. PR: writing of the manuscript. FB: interpretation of data and methylation analysis pipeline; critical revision of the manuscript for important intellectual content. AM, DT, KWM and MP: technical, and material support. SK: PanNEN DNA sequencing panel generation. DL: analysis and interpretation of FISH data. DC: analysis and interpretation of data, technical support. F.Ruoff and MT: proteomics experiments and initial analysis. BK: analysis and interpretation of proteomic data. UL: PanNEN DNA sequencing panel generation. KD: clinical analysis. F. Rossner: clinical analysis. JH: clinical analysis. MM: critical revision of the manuscript for important intellectual content. NB: analysis and interpretation of data. DH: analysis and interpretation of data. MP: clinical analysis and interpretation of data. CS: study concept and design; interpretation of data; writing of the manuscript; critical revision of the manuscript for important intellectual content; obtained funding; study supervision

Ethics Statements

The collection and use of human derived neuroendocrine neoplastic specimens and controls obtained from the Medical University of Graz was approved by the Ethics

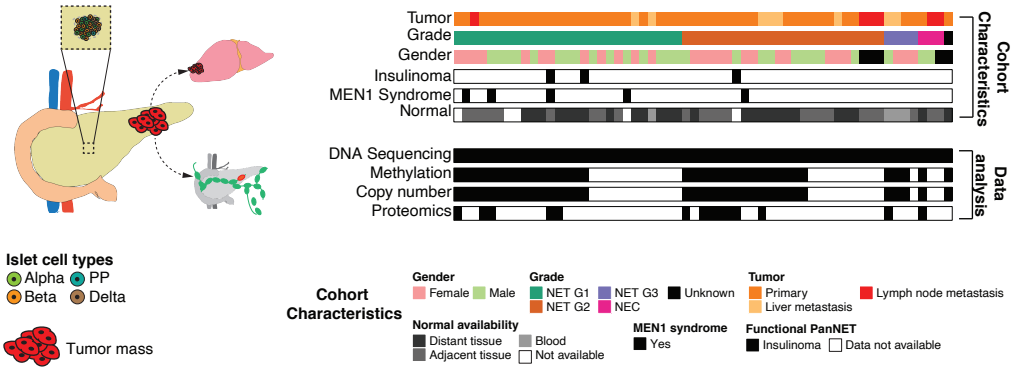
Committee of the Medical University of Graz, Graz, Austria, according to the ethical guidelines of the 1975 Declaration of Helsinki (EK 27-030 ex 14/15 and EK 29-455 ex 16/17). Tissue collection obtained from Charite Unversity of Medicine processed according to the Charité ethics vote EA4/022/15

Conflict of Interest:

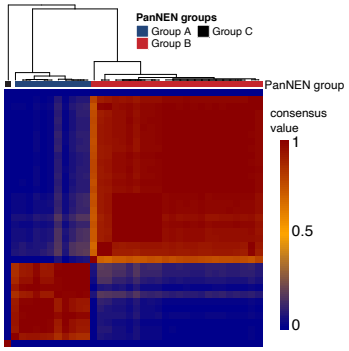
DC has a patent pending: DNA methylation-based method for classifying tumour species (EP16710700.2)

figure 1

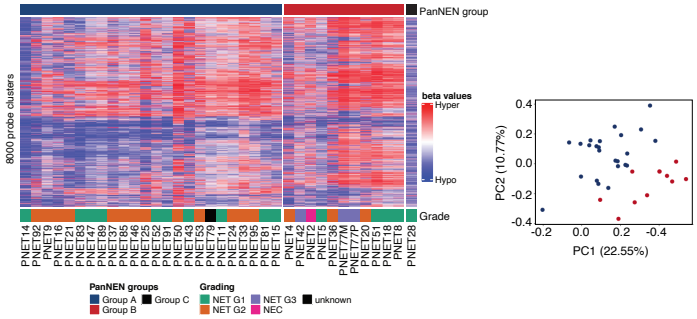
a



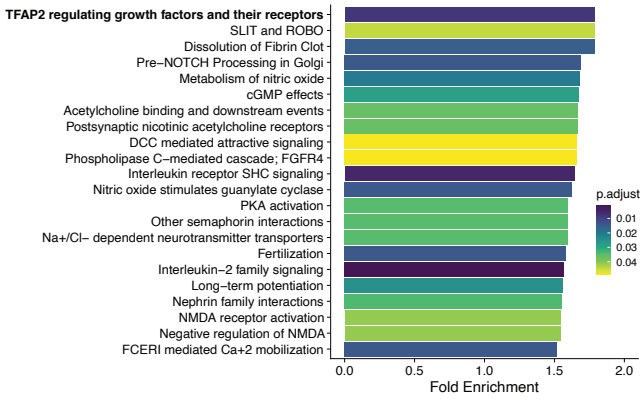
b



c



d



e

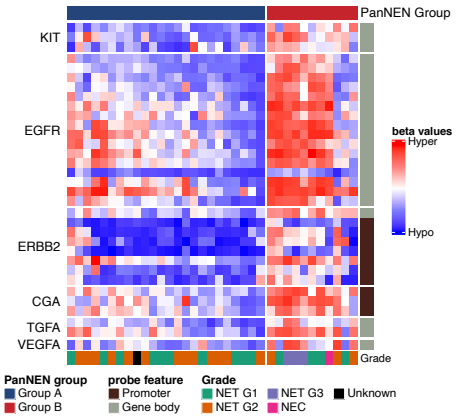
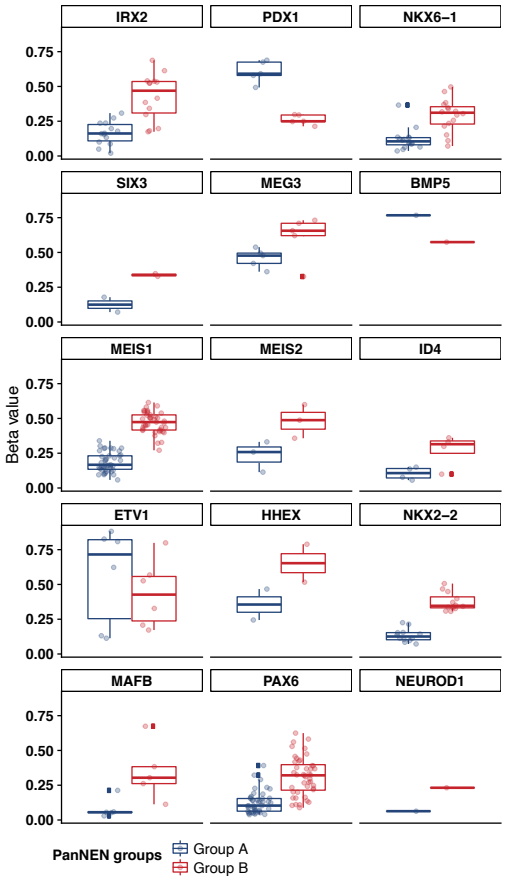


figure 2

a

b



c

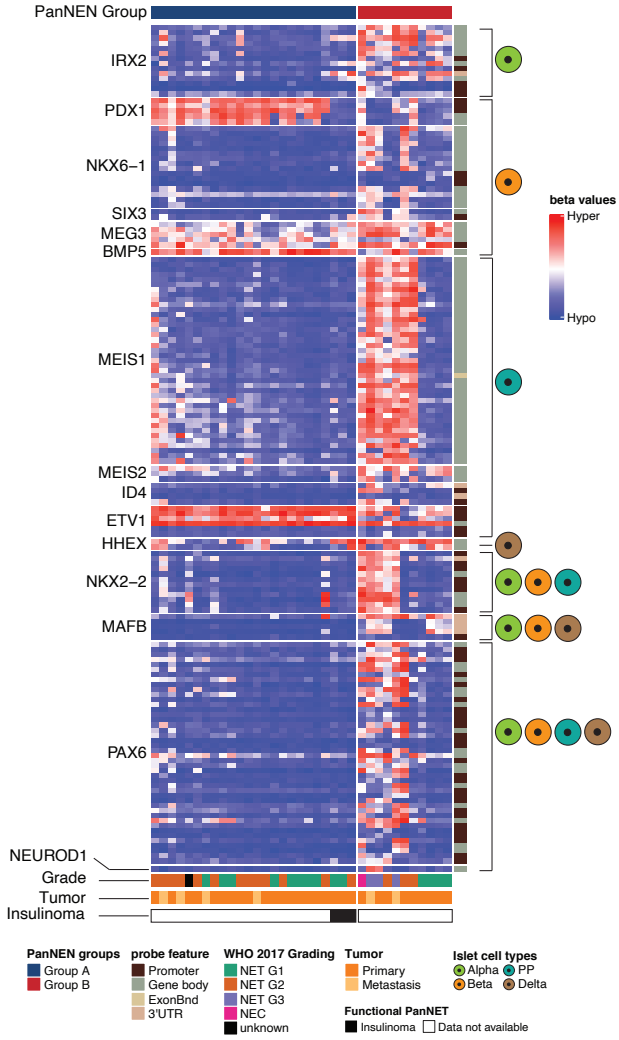


figure 3

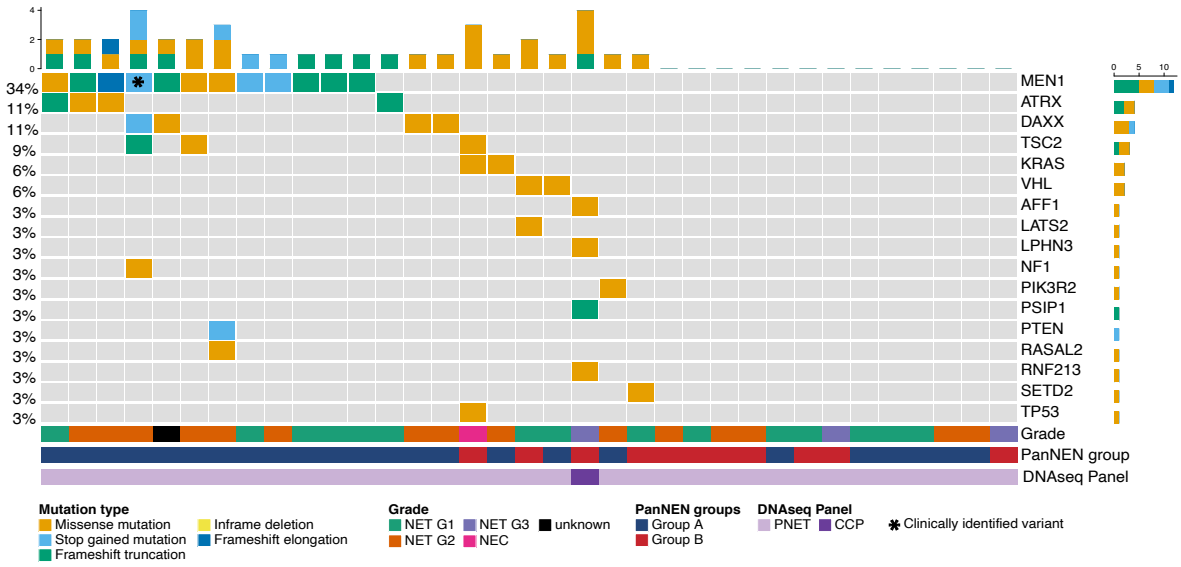


figure 4

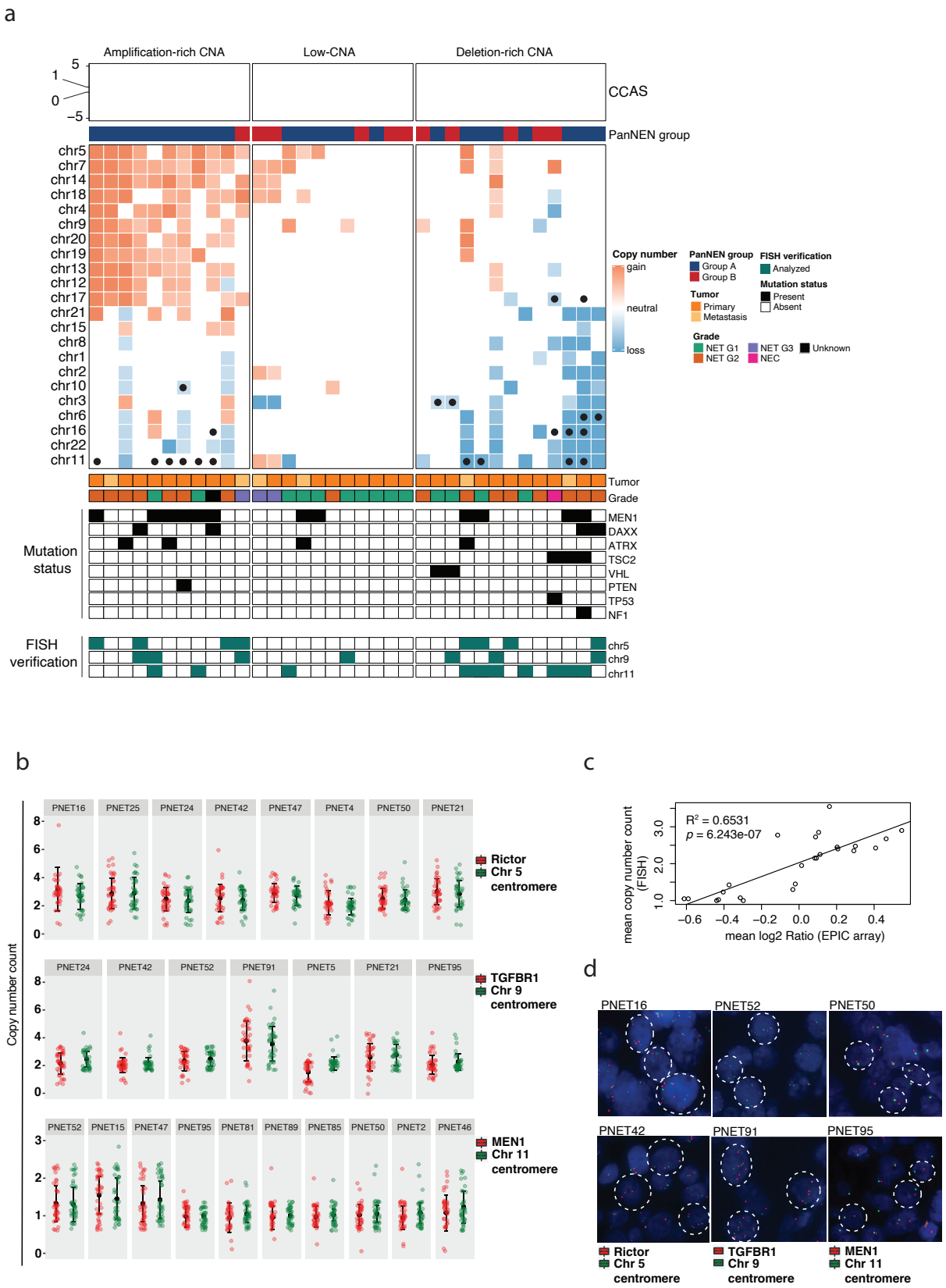
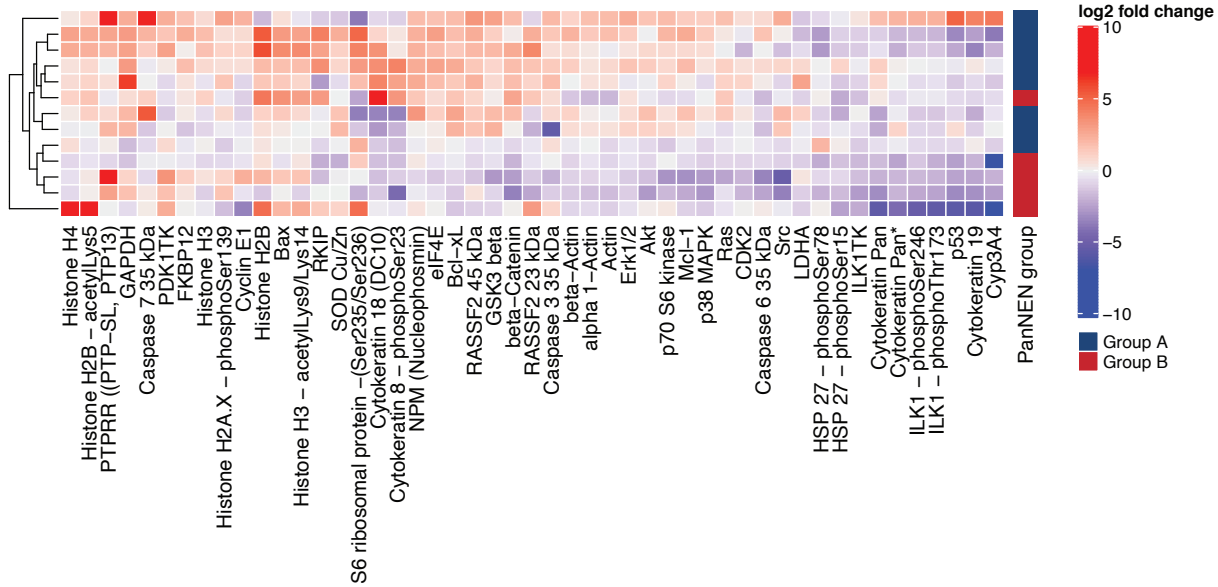


figure 5

a



b

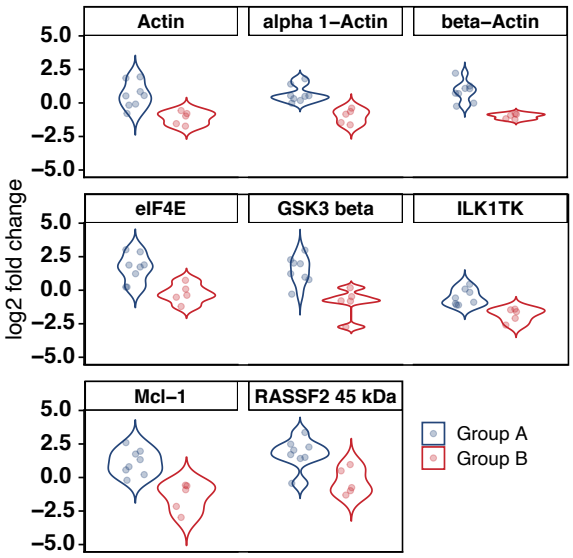
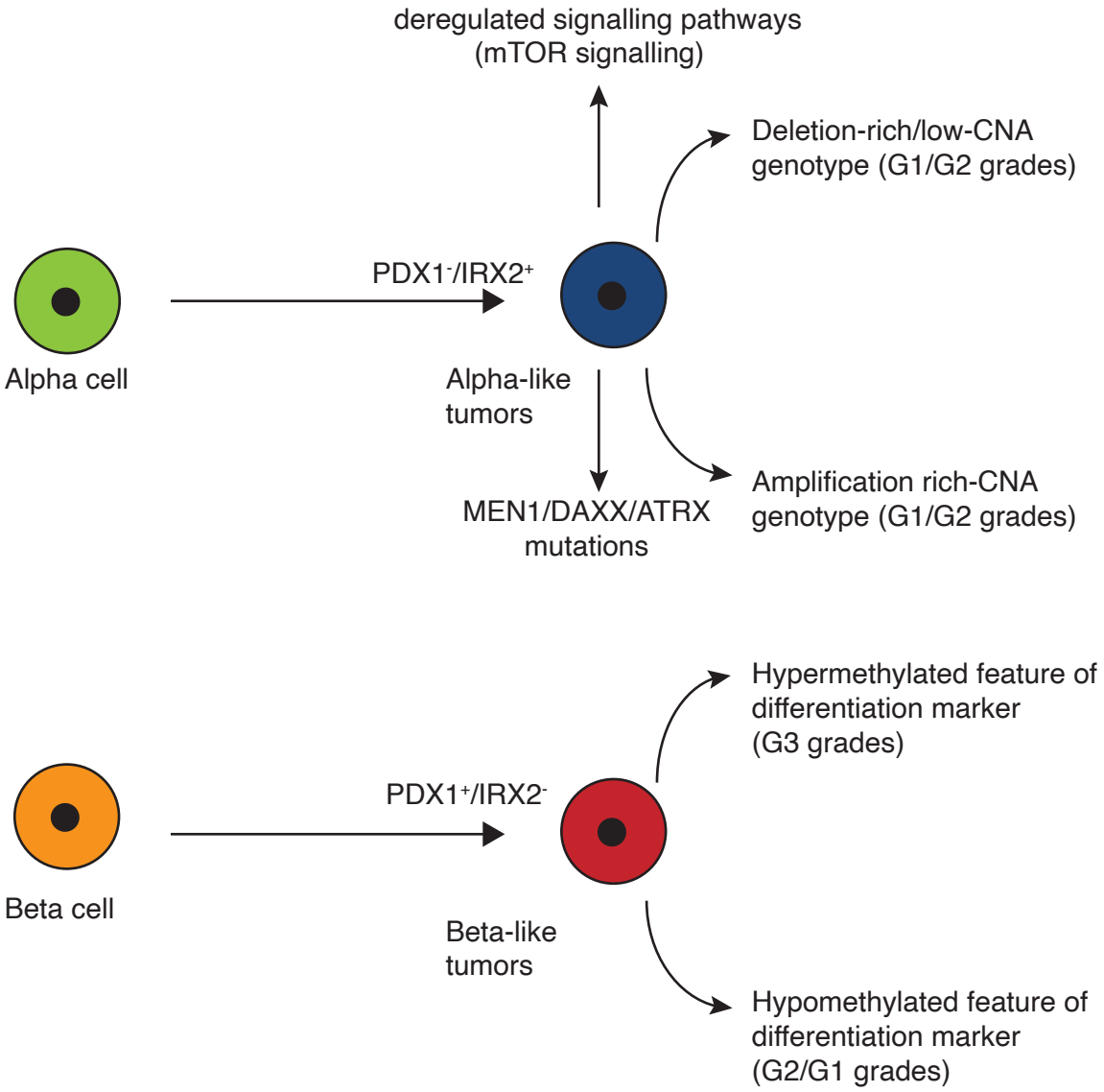
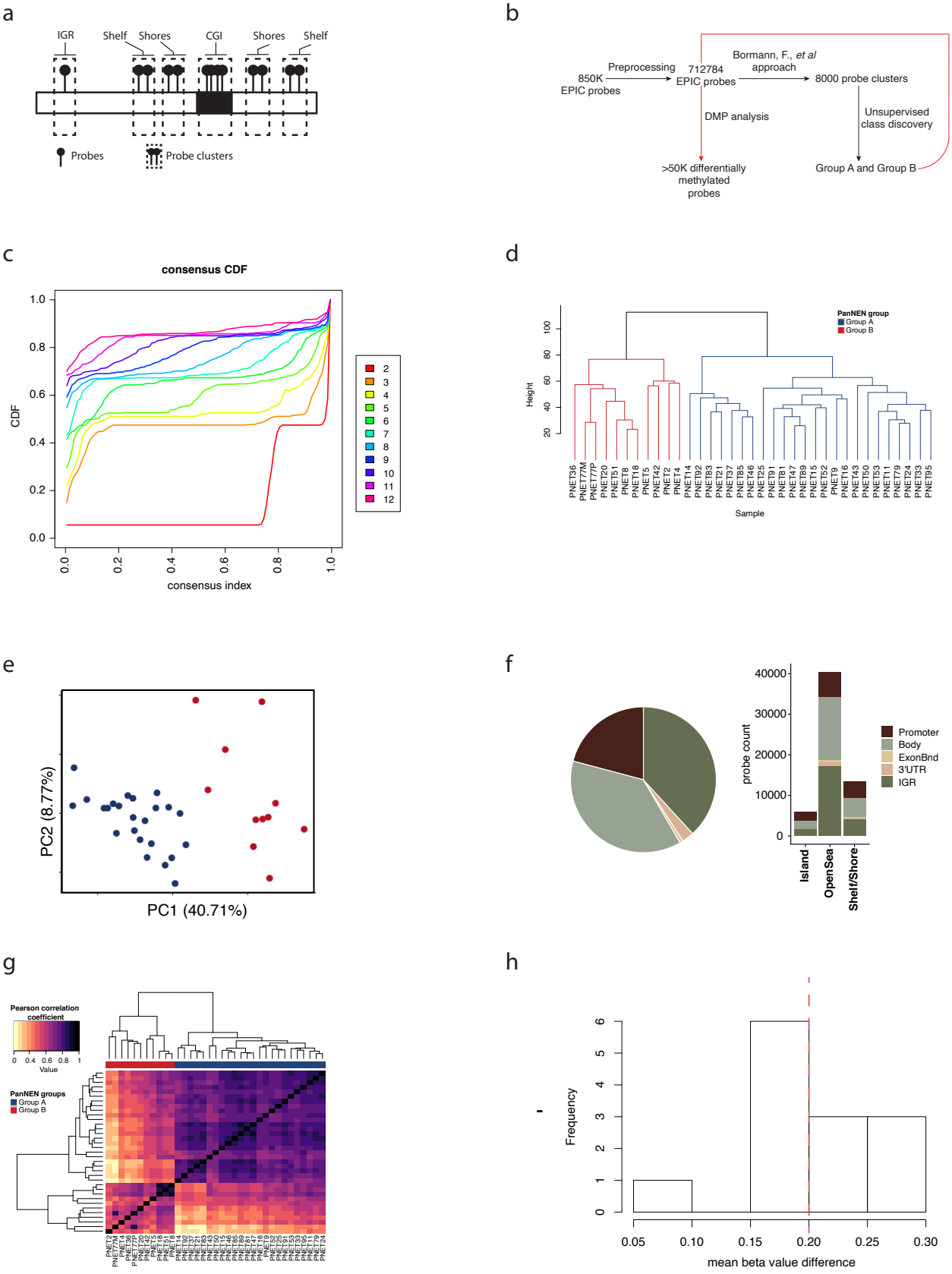


figure 6

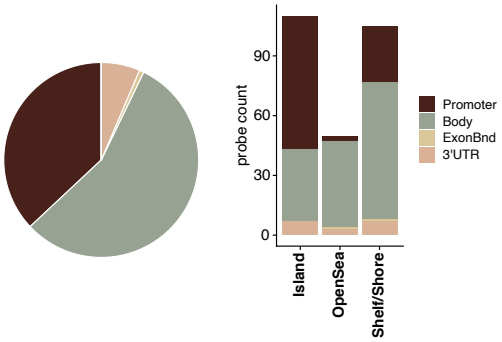


Supplementary figure 1

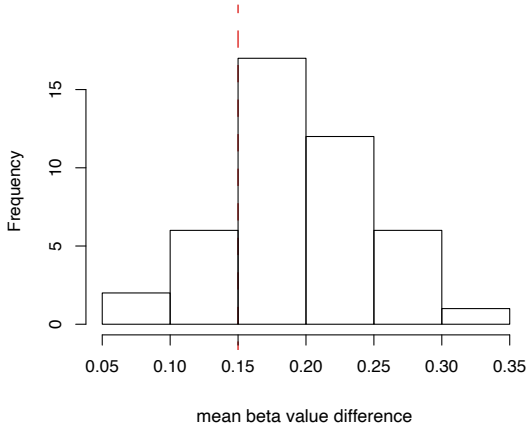


Supplementary figure 2

a

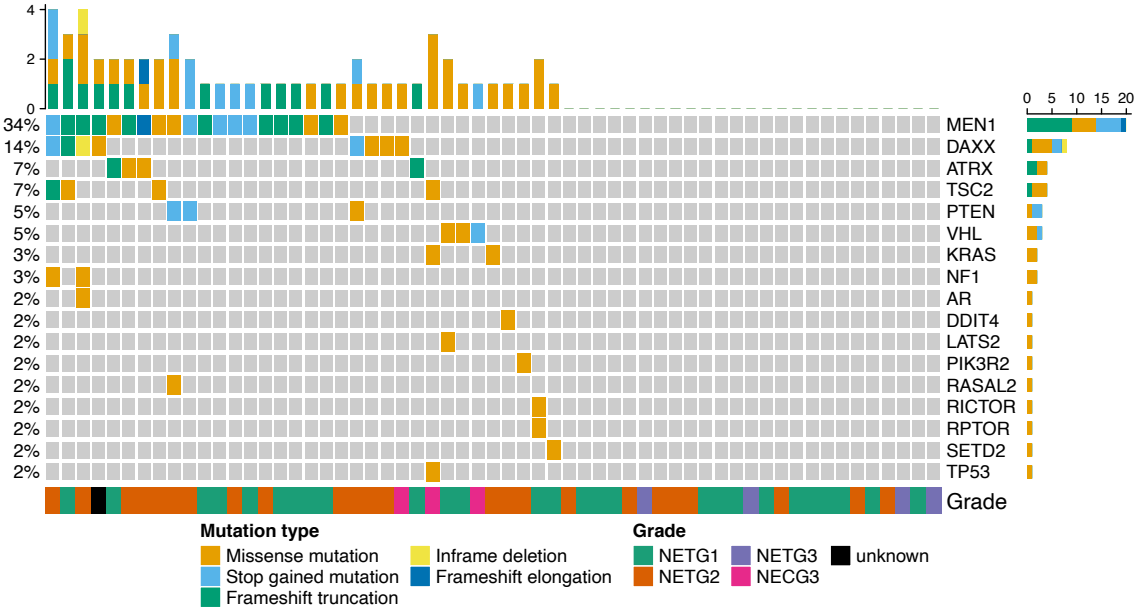


b

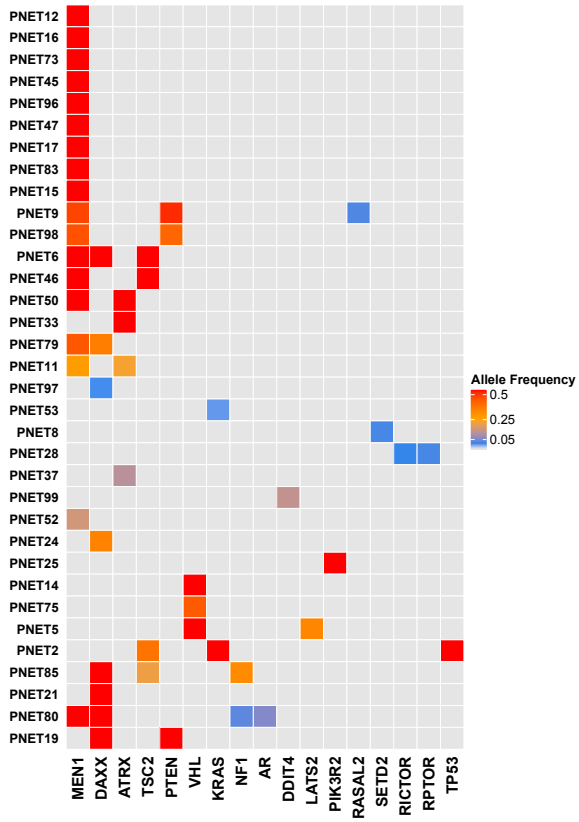


Supplementary figure 3

a

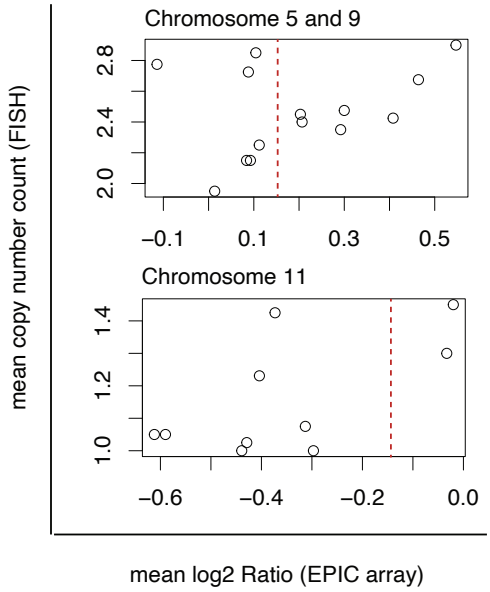


b



Supplementary figure 4

a



b